Dynamical System and Information Geometry

- A Complementary Approach to Complex Systems -

Masatoshi Funabashi

Centre de Recherche en Epistémologie Appliquée
Ecole Polytechnique

THESE présenté pour l'obtention du grade de
Docteur de l'Ecole Polytechnique

Discipline: Science des Systèmes Complexes

Octobre 2010

Reviewers

Jean-Jacques Szczeciniarz

Professor, Université Paris Diderot (Paris 7)

Frank Nielsen

Professor, Ecole Polytechnique
Researcher at Sony Computer Science Laboratory, Japan

Judges

Jean Petitot

Thesis supervisor
Professor, Ecole Polytechnique

Paul Bourgine

Professor, Ecole Polytechnique
Director of CREA

Giuseppe Longo

Professor, Ecole Normale Supérieure Paris

Mario Tokoro

President and CEO, Sony Computer Science Laboratories, Inc. (Japan)

Day of the defense

15 October 2010

色即是空 空即是色

Every form in reality is empty,

and emptiness is the true form.

-Heart Sūtra

ABSTRACT

Recently emerging complex systems sciences tackle the systems where complex interactions between components lead to the manifestation of emergent property linking different levels of organization. This thesis aims to reveal the mechanism of emergent property in complex systems, both in concrete modeling as well as comparative analysis between different systems. We tackle various subjects in complex systems science with newly proposed unified theoretical framework, based on the dialectic between dynamical system theory and information geometry. The thesis has therefore two levels of objectives: 1) Modeling and understanding of concrete complex systems with the use of constructive and interaction-analytical methodologies, and 2) comparison between different complex systems to characterize universal structure of emergence.

The thesis consists of 7 Parts, in which Part 2 to 6 correspond to the first objective, and the Part 7 to the second one:

In Part 1, we review the historical context of complex systems science and propose a dialectical strategy between the constructive and interaction-analytical methodology, based on the dynamical system theory and information geometry, respectively.

In Part 2, we treat a candidate model of brain cortex dynamics known as "chaotic itinerancy", and incorporate the effect of autonomous learning seeking for the creativity of intelligence as emergent property of neural system. The interpretation of emergence in terms of the internal measurement theory is extended to derive the concept of "chaotic itinerancy as catalyst of learning".

In Part 3, the dynamics of chaotic neural network is applied to emergent collective behavior of robots, so that to realize optimal intermittent search of sporadic information. The effectiveness of the collective infotaxis is analyzed on a simulator basis.

In Part 4, we define novel complexity measures from information geometrical point of view and apply to the analysis of social network data. The established complexity measures play a key role in comparative analysis between different systems in Part 7.

In Part 5, we apply the dialectical strategy between dynamical system and information geometry toward the understanding of morphogenesis during zebrafish embryogenesis. Theoretical propositions are tested with tentative experimental data from two european projects, Embryomics and BioEMERGENCES.

In Part 6, complex systems related to linguistics are investigated. We discovered novel invariants and geometrical relation between japanese vowels, as a system-level emergent property. Ecological modeling approach to multilingual environment is also proposed along the dialectical strategy between linguistic theory and mathematical modeling.

In Part 7, we review the obtained results in previous Parts with comparative perspective, seeking for a characterization of universal structure of emergence in terms of the organization of interactions that does not explicitly depend on the property of components. Comparison between Part 2 and 4, as well as 5 and 6, derived candidate qualitative dynamics of emergence and its detection strategy as the dynamics and constraint between functors and meta-functors. Further possibility of the proposed strategy is discussed.

Système Dynamique et Géométrie Informationnelle

-Une Approche Complémentaire aux Systèmes Complexes-

RÉSUMÉ

Un des défis majeurs de la science de complexité se situe à l'investigation de l'émergence, où les interactions entre les composants microscopiques d'un système produisent la propriété globale, et réciproquement, la dynamique globale influence le bas niveau. Cette thèse a comme ambition de 1) élucider le mécanisme sous-jacent des systèmes complexes par la modélisation concrète des systèmes réels, et aussi 2) comparer entre les différents modèles proposés pour détecter la condition universelle de l'émergence. Pour cela, nous développons la nouvelle méthodologie basé sur l'interaction entre la théorie des systèmes dynamiques et la géométrie informationnelle, afin d'avoir la dialectique entre la modélisation constructive/déterministe et l'analyse des interactions sous la formalisation stochastique.

La thèse se compose de 7 Parties, parmi lesquelles la Partie 2 à 6 correspondent au premier objectif, et la Partie 7 au seconde.

Dans la Partie 1, nous allons réviser l'histoire de la science de la complexité et proposer la stratégie dialectique entre les méthodologies constructive et interactions-analytique, basé sur la théorie de système dynamique et la géométrie informationnelle.

En Partie 2, nous traitons un modèle de réseau neuronal avec le comportement chaotique nommé "l'itinérance chaotique" comme un candidat de la dynamique du cortex cérébral, et analysons l'effet de l'apprentissage autonome sans superviseur comme une source de créativité qui est la propriété émergente du système neuronal. La théorie de la mesure intérieure est étendue afin de interpréter l'émergence des nouveaux attracteurs par "le chaos comme le catalyseur d'apprentissage."

En Partie 3, nous avons appliqué la dynamique du réseau neuronal chaotique aux robots qui manifestent la dynamique de recherche collective de manière émergente, au défi de la détection optimale des informations sporadiques. L'efficacité de la recherche collective est évaluée avec un simulateur virtuel.

En Partie 4, nous développons les nouvelles mesures de la complexité du point de vue de la géométrie informationnelle, et analysons les données des réseaux sociaux. Les mesures de la complexité jouera un rôle principal dans la Partie 7.

En Partie 5, nous appliquons la stratégie dialectique entre le système dynamique et la géométrie informationnelle vers la compréhension de la morphogenèse lors de l'embryogenèse chez le poisson zèbre. Quelques propositions théoriques sont établies et testées avec les données tentatives dérivées des projets européens Embryomics et BioEMERGENCES.

En Partie 6, nous analysons les systèmes complexes liés au linguistique. Nous avons découvert les nouveaux invariants et la composition géométrique entre les voyelles japonaises, qui sont les propriétés émergentes au niveau du système. Nous développons aussi la modélisation écologique de l'environnement multilingue dans un contexte de la dialectique entre la théorie linguistique et la modélisation mathématique.

En Partie 7, nous révisons les résultats obtenus dans les Parties précédentes sous une perspective comparative, en vue de détecter la structure universelle de l'émergence comme l'organisation des interactions qui ne dépende pas explicitement sur la propriété des composants. Surtout la comparaison entre les Parties 2 et 4, ainsi 5 et 6, nous indique la typologie et la stratégie de détection de la dynamique de l'émergence comme la relation et le contraint entre les foncteurs et méta-foncteurs. D'autre possibilité d'application de la stratégie établie est mise en discussion.

Résumé des Chapitres

Partie I Introduction Générale

Chapitre 1 Qu'est-ce que sont les systèmes complexes ?

Nous introduisons les concepts des systèmes complexes et les termes épistémologiques en révisant l'histoire de la science et de la modélisation. Nous proposons ainsi une nouvelle approche complémentaire basé sur les théories de système dynamique et géométrie informationnelle. Cette méthodologie jouera un rôle principal et transversal dans l'ensemble de cette thèse.

Mots clés: Système complexe, émergence, système dynamique de haute dimension, chaos, observation intérieure, méthodologie constructive, méthodologie interaction-analytique

Chapitre 2 Préliminaires de système dynamique

Préliminaires de système dynamique.

Mots clés: Système dynamique à temps discret, bifurcation, crise d'attracteur, itinérance chaotique, exposant de Lyapunov

Chapitre 3 Préliminaires de géométrie informationnelle

Préliminaires de géométrie informationnelle.

Mots clés: Variété statistique, métrique Riemannien, α-connexion, coordonnés double-plat, Kullback-Leibler divergence, théorème de Pythagore étendu

Partie II Modélisation synthétique avec réseau neuronal chaotique

Chapitre 4 Modélisation synthétique de l'apprentissage autonome avec réseau neuronal chaotique

Nous étudions un modèle de réseau neuronal chaotique (CNN) qui manifeste la dynamique intermittente et chaotique nommée l'itinérance chaotique, en interaction avec les règles d'apprentissage autonome sans superviseur. Premièrement, nous analysons la stabilité du modèle de manière hiérarchique selon la structure des espaces invariants. La transition irrégulière entre les ruines d'attracteur est observée avec le valeur positif de l'exposant de Lyapunov, provoqué par la blowout bifurcation des sous-espaces où les attracteurs sont situés. Cette crise est associée de la génération de la structure ridée des bassins. Ensuite, nous modélisons 2 sortes d'apprentissage autonome avec les règles classiques, de Hebb et STDP, afin d'observer l'effet de la dynamique de l'itinérance chaotique sur l'apprentissage. L'apprentissage de Hebb a augmenté le temps de résidence aux ruines d'attracteur et le degré de synchronisation entre les neurones, en produisant la modularité uniforme dans la dynamique et réalisa l'émergence des nouveaux attracteurs dans l'espace convergé des ceux mémorisés auparavant. L'apprentissage avec STDP a contrairement réduit le temps de résidence aux ruines d'attracteur, mais introduit une large variété de périodicité aux attracteurs émergents, indiquant la possibilité de l'apparition des attracteurs étranges. Les deux règles d'apprentissage autonome ont détruit et préservé les espaces invariants du modèle de manière sélective mais différente, selon les règles et la dynamique de synchronisation des neurones sur les sous-espaces où les orbites sont situées.

Mots clés: Itinérance chaotique, espace invariant, blowout bifurcation, bassin ridé, apprentissage de Hebb, apprentissage de STDP.

Méthodologie: Modélisation des apprentissages autonomes avec CNN comme un système dynamique → Analyse de stabilité linéaire de CNN comme un système dynamique → Analyse de corrélation neuronale émergente par la géométrie informationnelle

Chapitre 5 Mesure intérieur de l'itinérance chaotique comme catalyseur d'apprentissage

Nous révisons et étendre la "théorie orientée à la mesure" de Y.-P, Gunji, afin d'interpréter le processus émergent des nouveaux attracteurs dans le CNN étudié au chapitre précédent. Les résultats peuvent être interprétés avec une extension d'un fameux concept du "chaos comme catalyseur d'apprentissage," qui trouve aussi le lien avec la théorie de la mesure intérieure. L'émergence des nouveaux attracteurs peuvent être considérés comme un processus vers la réalisation de pseudo-solution, qui se génère de la dynamique d'apprentissage autonome basée sur la mécanisme intrinsèque du système, et qui effectue la modification de l'interface au sens de la mesure intérieure afin de résoudre l'état de contradiction que représente l'itinérance chaotique.

Mots clés: Mesure intérieure, physique orienté à la mesure, pseudo-solution

Méthodologie: Formalisation du chaos dans le système dynamique comme un observateur intérieur → Formalisation du CNN comme un observateur intérieur → Interprétation des résultats des apprentissages autonomes avec la théorie orientée à la mesure

Partie III Application de réseau neuronal chaotique en robotique émergente

Chapitre 6 Introduction de neurorobotique

Nous introduisons le concept de "neurorobotique" comme une source d'invention technologique et le champs de recherche interdisciplinaire situé entre les neurosciences et la robotique. Le contexte des projets de la partie 3.

Mots clés: Neurosciences, robotique, énaction

Chapitre 7 Équilibration entre l' autonomie et la réponse à l' environnement avec la dynamique chaotique hiérarchique

La structure hiérarchique entre les espaces invariants de la dynamique chaotique déterministe du CNN est analysée et mise en application dans la robotique. Il s'est révélé que la dynamique intrinsèque de l' itinérance chaotique du CNN permet de contrôler la synchronisation d'activité des neurones de manière sélective à la hiérarchie, aussi de réagir rapidement aux entrés sensoriels extérieurs. Avec les configurations des paramètres appropriées, le robot a réalisé le mouvement chaotique de recherche entre tous les combinaisons possibles des trois directions par le simple contrôle de synchronisation partielle des neurones. L' entré extérieur faible est montré suffisant pour avoir le dérive macroscopique à la direction visée avec la fluctuation chaotique. Le robot a démontré la capacité de contourner les obstacles par la simulation virtuelle.

Mots clés: Réseau neuronal chaotique, espace invariant, itinérance chaotique

Méthodologie: Modélisation d'un robot itinérant utilisant la dynamique hiérarchique du CNN comme le système dynamique complètement déterministe → Analyse stochastique du comportement et la réponse à l' entré extérieur du système

Chapitre 8 Recherche intermittente optimale avec le robot chaotique itinérant et infotaxis par l'interaction collective

Nous allons appliquer le robot modélisé au chapitre précédent à la recherche collective des informations dans leur environnement. Un modèle stochastique de recherche en 2 phases basé sur le comportement des animaux est étendu à la recherche des informations sporadiques sur la surface de 2 dimensions. Le réseau de contacte entre les robots est analysé, à partir duquel nous modélisons la recherche collective en partageant les informations détectés dans l'environnement par la communication locale entre les robots. L'efficacité de la recherche collective est examiné sur un simulateur virtuel.

Mots clés: Recherche d'information sporadique en 2 phase, infotaxis, interaction collective, partage d'information

Méthodologie: Modélisation de la recherche en 2 phase du robot comme le système dynamique complètement déterministe basé sur l'analyse stochastique du chapitre précédent → Analyse des paramètres optimales sur un simulateur virtuel sous la formalisation stochastique → Modélisation et simulation de la recherche collective des informations comme un ensemble des systèmes dynamiques complètement déterministes → Construction de la théorie d'évaluation de la recherche avec la géométrie informationnelle

Partie IV Mesures de la complexité et application à l'analyse de la dynamique de réseau social

Chapitre 9 Dynamique des corrélations ordonnées dans bases de données de texte

Nous appliquons la géométrie informationnelle pour l'analyse de cooccurrence du réseau des mots au-delà de la corrélation de seconde ordre. Afin de traiter tous les informations hiérarchiques contenu dans l'ensemble de corrélation entre les variables, la dynamique du méta-niveau est considéré. Le résultat de l'analyse des weblogs politiques démontre qu' il existe une structure riche de la dynamique à la fois au niveau de l'ordre de corrélation et aussi au méta-niveau de corrélation.

Mots clés: Clustering hiérarchique, géométrie informationnelle, réseau de cooccurrence, weblog politique, communauté épistémique

Méthodologie: Définition des corrélations à chaque ordre comme contexte et ses méta-contextes avec la géométrie informationnelle → Analyse des tendances de la dynamique des contextes et méta-contextes comme système dynamique → Analyse de la carte de retour et de la stabilité locale des contextes et méta-contextes avec la formalisation du système dynamique

Chapitre 10 Décomposition de réseau: Une approche de la théorie informationnelle

Nous considérons la représentation du graphe des modèles stochastiques avec en général n variables binaires, et construisons un cadre théorique qui nous permet de mesurer tous les degrés des interactions existants entre tous les sous-systèmes, aussi bien ceux représentés par chaque lien entre les noeuds. L'étude de cas sur les weblogs politiques est démontré.

Mots clés: Géométrie informationnelle, réseau complexe, clustering hiérarchique

Méthodologie: Formalisation de la décomposition du système et le coupure du lien par la géométrie informationnelle → Analyse des tendances des weblogs politiques comme système dynamique

Chapitre 11 Reconstruction des liens de graphe comprenant les interactions ordonnées et substantives en référence multi-noeuds

Nous étudions de nouveau les informations représentées par les liens du réseau définies comme "l'information du lien" dans le chapitre précédent, afin de reconstruire le réseau à grande échelle à partir des données. La définition de l'information du lien est étendue afin de distinguer les interactions de chaque ordre entre corrélation positive et négative, de manière compatible avec le théorème de Pythagore. Pour résumer, une nouvelle définition de l'information du lien qui réuni ses décompositions est établie. Du côté de la reconstruction du réseau à grande échelle, nous proposons une stratégie pour la réduction du temps de calcul en limitant les combinaisons des variables à ceux qui contient les interactions substantives. Les mesures de l'information du lien et la méthodologie de la reconstruction efficace du réseau sont mis en application avec les jeux de donnée des weblogs politiques.

Mots clés: Géométrie informationnelle, interaction de chaque ordre, information du lien, réduction des paramètres, weblog politique, réseau social

Méthodologie: Décomposition de l'information du lien à chaque ordre de l'interaction et formalisation de la reconstruction du réseau de corrélation en référence multi-échelle par la géométrie informationnelle → Application à l'analyse des weblogs politiques comme système dynamique

Chapitre 12 Mesures de la complexité par rapport à la facilité de décomposition de système

Plusieurs mesures de la complexité de point de vue de la théorie d'information sont inventées jusqu'à présent, qui sont en principe les produits arithmétiques des informations. Nous considérons une nouvelle classe de mesure de la complexité basé sur les produits géométriques du KL divergence, par nécessité de évaluer la facilité de la décomposition du système stochastique à plusieurs critères, la décompositionabilité du système. Cette nouvelle classe de complexité remplit aussi la condition aux limites des mesures de la complexité arithmétique, qui disparaît à l'état complètement déterministe ou aléatoire, et en outre par la présence d'un sous-système complètement indépendant. L'analyse du réseau sociale reconstruit des weblogs politiques a révélé la dynamique de chaque sous-système en relation avec le comportement global associé à l'augmentation de la complexité totale. Nous proposons une hypothèse théorique que la décompositionabilité de chaque sous-système peut être conçue comme une fonction de potentiel sur la variété statistique, qui caractérise certain aspect de la dynamique actuelle des données.

Mots clés: Mesure de la complexité, condition aux limites, décompositionabilité du système, moyen géométrique

Méthodologie: Définition des nouvelles mesures de la complexité par rapport à la décompositionabilité du système basé sur la géométrie informationnelle → Analyse des tendances des weblogs politiques comme système dynamique → Considération théorique sur les nouvelles mesures de la complexité comme le potentiel du flux géodésique avec le perspective système dynamique

Partie V Reconstruction d'embryogenèse

Chapitre 13 Description du projet

Le contexte du travail de la partie 5 qui est réalisé au sein des projets européens, Embryomics et BioEMERGENCES est présenté. Ce chapitre résume le défi des projets, et la stratégie globale qui définissent l' objectif des études développées dans cette partie.

Mots clés: Embryogenèse, poisson zèbre, reconstruction phénoménologique, reconstruction théorique

Chapitre 14 Vers la définition dynamique de tissu: Détection des corrélations signifiantes dans le déplacement de cellule

Nous établissons la définition du "tissu dynamique" qui se fond sur les paramètres dynamiques de résolution cellulaire dérivé des projets Embryomics et BioEMERGENCES, en contraste du concept traditionnel du "tissu statique" basé principalement sur les traits morphologiques et anatomiques. Pour ce but, nous construisons la théorie de test statistique par la géométrie informationnelle, afin de détecter les corrélations signifiantes et la non-uniformité dans le champs de vecteur de déplacement des cellules. Le résultat provisoire sur les jeux de donnée de l'embryon de poisson zèbre est obtenu, qui nous donne l'information sur la mode coordonnée de migration des cellules pour la futur reconstruction théorique. La mesure de la stabilité temporelle est aussi développée théoriquement à la base du tissu dynamique.

Mots clés: Tissu statique, tissu dynamique, déplacement de cellule, test statistique

Méthodologie: Définition du tissu dynamique par la géométrie informationnelle traitant les jeux de donnée expérimentals comme système dynamique → Analyse des jeux de donnée → Considération théorique sur la durée temporelle du tissu dynamique

avec le perspective système dynamique

Chapitre 15 Vers la caractérisation multi-échelle d'embryogenèse: Incorporation des variables biologiques dans clustering sémi-paramétrique

Nous considérons la stratégie qui fait le lien entre la reconstruction phénoménologique et théorique, par l' investigation du modèle phénoménologique avec les variables biologiques. Nous proposons une méthode de clustering basée sur les "descripteurs minimums" qui incorporent l'intérêt biologique pour caractériser la dynamique de l'embryogenèse. Les résultats provisoires sont obtenu, et la théorie pour la mesure de la variation inter-individuelle est développée à multi-échelle en extension spatiale et temporelle, afin d'évaluer l'effet du traitement des drogues sous condition expérimentale.

Mots clés: Descripteur minimum, algorithme EM, variation individuelle

Méthodologie: Définition du descripteur minimum pour la caractérisation des données primaires comme système dynamique à grande échelle → Clustering avec les descripteurs minimums de manière compatible à algorithme em de géométrie informationnelle → Proposition théorique pour la mesure de la variation individuelle des embryons basé sur la distance informationnelle et le perspective de système dynamique

Chapitre 16 Étude sur l' entropie morphogénétique par géométrie informationnelle

Nous développons l'analyse de l'arbre de lignage des cellules par la géométrie informationnelle. Nous considérons la probabilité de chaque cellule à devenir chaque champ morphogénétique, basé sur le concept de "l' entropie morphogénétique" défini par Miguel A. Luengo-Oroz, qui représente la quantité d'information nécessaire à spécifier complètement le profile de la futur différentiation d' une cellule. Telle mesure de différentiation est utile à comparer entre deux arbres de lignage avec différentes conditions comme avec/sans traitement des drogues. Nous démontrons théoriquement comment évaluer l'effet extérieur entre deux groupes qui chaqu' un est soumis des traitements différents, et proposons une série de quantité avec KL divergence, nommée "l' entropie morphogénétique relative" qui nous permet d'évaluer l'effet du traitement extérieur propagé à chaque génération. Ce chapitre est limité à l'étude purement théorique.

Mots clés: Entropie morphogénétique, entropie morphogénétique relative, arbre de lignage cellulaire, traitement polydrogue, géométrie informationnelle, KL divergence, information mutuelle

Méthodologie: Définition de l'entropie morphogénétique relative utilisant la distance informationnelle sur l'arbre de lignage cellulaire comme une mesure de la différentiation de cellule → Développement théorique de la décomposition de l'effet extérieur à chaque génération par la géométrie informationnelle

Partie VI Systèmes complexes dans linguistique

Chapitre 17 Invariance dans le système des voyelles

Nous modélisons le système des voyelles de la langue japonaise en appliquant la géométrie informationnelle de la distribution normale aux deux premiers formants. Nous étudions la distribution de KL divergence à plusieurs échelles et la relation entre les composants séparés par l'ordre de statistique, afin de détecter le trait invariant au niveau systémique. Le résultat implique que malgré l'existence d'une forte fluctuation au niveau individuel, la distribution de KL divergence au niveau de population converge à la distribution log-normale, de manière invariante entre homme et femme. Cette distribution de population peut être donc considéré comme la distribution invariante des locuteurs du japonais standard. En outre, nous montrons que la relation entre les composants logarithmique du moyen et de la variance de KL divergence est linéaire au niveau de population. La signification de cettes invariances est contextualisée par l'étude bibliographique dans la discussion.

Mots clés: Géométrie informationnelle, formant de voyelle, distribution Gaussienne 2-dimensionelle, distribution log-normale, KL divergence

Méthodologie: Formalisation par la géométrie informationnelle de la distribution des formants du système des voyelles → Analyse de la distance informationnelle entre les voyelles et ses composants de chaque ordre de statistique

Chapitre 18 Étude sur la composition géométrique de système des voyelles

Nous continuons l'étude sur le système des voyelles du côté de dépendance statistique par le moyen de géométrie informationnelle. Commençant par l'analyse de corrélation, la dépendance sur la m-géodésique entre les voyelles est examinée qui correspond au chemin de projection de KL divergence au chapitre précédent.

Nous montrons que les cinq voyelles du japonais se situent approximativement dans une surface de deux dimensions dans l'espace des coordonnés η, et la fluctuation du circoncentre défini sur la m-géodésique entre les voyelles contient le bruit incorrélé des voyelles.

La localisation des circoncentres sur les coordonnés θ est examinée par rapport au valeur du déterminant $|S|$ dans la distribution des formants.

Le résultat démontre l'origine de la distribution log-normale invariante de KL divergence découvert au chapitre précédent, qui implique la condition cognitive liée à la dépendance quadratique entre les voyelles sur les coordonnés η.

Mots clés: Géométrie informationnelle, formant de voyelle, circoncentre, déterminant de la distribution de formant

Méthodologie: Analyse de la composition géométrique sur les coordonnés double-plat de géométrie informationnelle → Définition du circoncentre entre les pairs des voyelles par géométrie informationnelle → Analyse de la composition géométrique et la fluctuation des circoncentres sur les coordonnés double-plat de géométrie informationnelle

Chapitre 19 Méthodologie pour la dialectique entre théorie linguistique et modélisation mathématique et application à la modélisation d' environnement multilingue avec processus de contacte

Nous cherchons le point de synergie entre la modélisation mathématique et la théorie conceptuelle des sciences humaines, notamment l'écolinguistique. Nous proposons deux étapes de l'interaction qui intègrent les deux côtés: Premièrement, nous formalisons la théorie linguistique de manière compatible à la modélisation mathématique en symbolisant proprement les observables comme les variables primaires, et ses fonctions avec suffisamment de profondeur logique. Ensuite, on étudie le modèle par le moyen mathématique et de simulation, et déduit les variables secondaires sur la dynamique simulée afin de étendre la théorie linguistique au niveau conceptuel.

Un exemple de cette modélisation dialectique est établi avec le processus de contacte, en vue de modéliser l'acquisition des langages dans un état multilingue. La dynamique de simulation a reproduit les patterns spatio-temporels de la propagation des états multilingues, qui contient les corrélations fortement signifiants de hauts ordres entre les langues. L'instabilité relative de l'état bilingue est observé. Les résultats trouvent certains points d'interprétation sur l'hypothèse de travail pour l'acquisition multilingue exercée par le Hippo Family Club.

Mots clés: Processus de contacte, environnement multilingue, multilinguisme, acquisition naturelle de langage, écologie

Méthodologie: Formalisation de la stratégie dialectique entre la théorie linguistique et la modélisation mathématique, de manière compatible à la modélisation avec le système dynamique et l'analyse par la géométrie informationnelle → Modélisation de l'environnement multilingue avec le processus de contacte comme le système dynamique stochastique → Analyse provisoire du processus de contacte sur la corrélation par la géométrie informationnelle → Développement théorique sur l'analyse de la propagation

de probabilité avec la perspective du système dynamique et la géométrie informationnelle

Partie VII Discussion générale

Chapitre 20 Discussion générale

Nous révisons les résultats d'étude effectué dans cette thèse et examinons les points de contribution au science des systèmes complexes par rapport à la stratégie complémentaire entre système dynamique et géométrie informationnelle introduit à la partie 1(Introduction Générale). Nous prenons comme exemples représentatifs la comparaison de la propriété émergente entre les organisations neuronales et sociales. La discussion est développée avec le perspective au plus large, en mettant l'accent sur la possibilité d' application y compris hors des sujets traités et la nouveauté de la stratégie proposée. Les résultats impliquent l'efficacité et le potentiel riche de cette stratégie au défi de la caractérisation universelle de l'émergence du point de vue relationnel.

Mots clés: Méthodologie constructive, méthodologie interaction-analytique, foncteur, méta-foncteur, organisation des interaction

Chapitre 20 Méthodologie dialectique entre système dynamique et géométrie informationnelle : Vers une typologie transversale des systèmes complexes

Ce chapitre résume en français la contribution méthodologique et métathéorique de cette thèse. Nous étudions d' abord l' ensemble des interactions possibles du système qui deviennent les paramètres dominants par rapport à l' augmentation de sa taille. Ensuite, nous développons la méthodologie dialectique entre la théorie de système dynamique et géométrie informationnelle afin de analyser l' effet des interactions comme origine des phénomènes émergentes. Application de cette méthode aux plusieurs systèmes à différentes échelles implique la possibilité de construire la typologie transversale de l' émergence, basé sur la nature statistique et la dynamique de l' organisation des interactions qui ne dépendent pas explicitement de la propriété des composants.

To mother nature, who offered me a son

and gave birth to my wife.

Remerciements

Je me permets de rédiger cette petite partie de thèse en français, la langue que j'ai vécu avec pendant la plupart de mon étude et qui était le moyen principal de la discussion au labo. Tout d'abord, merci à Paul de m' avoir initialement accepté de venir au CREA, j'étais heureux d'être dans ton équipe qui me laisse le maximum de liberté de pensé et de la choix, qui m'a permit de établir l'identité de chercheur et mener la recherche en grande autonomie. Je te trouve un représentant des trois valeurs de la république, le troisième étant le plus fort.

Je remercie sincèrement à mon directeur de thèse, Jean, de m' avoir aidé à accomplir ce travail toujours divergent jusqu' à la fin, tout en montrant le plaisir de théoricien. Entre les sciences et l' épistémologie, j'ai pu pleinement apprécié la richesse intellectuelle du CREA.

Je m'exprime ma plus grande gratitude aux chers membres de jurys, qui ont littéralement donné la peine de lire le manuscrit et élaborer la discussion lors de la soutenance et de la pré-soutenance. Vos critiques et confiances m'ont donné le plus de courage à franchir ma carrière dans les domaines scientifiques.

Je devrais noter sur la vraie dualité de cette thèse maintenant, d'une part la mode de recherche sérieux et intensif, et d'autre part l'ambiance conviviale et fêtard, qui restaient toujours "invariant" entre mes collègues. S'il y a l'universalité à la définition de l'humanité, mon hypothèse est de trouver des traits qui ne changent pas par ses profiles socioculturels. Les gens aiment savoir sur la différence, par exemple entre le Japon et la France, mais l'essentiel est ce qui ne changent pas entre les japonais et les français, les qualités que je trouve au profond d'un individu. Je remercie donc vivement à mes collègues au CREA qui m' ont laisser vivre ses moments de l'humanité universelle ou le méta-foncteur de l'humanité si je le formalise, plus particulièrement:

A David Chavalarias pour ta générosité océanique, à Jean-Philippe pour me faire découvrir les argots, parfumés de cynisme à la hauteur de l'élégance, à Camille, ton esprit pénétrant qui ne manque pas d'humeur, à Carla, le grand voyage en Amérique latine et le rhapsodie à Berlin est mon meilleur souvenir. A David Colliaux pour ta fascination au japon, j'espère que j'ai pas trop démistifié mon pay. A Raphael pour ta transformation totale avant / après la visite au japon. A Emmanuel pour tes bonnes humeurs fulgurantes et l'apaisement de culpabilité quand j'ai manqué l'avion à Malaga, à Benoît Lombardot pour nos rendez-vous chaotique à Shanghai, à Thierry pour tes bisous un peu piquants. Merci Lionel de partager discrètement le moment de concentration, à Gabriel et Dorothée pour les séances de cinéma.

Un grand merci aux équipes des projets Embryomics et BioEmergences, Louise, à tes petits oursins, à Sophie pour tes humeurs éclatantes, à Benoît Maury à ton aventure à Bratislava, à Miguel pour ta gentillesse et curiosité gourmande, et à Nadine d'avoir dirigé nous les grands gamins. D'akujem, molto grazie, muchas gracias, les équipes de Bratislava, Bologne, et Malaga pour les nuits blanches.

Je remercie également les collègues de l'ISC-PIF, Les habitants et les visiteurs qui ont discuté ensemble, surtout aux membres du projet MOMA et TINA d'avoir s' occuper de Enzo, et à René pour sa direction au bouteur anglo-saxon. Merci les membres de Gdt-ELA pour les jolis souvenirs à l'île de Berder, à Frédéric pour ta terminologie sportivement philosophique, à Alexandra pour ta charme, à Sven pour ta modestie hors-norme, et à Nadège pour les "étoiles toi et moi".

Je remercie collectivement les secrétariats du CREA et l'ISC-PIF de m'avoir soutenir du côté de l'administration et du voyage en mission, nottament à Marie-Jo pour ton encouragement chaleureux, à Nadiège à ta patience et ta sincérité, à Noemi pour ta caprice italienne, à Geneviève pour ton accueil toujours souriant et généreux.

Je remercie personnellement les amis hors de l'étude, les amis de mes parents, Francis Pavé, Marie-Christine Pouchelle, pour ses amitiés trans-générationelles, les dîners chez vous ont littéralement nourrit ma thèse et la compréhension sur la culture française.

Arigatou les amis de Hippo en France, pour me faire découvrir les régions et m' avoir invité chez eux comme un membre de la famille.

Merci les résidents de la Maison du Japon, le moment de jouer au foot avec les gamins de quartier était la meilleure détente et une source d'inspiration importante.

Sincère gratitude à mes parents de soutenir ma traversée nomade, à mon grand-père qui depuis ma naissance préparait mon compte pour accomplir mes études.

Mille mercis à Chizuru et Enzo d'être à côté.

Acknowledgement

I would like to acknowledge the grants and projects, and express my sincere gratitude to members out of french community that supported this study.

This study was supported by the long term study abroad support program of the university of Tokyo for the first 3 years, and by the French government (Promotion Simone de Beauvoir) for the first 2 years. Other short-term contracts of CNRS were provided from CREA for the rest of the working period, and several trip supports were provided from CROUS.

Works in Part III are the fruits of collaboration with Syunsuke Aoki, and were financially supported by Information-technology Promotion Agency in JAPAN.

Part IV was partially supported by MOMA, TINA projects at ISC-PIF and an ANR grant 'Webflunce' #SYSC-009-03.

The study in Part VI was greatly inspired by the discussion with Yo Sakakibara and members of Dragon Project. Discussion with Hiroto Kuninaka brought important insight on the origin of the observed lognormal distribution. Author would also thank members of Hippo Exchange Club in Singapore for answering to the questionnaire, Masanobu Shiratori for the blueprint of the questionnaire, Nobuko Yokokawa, Yeo Choon Hin Melvin, KK Lew, Shea Nee, Steve Lau, Akiko Okada and Hiroyasu Kawakami for supporting the survey, Yukiko Sakakibara, Tomoyuki Nogita and Eiichiro Osawa for the summary of data.

Contents

List of Figures

List of Tables

Part I

General Introduction

Chapter 1

What are Complex Systems ?

Abstract

In this chapter, we introduce the concept of complex systems and related epistemological terms looking back the history of science and modeling based mainly on [3] [4]. A complementary approach with the constructive methodology and interaction-analytical methodology is proposed with the foundation of dynamical system modeling and information geometrical analysis, which provides the consistent theoretical axis of this thesis.

Keywords: Complex systems, emergence, high-dimensional dynamical system, chaos, internal observation, constructive methodology, interaction-analytical methodology

1.1 What are Complex Systems ?

1.1.1 Multi-scale Hierarchy of the World

Let us illustrate the world around us according to the micro-macro hierarchy in a schematic way (Fig.1.1). Each level is distinctively different from others, and follows their proper rules. It does not mean, however, that each level is independent. The upper level consists of lower ones. For example, molecules are combinations of atoms, individual organisms are made of cells, and so on.

Such hierarchy being so apparent, usually sciences have been taking the reductionist approach to understand the system. Reductionism is based on the thesis that to understand the lower level would naturally lead to the understanding of the upper one.

Though, in complex systems such as living organisms, society, etc., it is the way of organization itself that is essential, which severely limits the effectiveness of the reductionism. This does not mean simply that the system is too complicated to understand, but the reductionism is not sufficient methodologically to understand them. Typically, the "living" systems can not be fully understood by decomposing them into parts. The aim of complex systems sciences is to tackle these areas that usual reductionism fails to investigate (Fig.1.2).

1.1.2 Hierarchy in Sciences

The word "science" in english and french is derived from a latin word "scientia", which in a broad sense means knowledge. "Science" originally stood for the entire scholarship, deeply connected to what we call philosophy today, and investigated a wide range of problems in a comprehensive manner, such as the relation between human and society. From the 19th century, it became segmentalized into various professions and established their own independency. From this time on appears the usage of plural form "sciences".

In the early 19th century, french philosopher Auguste Comte represented the hierarchy among sciences as follows: Beginning from mathematics being most abstract and independent from others, follow the astronomy, physics, chemistry, biology, then finally sociology as "the queen of sciences",

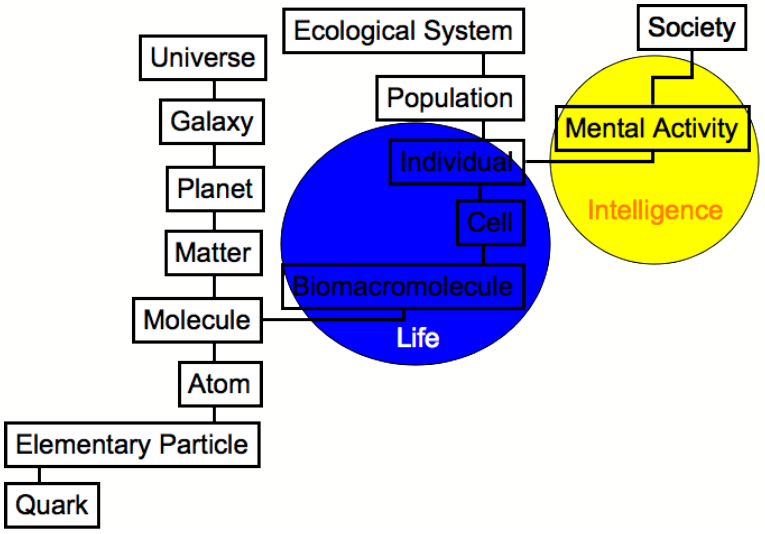


Figure 1.1: **An illustrated hierarchy of the world.** Based on [4] .

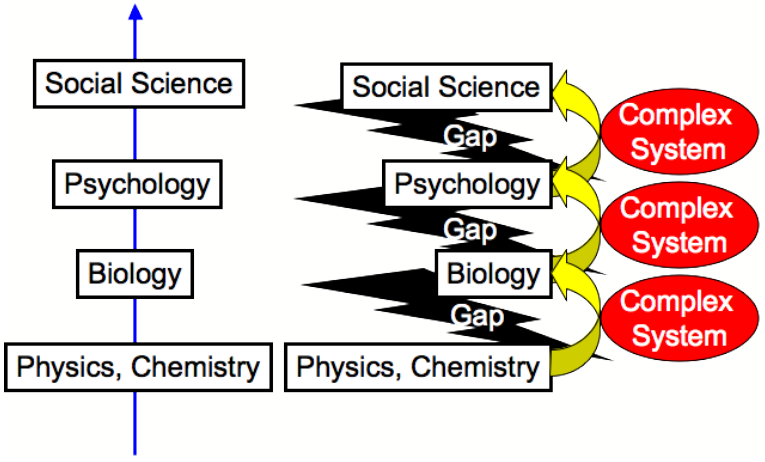


Figure 1.2: **Reductionism approach (Left) and complex systems approach (Right).** Based on [4] .

in this order augmenting the complexity (Fig. 1.3 Left). In this case, each level is supported by the rest of lower levels.

Based on this ordering, british management scientist Peter Checkland established the hierarchy of today's major experimental sciences (Fig. 1.3 Right). In this alignment, Checkland insists that there exist certain gaps between the neighboring levels of sciences. Although physical laws are supposed to explain chemical reactions, which would then explain biological process, in reality there exists too much complexity to connect different levels coherently.

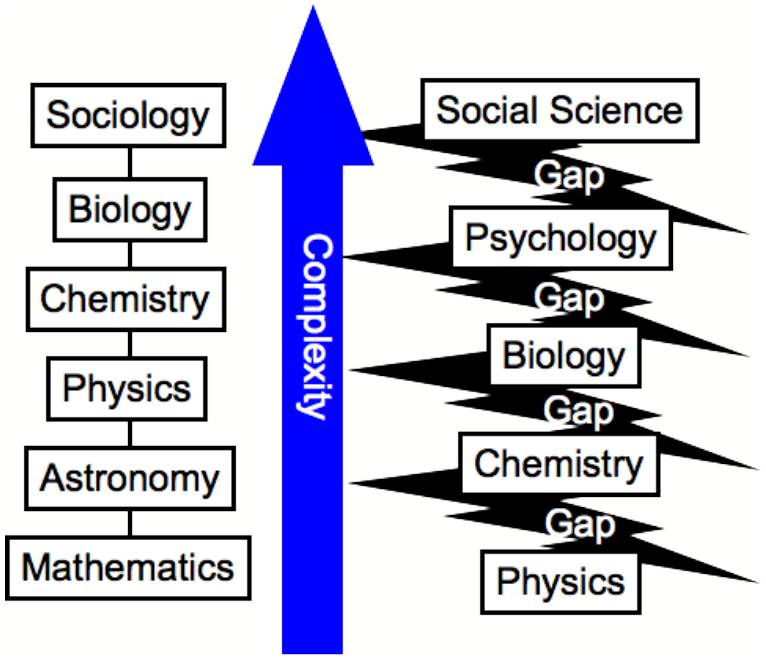


Figure 1.3: **Hierarchy of sciences by Auguste Comte (Left) and Peter Checkland (Right).** Based on [4] .

These hierarchies of scientific knowledge have its root in the principal language of science, logical and axiomatic reasoning based on the symbolization of concepts. As is positioned to the most basic line of science in both Compte's and Checkland's hierarchy, mathematics is the most refined and fundamental form of scientific language. Although the separation between different levels of sciences becomes unavoidable, the symbolization of concepts and investigation on logical structure between them is what established modern mathematics in algebraic aspect, and simultaneously affected philosophical thought. Jules Vuillemin analyzed and conditioned the transition of formal logic into mathematics, and characterized the kind and dynamics of cognition underlying this process [5]. In this perspective, the pure reason of Kant, for example, is a matrix of both formal logic creating the gap between different levels of sciences, and the intuitive cognition of the logical gap between them. The phenomenology of Edmund Husserl, which escape from naive objectivism to transcendental subjectivism, also includes ambiguous but constructive relation with mathematics, expanding our faculty of cognition involved in philosophical conceptualization. If complex systems sciences aim to go beyond the formal limit of logic that has created the existing hierarchy of sciences, we have to proceed into phenomenological perspective to set up an effective language with logical form describing ignored gaps.

1.1.3 Complex Systems Paradigm

What is the stand point of complex systems sciences with respect to the other area of sciences ? One important perspective is that usually sciences have been investigating the substance of a phenomenon, while complex systems deal with their synthesis. Therefore, complex systems sciences are something which bridge the gaps between preexisting sciences, such as physics, biology, psychology, economics, and so forth.

Until recently, it seems that we do not necessary have strict common definition on complex systems, but some kind of common perceptions are shared among researchers. We take here as an example the result of questionnaire performed by the Japanese Electronic Industry Development Association:

The complex systems are the nonlinear systems that consist of various kind of, multiple, nonlinear components, among which there exist dynamically altering nonlinear interactions. Each component is under effect of superior global structure, at the same time global orders emerge from the interaction of the components. (Translation by M. Funabashi)

They also classified the complex systems as follows:

1. Natural systems which consist of simple components. *e.g.* Fluid, meteorological phenomenon.
2. Natural systems which consist of complex components. *e.g.* Brain, immune system, ecological system.
3. Artificial systems which consist of simple components. *e.g.* Plant, computer network, distributed robotic system.
4. Artificial systems which consist of complex components. *e.g.* Economical system, transportation system.

As we consider the history of complex systems science, we should review the work of W. Weaver who brought intellectual insight in this domain [6]. In his pioneering work of 1948, Weaver classified the history of natural sciences after 17th century into three categories:

1. Problems of Simplicity, investigated from 17th to 19th century, which seek for the deterministic laws among a small number of variables.
2. Problems of Disorganized Complexity, generated around 1900, which seek for the averaged statistical laws of systems with a huge number of variables.
3. Problems of Organized Complexity, problems to be tackled seriously hereafter, which corresponds to the today's complex systems science.

He pointed out living systems, economical systems, and social systems as examples of 3., and predicted that the electronic computer drastically developed during the world war II would play an essential role for the research of these systems.

The modeling paradigm is also deeply associated with the genesis of complex systems science [7]. There exists a classification such that:

1. Newton paradigm (17th - 19th century): Both physical laws and the solutions are expressed with mathematical formula.
2. Poincaré paradigm (early 20th century): Geometrical and qualitative analysis. For example, problems such as "Is this system stable ?", "Does the dynamics asymptotically converge to periodic state ?", "How the structure of the solution change when the parameters are modified ?", were explored without explicitly showing the solution in mathematical formula. Only formula of the physical laws are given.
3. Algorithmic modeling paradigm (present): Both physical laws and the solutions are approximately calculated with the use of computer.

The 3. Algorithmic modeling paradigm has become first realizable with the progress of computer technology, and is revealed to be fruitful to simulate the complex systems such as blast furnace and biological system, which is difficult to establish the model from the first principles.

As previously mentioned, the definition of complex systems is not necessary clear, but in a large sense it is considered as "large-scale nonlinear systems". We can further point out some important features as follows:

- Systems that are difficult to apply reductionism (applicable to the Problems of Simplicity and Complicated Systems) and statistical averaging (applicable to the Problems of Disorganized Complexity) for the understanding.

- Systems which themselves are the best simulators. We never know what would happen unless we actually operate the system in real world. In this sense, such systems are unpredictable, difficult to compress information for control, and possibly undecidable.

We will further overview the history of statistical and algorithmic complexity in Ch. 12.

1.1.4 Emergent Property of a System

The concept of "emergence" is also essential in complex systems science. The emergence is generally defined as the generation of global dynamics and functions that are self-organized from non-linearity, interactions, and feedbacks between the whole and the components, which can not be understood by the simple combination of the components dynamics. In wider sense, it can be considered as the spontaneous generation of non-trivial properties that are not incorporated *a priori* in the system.

The word "emergence" has mainly been derived from two disciplines: Systems methodology and artificial life. Let us briefly introduce their meaning in each context.

In systems methodology, we investigate the phenomena according to the levels of description, as shown in Fig.1.1. There exists specific characteristics in each level, such as atom, matter, society, etc., which is described as "emergent property". Usually, the laws which describe certain level seem to contain more information than the ones in lower levels. This fact is the very core of the concept of emergence in systems methodology [8]. The emergence occurs when the characteristic property of a given level can not be reduced in the nature of lower levels. Take an apple, for example, which is a physical form with biological function. Such global characteristics lose its meaning if we look the apple in atomic level. Therefore, the property of the apple is something emerged in the level of matter beyond the atomic level. If we follow this definition, the emergence exist in every level of description, regardless of its complexity.

The definition of emergence in systems methodology belongs to an bottom-up way of interpretation. In artificial intelligence, on the other hand, the word emergence also contain the top-down feedback Fig.1.4. Here, the emergence is the mechanism in which plural components produce global order by interacting locally with each others, at the same time global dynamics affect the property of each component. This scheme is named as "collectionism" by Christopher Langton. This mechanism can be easily understood in social life. Taking generally the dynamics of culture, people behave only with limited information and communication, which forms a global entity called culture, beyond individual's intention. At the same time, each person is in cultural context, affected by the impersonal and global culture, which eventually reproduce novel culture along the bi-directional dynamics.

The concept of emergence in complex systems sciences include these two definitions. In complex systems, rules dominating the components change according to the context of entire system, which leads to the change of the global context itself. This corresponds to the concept of emergence in artificial life. To investigate the relation between the global rules and its components is one of the major tasks in investigating complex systems.

We should also seek for the origin of such systems. How does the global property emerge from lower levels ? How the coupling of bottom-up and top-down feedback is generated ? In other words, how the complex systems came into existence ? These correspond to the concept of emergence in systems methodology.

1.1.5 The Limit of Linear Approximation in Complex Systems

Behind the emergence of complex systems science, there exists an expectation to discover the universality in a large variety of natural phenomena by modeling them as high-dimensional nonlinear systems.

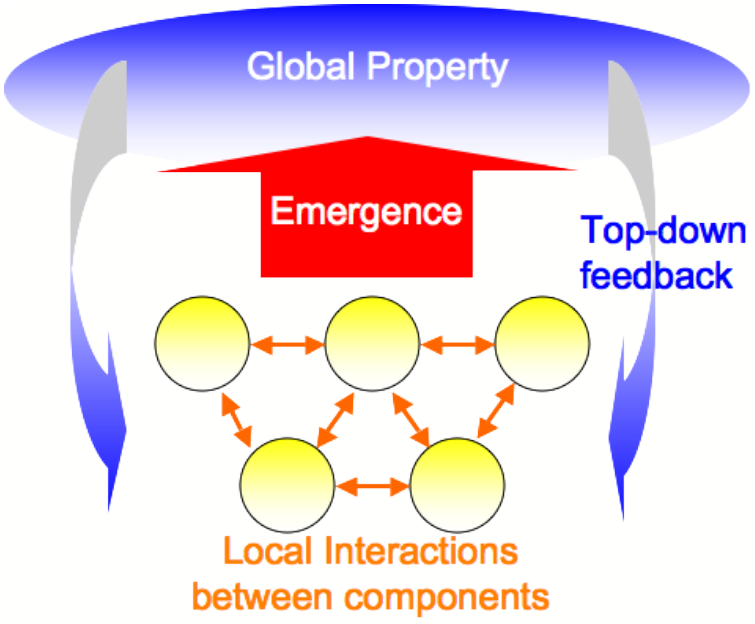


Figure 1.4: **The concept of emergence in artificial life.** Based on [4] .

During the 20th century, natural sciences have been successful in pursuit of the reductionism principle. The reductionism is based on decomposing the target system into the smaller components which are easier to analyze, and trying to understand the whole property of the system by assembling the discovered characteristics of the components. This strategy is supported by the conviction of the modern science that the nature is ultimately dominated by simple laws. To discover the simple law lying beneath the complex phenomena was one of the goals in modern science. The reductionism is also associated with the linear system theory, which is a strong theoretical tool of 20th century sciences. Technically, the superposition principle in linear system confirmed this belief, which assure that the superposition of solutions is also a solution.

However, generally in non-linear systems, the superposition principle does not hold. Therefore, it is basically impossible to grasp the global property of non-linear system with the reductionism strategy, which ignore the effect of non-linear interaction. Furthermore, recent development of technology also raised the problem of the system control based on the reductionism. There are cases that the unintended interaction in the system can produce uncontrollable disturbance, even if we are able to construct basic units and parts properly. We are now confronted to the necessity of science and technology which can handle complex systems for a given purpose. The complex systems sciences show us the limit of analytic approach based on the linear system theory and the reductionism, and bring a great influence on today's science and technology.

1.1.6 High-Dimensional Dynamical Systems and Modeling of Complex Systems

Several types of mathematical model have been proposed for the modeling of complex systems [9] [10]:

1. C.A.: Cellular Automaton or Cellular Automata
2. CML: Coupled Map Lattice
3. GCM: Globally Coupled Map
4. CNN: Chaotic Neural Network
5. ODE: Ordinary Differential Equation

Table 1.1: **Classification of models in complex systems science.** The D,C,L,G, and A stand for Discrete, Continuous, Local, Global, and Arbitrary, respectively.

Model	Time	State	Space	Interaction
CA	D	D	D	L
CML	D	C	D	L
GCM	D	C	D	G
CNN	D	C	D	A
ODE	C	C	D	A
DDE	C	C	D	A
PDE	C	C	C	L

6. DDE: Difference-Differential Equation or Delay-Differential Equation

7. PDE: Partial Differential Equation

In general cases, complex systems consist of a large number of components. The modeling naturally require a corresponding number of variables. Therefore, it is important to consider how to describe the complex systems as high-dimensional dynamical systems.

Mathematical models of complex systems can be classified loosely with the following properties:

- Whether temporal development that determines the state in next step is discrete or continuous: Discrete-time dynamical system or continuous-time dynamical system.
- Whether the variables that define the state of the system is discrete or continuous: Integer variables or real variables.
- Whether the components spatially exist as discrete units or continuous (as a fluid).
- Whether the interactions between the components are local or global.

The Tab. 1.3 shows the classification of the previously mentioned models with these criterions.

Recent study on chaotic systems has achieved not only the deep understanding of low-dimensional chaos, but even realized engineering application. On the other hand, the rapid development of computation technology enables us to perform large scale simulations, which greatly promote the analysis of the dynamics of high-dimensional chaotic systems. Such area of research concerning the high-dimensional chaos is situated at the core of complex systems science.

The understanding of complex systems in the real world is expected to realize a new paradigm of science and technology of 21st century by integrating previously opposing concepts in the history of science, such as determinism *vs.* stochasticity, chance *vs.* necessity, order *vs.* chaos, whole *vs.* part, universality *vs.* specificity, subjectivity *vs.* objectivity, and mechanistic theory *vs.* teleology.

1.2 Internal Observation

1.2.1 Introduction of Internal Measurement

The observation of complex systems requires a more refined understanding of "measurement". Above the fact that the system is huge and complex, we have to reconsider why we can determine the "state" of the system. Generally, the word "measurement" stands for the observation of an object by a person. Though, if we consider the hierarchy of emergence in this world, it is impossible to hold a complete understanding of an object in arbitrary resolution.

In internal measurement theory, we decline in principle the possibility to observe the system with the accuracy of Maxwell's Demon, but define the measurement as the interaction between object and observer. In this framework, the object and the observer can not be separated independently, and the observer always acts as an agent, in unceasing interaction with the object. The internal measurement is expected to play a key role to interpret the process of emergence in complex systems.

We introduce here the discussion by Kouichiro Matsuno, based on material science [11] [4]. Another stream of internal measurement is developed by Pegio-Yukio Gunji, taking after the discussion of Wittgenstein and Kripke, which will be introduced in Part 2.

1.2.2 Reconsidering Measurement

How can we know the "present" state of a distant object ? To see something means that there exist some photons reflected from the object to reach our eyes. Even though the speed of light is as high as about 300 km/s, in principle there exists temporal delay in any observation. This is also the principle of relativity theory.

The temporal delay in measurement is more apparent when we look up the stars in the night. What we see as the stars do not represent their present state. Even the closest fixed star from our solar system is as far as 4.3 light years, which shows us the state of 4.3 year before. Augmenting the resolution of measurement does not change this delay, unless we develop a way to measure faster than with the light.

Such problems have not been discussed in central issue of physics, since the delay of measurement appears to be significant only in astronomic scale. When we observe a system in an ordinary scale, we usually consider its entire state. Especially in the modeling, we are based on an assumption that we can determine well the state of the system (except for uncertainty principles under theoretical assumption such as in quantum level). The speed of light can be approximated as infinity. In this sense, present physics can be called as *state-oriented physics*.

On the other hand, in complex systems such as the chemical reaction inside of a cell, the propagation of energy is mediated by macrobiomolecules, which significantly limit the speed of interaction. Here, it becomes important to consider the interacting process including its time delay.

1.2.3 Local Interaction of Components and Internal Measurement

Matsuno formalized the energy conservation law as an internal measurement process as follows. Take for example a many-particle system, and simulate the relaxation process after certain amount of energy was added externally (Fig. 1.5). We consider the case only a part of the particles receive the external energy. For an experimenter observing the system from the outside, the relaxation process proceeds toward equilibrium where the total amount of energy remain unchanged. The experimenter observed the two state of the system in equilibrium, before and after adding the energy, as an *external observer*.

On the other hand, if we consider the transition of energy between each particle as *internal observer*, the process of measurement becomes qualitatively different. Particles only react with their neighbors, and never obtain global information. In other words, there exists no signal to let the particles know the definitive state of equilibrium, and energy transition occurs only through the local interaction. The particles therefore change their own state along the measurement or interaction with others, and as a result of collective internal measurement, realize a global equilibrium state without knowing the total amount of energy.

The particles in this example are interacting, at the same time identifying the present state of others to compare the amount of energy. In this sense, identification or recognition precedes the interaction. Matsuno then defined the internal measurement as such identification/recognition process inevitably associating with the interaction [12]. Internal measurement therefore means the act of a system's components recognizing other ones through interaction, and is in opposition to the external measurement where hypothetical observer detects all the details of the system instantaneously.

External measurement is based on the assumption that the observer and the object can be separated and does not affect with each other. In internal measurement, however, it is in principle impossible to separate or ignore the interaction between them. Since internal measurement is defined based on the interaction, internal observer is in constant change affected by the measurement. Internal measurement theory places particular emphasis on this framework, and tries to investigate complex systems from the inside.

Such framework of thinking becomes non-trivial when investigating complex systems where the propagation of interaction is significantly constrained with respect to the system size. How does each component realize a global state with complex structure, by means of only local interaction, without knowing the information of other distant components ? Naive subjective observation on emergent phenomena in living systems gives us the impression that as if some global context determines the behavior of each component, and historically provided the long-held trend of teleology in organicism.

We will further develop and apply internal measurement in Part 2 in order to interpret the emergent attractors observed in chaotic neural network.

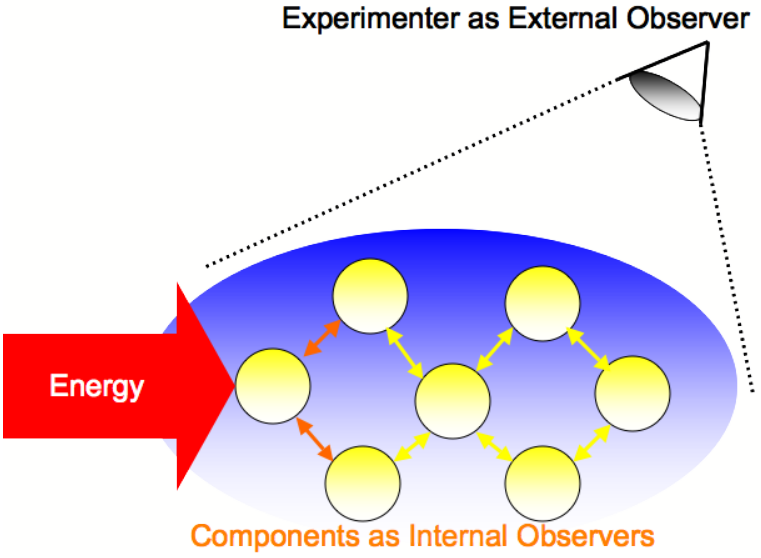


Figure 1.5: **Schematic representation of internal measurement and external measurement in many-particle system.** Internal observers interact only locally and propagate energy with time delay. In the beginning, other end of the system does not even know whether the energy was added or not. External observer measures the resulting equilibrium where total amount of energy is observed. Based on [4] .

1.2.4 Mathematical Closure of Open System and Internal Measurement

Physics have been succeeded by using the mathematics as the principle language to describe natural phenomena. The logical perfection and empirical symbolization are the strong property to establish scientific thinking, but requires to be extremely rigorous and precise to observe the state of a system. For this reason, mathematical modeling tends to require external measurement due to both its form of description and to ensure non-tribial logical autonomy.

Still, some principles in modern physics can be found accessible to the internal measurement. Even though the eventual model need to form a mathematically coherent closure, local property of the model tolerates certain uncertainty in the way of measurement under invariance property of the model. For examples, In information theory, the topological entropy does not depend on the way to make partition of the state space. The Fisher information matrix which defines the information theoretical metric in information geometry does not depend on the way of observation as long as the variables are sufficient statistics. The Gauge invariance in particle physics also stands on the same principle [13]. With these invariance properties, one does not need to have the global coordinates of observation to identify the system on the model. The data from internal measurement with sufficient degrees of freedom with respect to the whole system may find its way to be expressed in a state-oriented model.

Dynamical system is another example of mathematical closure describing open system. Although the overall kinds of components need to be symbolized with interactions, the quantitative

Table 1.2: **Complex systems and their major emergent properties investigated in this thesis.**

Part	Object	Emergent property
Part 2	Chaotic neural network (CNN)	Chaotic Itinerancy Synthesis of novel attractors
Part 3	Collective robots with CNN	Optimal intermittent search
Part 4	Social network	Order-wise correlations dynamics Dynamics of complexity measures
Part 5	Embryogenesis	Morphogenetic fields
Part 6	Vowel system	Formants distribution
	Multilingual environment	Multilingual community patterns

change of energy and material flows including external environment is possible to simulate as a non-equilibrium open system. The modeling of dissipative structure in non-equilibrium thermodynamics is a historical landmark of this possibility, which is also extended to model dynamical brain activity on the basis of neural network.

1.3 Complex Systems Treated in this Thesis

We list the complex systems and their major emergent property which will be investigated in this thesis in Tab. 1.3. The emergent property will be investigated with the use of complementary analysis between the constructive and interaction-analytical methodologies explained in the following section.

1.4 Methodologies toward the Understanding of Complex Systems

1.4.1 Constructive Methodology

Analysis of complex systems requires not only the identification of the components, but the investigation on the interactions between them. The constructive method is an engineering approach to model this coupling between the components and the interactions in its entity. The methodology is based on the construction of a virtual model in computer, and try to simulate the phenomenon of interest. We try to extract the essential property of the system by ameliorating the model. For example, models such as multi-agents systems simulating economical exchange and coupled chaotic oscillators to observe emergent global dynamics are part of it. We will apply constructive method in Part 2 and 6.

"Simulation" and "analogy" are two important tools in constructive method. Simulation is the construction and heuristic operation of a model which encapsulate the components and their relations of the target system. In most case, simulation is performed by computing *in silico* derived mathematical model.

The objectives of simulation can be classified into two reasons: One is the case that we know rather well about the structure of the system, but are difficult to perform experiment for reason of safety and cost. Examples are crash experiment of cars and flight simulator. The second one is the case that we do not know about the internal structure of the system, and try to understand that by simulating the components with interactions. Simulation in complex systems sciences corresponds to the latter case.

How is simulation different from experiment ? Experiment is performed in real world and is a way to verify a theory. It inevitably includes uncertainty factors out of theoretical consideration. On the other hand, simulation only shows the dynamics of the model following strictly the defined logical operation. We can simulate the pure dynamics of hypothetical theory in itself with idealized condition. This is a strong point of simulation when investigating complex interactions in the system. Recent physics also call for the simulation as the third methodology besides the

traditional dialectic between theory and experiment. Computational physics try to investigate complex phenomena by simulating them from a series of fundamental equations of basic laws.

Modeling precedes simulation, as hypothesis precedes experiment. The way of modeling generally varies, such as simplified, constructive, replicate, homologous ones. In complex systems sciences, the constructive modeling is especially important. It is used to model a system that we do not know its internal structure in detail, by putting certain assumption. Constructive model is therefore hypothetical, and does not guarantee a complete understanding of the object. It is necessary to evaluate its plausibility through simulation. This process is called as the "identification" of the model, and can be described as the following procedures:

1. Observation of the System

   Observe the system in detail and try to investigate the internal structure as clear as possible.

2. Extraction

   Based on the observation, extract the property which are supposed to be essential for the reproduction of the phenomenon. In general, there exist plural possibilities in the way to extract from the same observation.

3. Constructive Modeling

   Create a constructive model based on the extracted property. The model generally combines not only the revealed fact but hypothetical notions. We should compensate the missing information with provisional assumption.

4. Simulation

   Compute the model dynamics with necessary range of parameters and initial conditions, and register the results.

5. Investigation of Simulation Result

   Visualize and investigate the results. Focus on important characteristics and replan the simulation to better characterize them. Reconsider the hypothesis and model structure.

6. Evaluation with the Comparison between the Model and the System

   Compare between the dynamics of the model and that of the system in real world, and evaluate the plausibility of the model. Taking analogy between the simulation and real world is an important methodology. Though we do not have yet an objective measure for the evaluation.

Simulation is a method to understand the object with the use of hypothetical constructive model. Simulation forms a closure of virtual world, where the model and its computed dynamics are logically coherent on the basis of programming language. However, simulation does not directly tell us about the plausibility of the model. The model seems to be "correct" if the dynamics is similar to that of real world, though there exist no objective criteria to judge its authenticity. We can only notice the similarity, which does not necessary mean to elucidate the mechanism of the object. We should call for a conceptual methodology for the evaluation of the model. In this sense, considering the analogy between the model and the object becomes important.

Constructive modeling requires analogy for the evaluation. Analogy is the comparison between two different systems, which evaluate the similarity and dissimilarity between them. The correspondence between economical and ecological systems is an example. M. Hesse pointed out that there exist three kinds of analogy in scientific models based on the similarity/dissimilarity relation:

1. Positive analogies. The features which are known or thought to be shared by both systems. For example, similarity of the mechanism between the competition of companies and the natural selection of species can be considered as positive analogy.

2. Negative analogies. The features which are known or thought to be present in one system but absent in the other. The fact that the components of business company are humans with thinking ability and that the components of living organisms are the cells without such mental capacity is an example of negative analogy.

3. Neutral analogies. The features whose status as positive or negative analogies is uncertain at present. Neutral analogies will be further classified into positive or negative ones through further investigation. The change of business companies and biological evolution does not hold clear correspondence in its mechanism, therefore belongs to neutral analogy. Neutral analogies are by far the most interesting of the three types of analogies, for they suggest ways to test the limits of our models, guiding the way for scientific advancement.

Usually sciences emphasize the positive analogies when we describe a theory or law. In constructive modeling, we also include neutral analogies in the model. It is by incorporating the neutral analogies into the model and investigating by simulation that we can further classify them into positive/negative analogies. This is the very understanding process of constructive modeling.

Taking analogy also raises the problem of the appropriate level of comparison. It is important to consider in which features of simulation we should discuss analogy. In the following we quote an interesting dialogue concerning the interpretation of a hurricane simulation [14]:

> Sandy: *I mean that to see the winds and the wetness of the hurricane, you have to be able to look at it in the proper way.*
>
> Chris: *No, no, no! A simulated hurricane isn't wet ! No matter how much it might seem wet to simulated people, it won't ever be "genuinely" wet! And no computer will ever get torn apart in the process of simulating winds!*
>
> Sandy: *Certainly not, but you're confusing levels. The laws of physics don't get torn apart by real hurricanes either. In the case of the simulated hurricane, if you go peering at the computer's memory expecting to find broken wires and so forth, you'll be disappointed. But look at the proper level. Look into the "structures" that are coded for in the memory. You'll see that some abstract links have been broken, some values of variables radically changed, and so forth. There's your flood, your devastation - real, only a little concealed, a little hard to detect.*

In this conversation, Chris sees the simulation as a part of natural phenomenon in real world, and cares if the simulation is "wet" as the actual hurricane is. Chris is trying to take analogy between the behavior of computer processor and the hurricane all in the same actual world. While Sandy conceptually distinguishes between what is happening in the virtual world of computer and the phenomena of real world. Sandy is taking the analogy between the representation of the model inside of computer and the hurricane of actual world.

We should always pay attention on which levels of the object and simulation we can find meaning in comparison. With this condition, we can for the first time appropriately discuss the classification of neutral analogies into positive and negative ones.

Complex systems such as life, intelligence, society, etc., have been investigated in various way for a long time. We do not practically know however, even the certain way to understand them. Complex systems sciences therefore need also to seek for the methodology to understand them.

Constructive modeling in Complex systems sciences does not necessary require to produce a copy of the real world. We are investigating the mechanism of the object, and not the object itself. It is similar to creating a map does not require to realize with the same scale and the real substance. We construct a model to understand the object, and not to create a precise approximation model.

In this thesis, we mainly use dynamical systems or stochastic dynamical systems for the constructive methodology.

1.4.2 Interaction-Analytical Methodology

The difficulty of the constructive methodology is situated not only in the complexity of the components, but the interactions. Traditional physics has been principally succeeded in the system with small number of variables, or in the large system with simple interactions such as linear or random relation. The effectiveness of the reductionism reduces quickly as the system size augments with complex effective interactions. Let us consider this fact in terms of the degrees of freedom in a stochastic system.

We consider a system with n stochastic variables allowing all orders of many-body correlation. We take the measurement-oriented theory viewpoint and suppose that the propagation of the interaction is limited to a certain speed v. Then the relative propagation scale of the effect of a component is given by v/n as depicted in the top figure of Figs. 1.6, which represents the domination degree of a component over the whole system. The value of v/n qualitatively reflects the importance of elementalism on the system and the effectiveness of the reductionism with respect to the system size n.

On the other hand, the possible many-body correlations exist at the order of combinatorics as $\sum_{k=1}^{n} {}_nC_k$. By renormalizing it with the system size n as $\sum_{k=1}^{n} {}_nC_k/n$, this quantity represents the domination degree of the interactions over a component, and qualitatively reflects the importance of relationalism. As depicted in the bottom figure of Figs. 1.6, the combinatorial explosion occurs to the renormalized number of interactions. The degrees of freedom necessary to identify the model is therefore almost the parameters of interactions which augments in factorial order. For example, at $n = 7$ the degrees of freedom of each component's marginal distribution become less than 1 % of those of joint distributions. Even if we define the interaction propagation speed as infinity, the effectiveness of the reductionism inevitably reduces with the presence of these interactions. The role of relationalism becomes dominant.

It is therefore quite important to consider how to analyze the interactions in complex systems. The constructive methodology should put relatively greater emphasis on the analysis of interactions than the simple elements. Furthermore, the hierarchical multi-body interactions also includes the relations of the relations between the subsystems. To encapsulate the variation and complexity of the interactions, we consider the following "interaction-analytical methodology" as a complementary strategy of the constructive methodology.

In this thesis, we mainly use the information geometry as a theoretical foundation of the interaction-analytical methodology.

We first consider the dual expression between the variables and the interactions of a model. Schematic representation is depicted in Fig. 1.7. The interactions are usually described as a function of the variables. For example, the interactions between a set of the variables $\{x_i | 1 \leq i \leq n\}$ can be formulated as $\{f_j\{x_i\} | 1 \leq j \leq \sum_{k=2}^{n} {}_nC_k\}$. The network representation usually takes the variables as nodes, and the interactions as edges between nodes. We call this orthodox representation as the "variable network". However, we can also describe the model with a network whose nodes represent the interactions and the edges correspond to the variables. We call this the "dual expression of variable network". Here, we can formulate the variables as a function of the interactions such as $\{x_i(\{f_j\})\}$.

The dual expression is itself a trivial translation of the variable network, but becomes the start point to investigate the interactions. We call the symbolized functions $\{f_j\{x_i\} | 1 \leq j \leq \sum_{k=2}^{n} {}_nC_k\}$ as *functor*, taking after the same word representing "function object" in C++ language. By materializing the interactions as functors, we can for the first time treat the interactions between the interactions, which are formalized as the interactions between functors as $\{f'_i(\{f_j\})\}$. Since the $\{f'_i\}$ are the functors of functors, we call them as "*meta-functors*". We then define the "functor network" whose nodes consist of the functors $\{f_j\}$ and edges are the meta-functors $\{f'_i\}$. This representation gives a non-tribial information on the way emergent property of a system is organized via interactions.

We can further construct a meta-level expression as meta-functor network, meta-meta-functor network, etc., according to the necessity of investigation. We describe these meta-level networks with a general expression "meta^l-functor network", where $l \geq 1$ represents the hierarchical superscript of meta-level.

Note that the nodes number of meta^l-functor network augment in factorial order with respect to meta-order l if the system size is $n \geq 4$.

This strategy aims to contribute to discover the universality between different scales of emergent phenomena. Fig. 1.8 represents the expected universality we aim to discover with the use of the meta^l-functor network model. Suppose we have a series of emergent phenomena in a nested structure ranging from micro to macro scale. For example, human society consists of individuals who has brain, which consists of neurons. The ways human organize society and neurons express brain activity are typical examples of emergent phenomena in complex systems. If we are satisfied to independently model these two different levels of emergence and do not discuss on the commonality

between them, we loose a big picture of the universal theory in complex systems. Complex systems sciences try to deal the way of organization, or the way the interactions contribute to emergent phenomena. The universal theory to understand the emergence should be situated in the highest level of abstraction which does not primarily depend on the specific property of components. If there exists transversal universality of interactions in different levels of organization, such physical law is a strong candidate for the general theory of emergence.

The analysis and comparison of the meta^l-functor network is a concrete strategy toward this possibility. Simple modeling with variable network can reveal the organization of complex interactions to some extent, but is not accessible to the comparison of different systems with different scales. Functor networks, on the other hand, deal with the interactions of the interactions, which do not explicitly depend on the property of the components. If there exists some universal structure between the functor networks of different scale systems, such feature is a candidate for the general theory of emergence. If the functor network still depends too much on the components particularity of the systems and is not sufficient to abstract the organization of interactions, we can augment the level of relationalistic abstraction $l \geq 1$ to obtain the sufficient meta^l-functor network, until we can judge the existence of a transversal universality.

1.4.3 Dialectic Between Constructive and Interaction-Analytical Methodologies

The constructive methodology and the proposed interaction-analytical methodology are complementary with each other toward the understanding of complex systems. Constructive methodology usually utilizes dynamical or stochastic systems for mathematical modeling. On the other hand, a strong theoretical framework is needed for the interaction-analytical methodology, which should be also compatible to constructive modeling.

For a stochastic system, analysis of interaction can be performed with the use of a statistical model. In a completely deterministic dynamical system, chaotic dynamics is situated at one of the cores of complex systems sciences. Since chaotic dynamics both has the deterministic and stochastic aspects, stochastic model is one of the important approach to analyze the interactions in dynamical system with chaos.

In this thesis, we use the information geometry on stochastic models describing the interactions of both dynamical and stochastic systems as a concrete strategy of the interaction-analytical modeling. The dialectic between the two methodologies are based on the complementarity between the dynamical system and information geometry in actual analysis.

The strategy of the investigation is schematized in Figs. 1.9. The interaction-analytical strategy can serve both to construct and evaluate a constructive modeling. Usually a constructive model of complex systems, such as chaotic dynamical system and contact process, expresses quite complex dynamics, which itself is difficult to analyze and contains huge theoretical interest. The formalization of the interactions with the use of a stochastic model and the analysis with the information geometry is expected to reveal and characterize the dynamics of the model, which would serve to evaluate and identify the model with respect to the actual phenomenon.

The analysis of interactions with the use of the information geometry is also valid for actual data. If we establish a constructive model based on the analysis of a functor network of actual data, it would be possible to attribute more relationalistic property to the model. This approach is especially fruitful when we do not have a priori information on the components, or the components are themselves extremely complex such as social dynamics. In such cases where the construction of the variable network is difficult, the constructive modeling would still be possible with the use of the functor network extracted directly from stochastic interactions of data.

The concrete complementarity between the analyses with dynamical system and information geometry in this thesis is listed in Tab. 1.3. The vertical axis aligns the part number which corresponds to the different objects in complex systems as listed in Tab. 1.3. The horizontal axis aligns the theoretical concepts related to the dynamical system and information geometry. Note that we used the term “dynamical system” in a large sense including stochastic dynamical systems such as contact process.

The comprehensive objective of this thesis is to apply the proposed complementary strategy to a wide range of complex systems, seeking for a universal characteristics of emergence in a

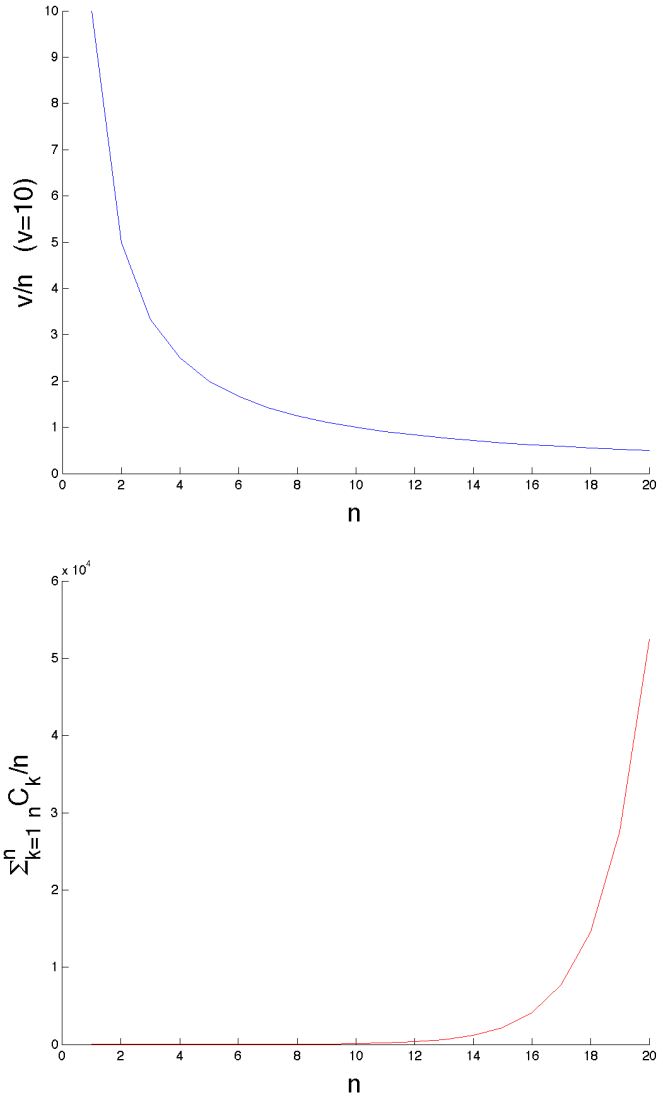


Figure 1.6: **Domination degree of components and interactions.** Top: System size n vs renormalized propagation speed of interaction v/n (blue line) representing the domination degree of a component over the whole system. Bottom: System size n vs renormalized interaction numbers $\sum_{k=1}^{n} {}_nC_k/n$ (red line) representing the domination degree of interactions over a component. The effectiveness of the reductionism/elementalism qualitatively reduces as the role of relationalism becomes dominant in large system.

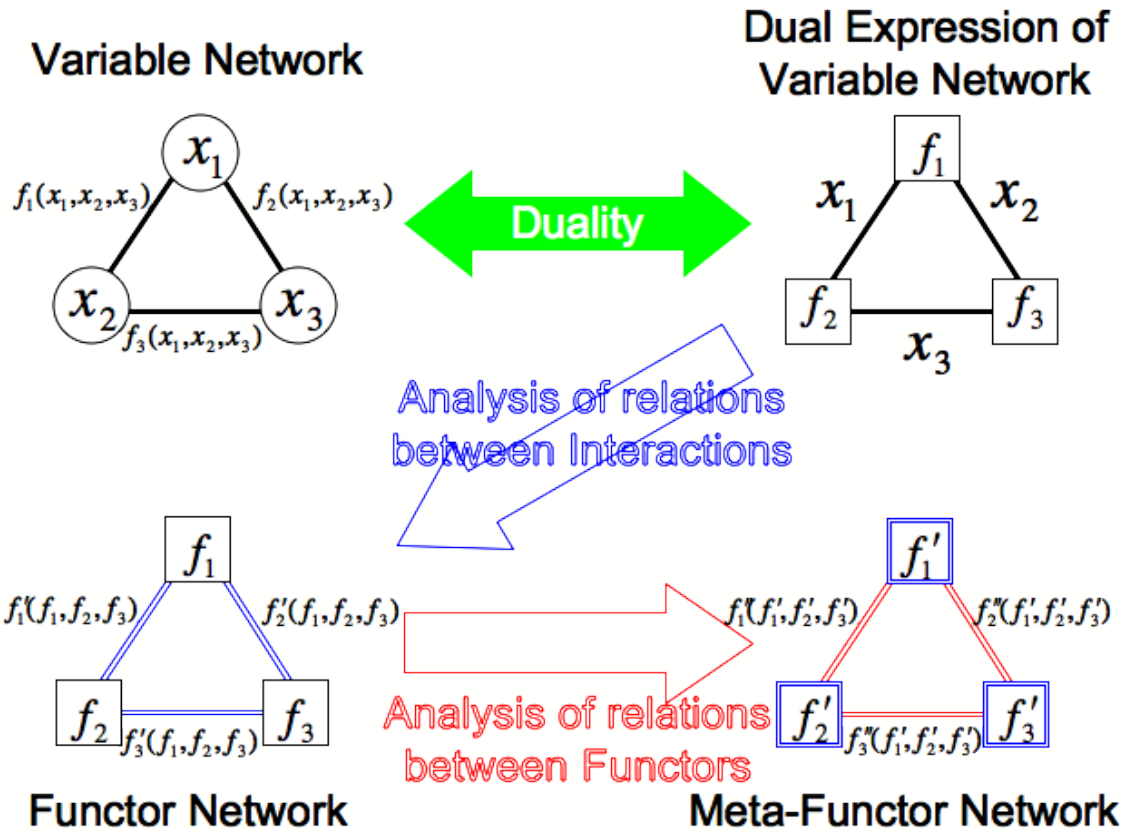


Figure 1.7: **Duality between variables and interactions, and reconstruction of (meta-)functor network.** x_1, x_2, x_3 stand for the variables of the system, and f_1, f_2, f_3 are their functions representing the interactions. We take these function as objects and call them as functors in the dual expression of the usual variable network. This leads to a formalization of functor network, where we consider the interactions f'_1, f'_2, f'_3 between functors. This abstraction will continue to create another meta-functor network until we would reach to an sufficient understanding of the system organization.

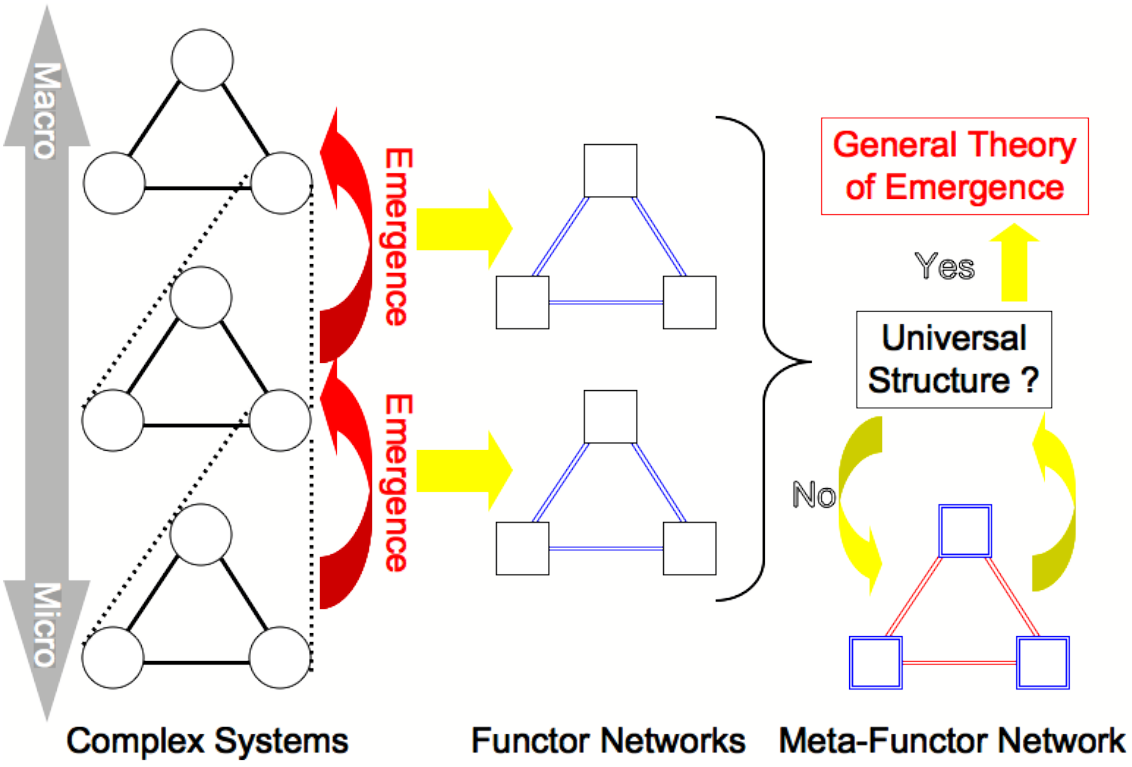


Figure 1.8: **Expected universality in functor networks structures in different levels of emergence.**

relationalistic viewpoint.

1.5 Conclusion

We briefly reviewed the objective of complex systems sciences including historical context. The novel interaction-analytical methodology based on information geometry is proposed, as a complementary approach to the constructive methodology based on dynamical system, as well as an unified strategy to tackle the complex systems treated in this thesis.

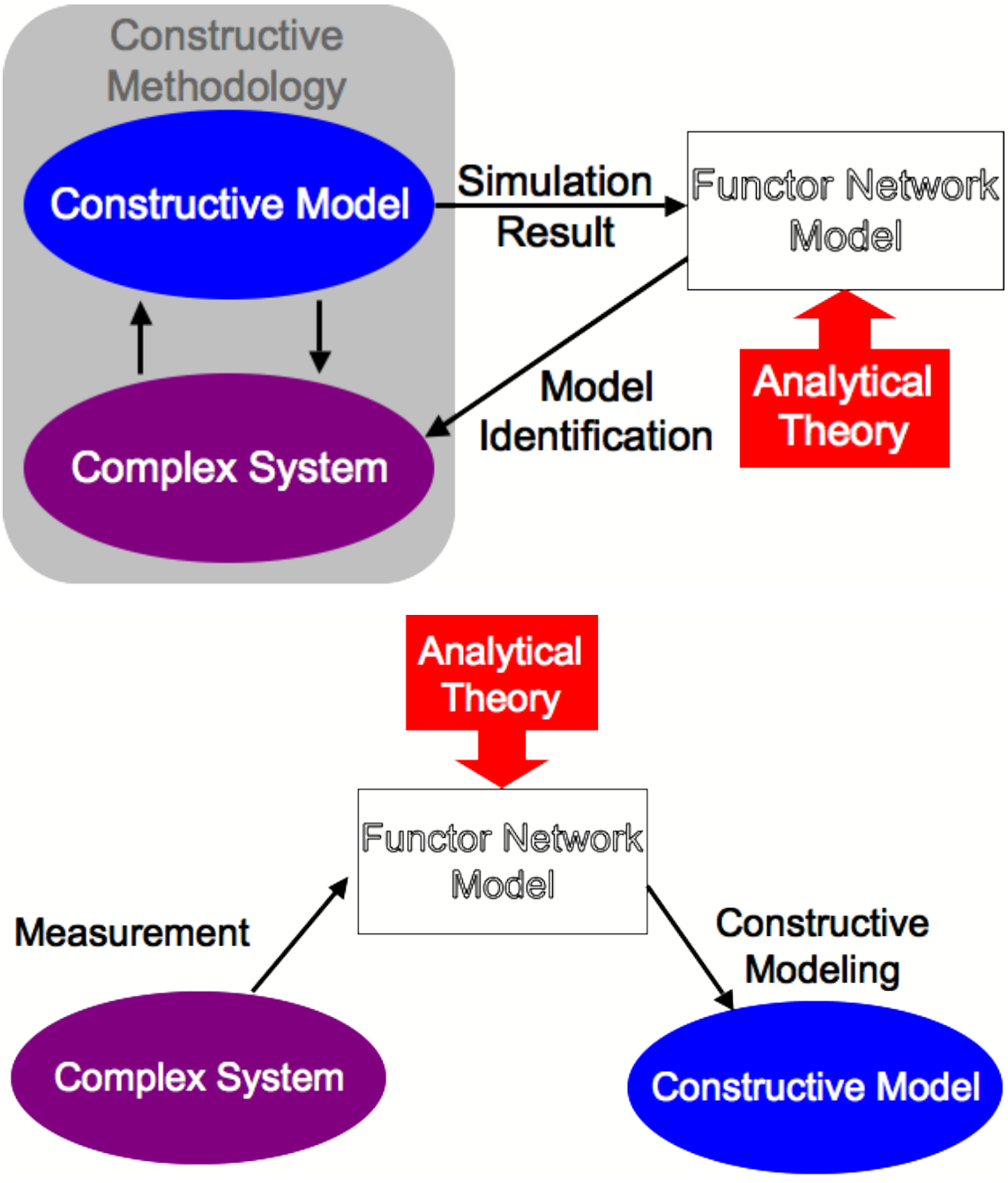


Figure 1.9: **Dialectic between Constructive and Interaction-analytical methodologies.** Top: Identification of constructive model. Bottom: Derivation of constructive model.

Table 1.3: **Theory *VS* Object matrix for the investigation of complex systems in this thesis.**

	Dynamical System	Information Geometry
Part 2	Chaotic itinerancy in CNN Hierarchical stability analysis	System decompositionability
Part 3	Simulation of chaotic roving robot Optimal intermittent search strategy	Multi-scale evaluation of infotaxis
Part 4	Return map analysis Dynamics of system decompositionability and complexities	Order-wise correlation Network decomposition Edge information Complexity measures
Part 5	Cell movement vector field Embryo development in clustering space	Dynamical tissue detection EM-clustering Analysis on morphogenetic entropy
Part 6	Invariance of KL div. distribution Formant localization Non-linear contact process	Decomposition of KL divergence Vowel center Multilingual accumulation effect

Chapter 2

Preliminaries of dynamical system

Abstract

In this chapter, we review the preliminary knowledge of dynamical system used in this thesis based on [1] [15].

Keywords: Discrete-time dynamical system, bifurcation, crisis, chaotic itinerancy, Lyapunov exponent

2.1 Introduction

Dynamical system is a set of variables changing its values along the time line according to an interaction rule. The mathematical definition of dynamical system can be defined both with continuous and discrete representation of time. Typical examples of continuous-time dynamical system are differential equation and vector field. Discrete-time dynamical system is mostly expressed as difference equation or **map**. In this chapter, we introduce basic concepts of discrete-time dynamical system as a preliminary of this thesis.

2.2 Discrete-Time Dynamical System: Map

Definition.

1. We represent a point $\mathbf{x}$ in n-dimensional real value space $\mathbf{R}^n$ with $\mathbf{x} = (x_1, \cdots, x_n)^T$, where the superscript T means the transpose of a vector.

   We define a continuous map $f : \mathbf{R}^n \to \mathbf{R}^n$ as follows:

$$\begin{aligned} f(\mathbf{x}) &= (f_1(\mathbf{x}), \cdots, f_n(\mathbf{x}))^T \\ &= (f_1((x_1, \cdots, x_n)), \cdots, f_n((x_1, \cdots, x_n)))^T. \end{aligned} \tag{2.1}$$

   Then we call the following discrete-time difference equation as **discrete-time dynamical system**:

$$\begin{aligned} \mathbf{x}(t+1) &= f(\mathbf{x}(t)),\ (t = 0, 1, 2, \cdots) \\ &\Leftrightarrow \begin{cases} x_1(t+1) = f_1((x_1, \cdots, x_n)) \\ \cdots \\ x_n(t+1) = f_n((x_1, \cdots, x_n)) \end{cases} \end{aligned} \tag{2.2}$$

   This is nothing but a renaming of $f : \mathbf{R}^n \to \mathbf{R}^n$. In the context of dynamical system, we simply call discrete-time dynamical system as **map**.

2. For a point $\mathbf{x}_0 \in \mathbf{R}^n$, we define the positive semi-orbit of $\mathbf{x}_0$ with a sequence $\{\mathbf{x}_t : t = 0, 1, 2, \cdots\}$ as follows:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t), \ (t = 0, 1, 2, \cdots). \tag{2.3}$$

3. We call f **invertible** if f is homeomorphism(there exists continuous inverse mapping f^{-1}), if not **noninvertible**. In case f is invertible, we can obtain the time development in the negative direction(past). For $\mathbf{x}_0 \in \mathbf{R}^n$, we consider the following sequence $\{\mathbf{x}_t : t = -1, -2, \cdots\}$:

$$\mathbf{x}_{t-1} = f^{-1}(\mathbf{x}_t), \ (t = 0, -1, -2, \cdots). \tag{2.4}$$

Combining with positive semi-orbit $\{\mathbf{x}_t : t = 0, 1, 2, \cdots\}$, we call the sequence $\{\mathbf{x}_t : t = 0, \pm 1, \pm 2, \cdots\}$ as **orbit** of $\mathbf{x}_0$.

For example, the chaotic neural network model in Part 2 is noninvertible.

2.3 Bifurcation of Periodic Point in Discrete-Time Map

We consider the dynamical systems with parameters. If the dynamics of a dynamical system qualitatively changes when the parameters are changed, it is called **bifurcation phenomenon**. In this section, we introduce the example list of bifurcation phenomena in 2-dimensional map.

We consider the bifurcation of periodic point in discrete-time map. Consider the following map with parameters $\mu \in \mathbf{R}^p$:

$$\mathbf{x} \mapsto f(\mathbf{x}, \mu), \ \mathbf{x} \in \mathbf{R}^n. \tag{2.5}$$

Then define that $\mathbf{x} = \mathbf{x}_0$ is a fixed point when $\mu = \mu_0$.

$$f(\mathbf{x}_0, \mu_0) = \mathbf{x}_0. \tag{2.6}$$

Definition.
The linearized map at $\mathbf{x}_0$, consider the following Jacobian A:

$$\mu \mapsto A\mu, \ \mu \in \mathbf{R}^n, \tag{2.7}$$

$$A = D_{\mathbf{x}} f(\mathbf{x}_0, \mu_0) = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0, \mu_0) \right)_{1 \le i,j \le n}. \tag{2.8}$$

If all eigenvalues of A do not exist on unit circle $S = \{\lambda \in \mathbf{C} || \lambda | = 1\}$, the fixed point $\mathbf{x}$ is called **hyperbolic**. If all eigenvalues of A exist inside of the unit circle $\{\lambda \in \mathbf{C} || \lambda | < 1\}$, the fixed point $\mathbf{x}$ is called **stable**. If there exists at least one eigenvalue of A at the exterior of the unit circle $\{\lambda \in \mathbf{C} || \lambda | > 1\}$, the fixed point $\mathbf{x}$ is called **unstable**.

Definition.
Fix the parameter μ and represent the p times composite mapping of $f(\cdot, \mu) : \mathbf{R}^n \mapsto \mathbf{R}^n$ as f^p.

$$\mathbf{x} \mapsto f^p(\mathbf{x}, \mu), \tag{2.9}$$

$$f^p(\cdot, \mu) = f(\cdot, \mu) \circ \cdots \circ f(\cdot, \mu). \tag{2.10}$$

If a point $\mathbf{p}$ is the fixed point of f^p and is not for f^i with arbitrary i $(1 \leq i < p)$, $\mathbf{p}$ is said to be the p-periodic point of f.

$$f^p(\mathbf{p}, \mu) = \mathbf{p}, \tag{2.11}$$
$$f^i(\mathbf{p}, \mu) \neq \mathbf{p} \ (1 \leq i < p). \tag{2.12}$$

The meaning of being **hyperbolic**, **stable**, and **unstable** of the p-periodic point $\mathbf{p}$ corresponds those of the fixed point f^p.

The bifurcation of the periodic point of a map f is equivalent to the bifurcation of the fixed point of f^p. Then it suffices to consider the bifurcation of the fixed point of a general map f. The following theorem gives the sufficient condition that a fixed point of a map do not trigger bifurcation.

Theorem.
If a fixed point $\mathbf{x}_0$ at $\mu = \mu_0$ is hyperbolic, the fixed point does not disappear with respect to the perturbation of parameter μ in the neighborhood of μ_0, and the type of stability does not change.

Therefore, to examine the bifurcation of a map, it suffices to consider a general form that a map hold the non-hyperbolic fixed point $\mathbf{x}_0 = 0$ when parameter $\mu = 0$.

2.4 Bifurcation List of 2-Dimensional Map

We introduce the bifurcation list of 2-dimensional discrete-time map. Consider the following map

$$\mathbf{x} \mapsto f(\mathbf{x}, \mu) = (f_1(x, y, \mu), f_2(x, y, \mu)), \tag{2.13}$$
$$\mathbf{x} = (x, y) \in \mathbf{R}^2, \ \mu \in \mathbf{R}, \tag{2.14}$$

and examine the case of the non-hyperbolic fixed point $\mathbf{x} = 0$ with parameter $\mu = 0$. If the Jacobian

$$A = \begin{pmatrix} (f_1)_x & (f_1)_y \\ (f_2)_x & (f_2)_y \end{pmatrix} (\mathbf{0}, 0), \tag{2.15}$$

takes real eigenvalues λ_1 and λ_2, the bifurcation list is as follows:

Since we examine the non-hyperbolic case, suppose $\lambda_1 \neq 1$. Then the condition that $\mathbf{x} = 0$ is a non-hyperbolic fixed point is $\lambda_2 = 1$ or $\lambda_2 = -1$.

1. If $\lambda_2 = 1$, **saddle-node bifurcation** occurs. If the map $f(\mathbf{x}, \mu)$ always has a fixed point at $\mathbf{x} = \mathbf{0}$ as $f(\mathbf{0}, \mu) = \mathbf{0}$, **transcritical bifurcation** occurs. If the map $f(\mathbf{x}, \mu)$ is an odd function of $\mathbf{x}$ as $f(-\mathbf{x}, \mu) = -f(\mathbf{x}, \mu)$, **pitchfork bifurcation** occurs.

2. If $\lambda_2 = -1$, **period-doubling bifurcation** occurs.

Figs. 2.4 depict these bifurcation phenomena in 2-dimensional map. The top middle figure shows the moment of saddle-node bifurcation at $\mu = 0$ and the fixed point $\mathbf{x} = \mathbf{0}$ belongs to **Milnor attractor** that has both unstable direction and Lebesgue measure positive basin.

These bifurcations and conditions are analogous to the case of 1-dimensional map. Generally, bifurcation phenomena of low-dimensional map naturally appear in high-dimensional cases.

If the Jacobian A takes complex conjugate eigen values, the stable fixed point $\mathbf{x} = \mathbf{0}$ becomes destabilized according to the parameter change, and a stable invariant circle appears around unstable fixed point $\mathbf{x} = \mathbf{0}$. This is equivalent to the time reversal case where the unstable fixed point $\mathbf{x} = \mathbf{0}$ changes to the stable one with the apparition of an unstable invariant circle. These bifurcations are called **Naimark-Sacker bifurcation**.

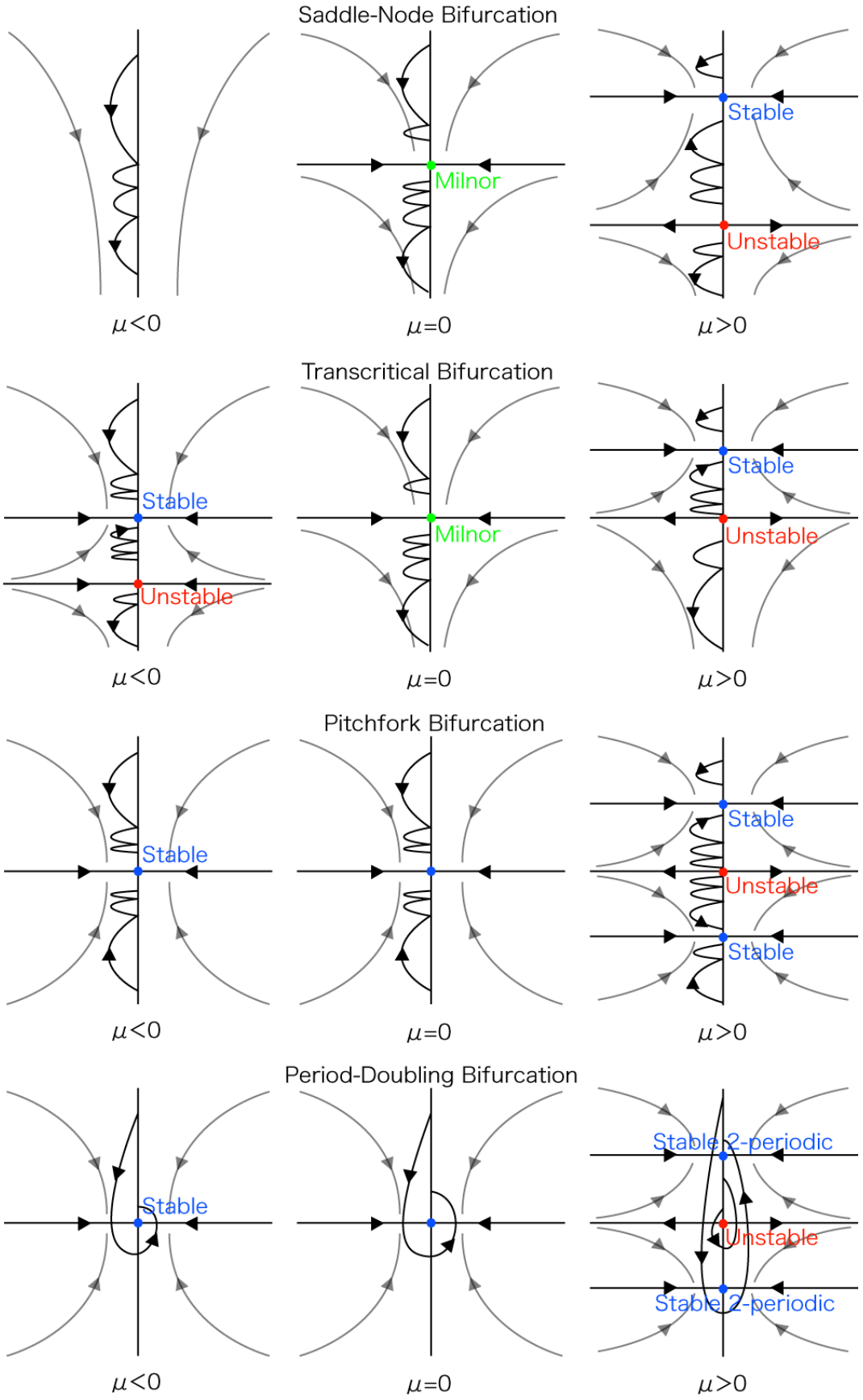


Figure 2.1: Bifurcation list of 2-dimensional discrete-time map on $\mathbf{x}$-plane. From top to bottom: Saddle-node bifurcation, transcritical bifurcation, pitchfork bifurcation, and period-doubling bifurcation. $0 < \lambda_1 < 1$. Blue points represent stable fix points or 2-periodic attractor, red points are unstable fix points, and green points are Milnor attractors. Naimark-Sacker bifurcation is omitted. Based on [1].

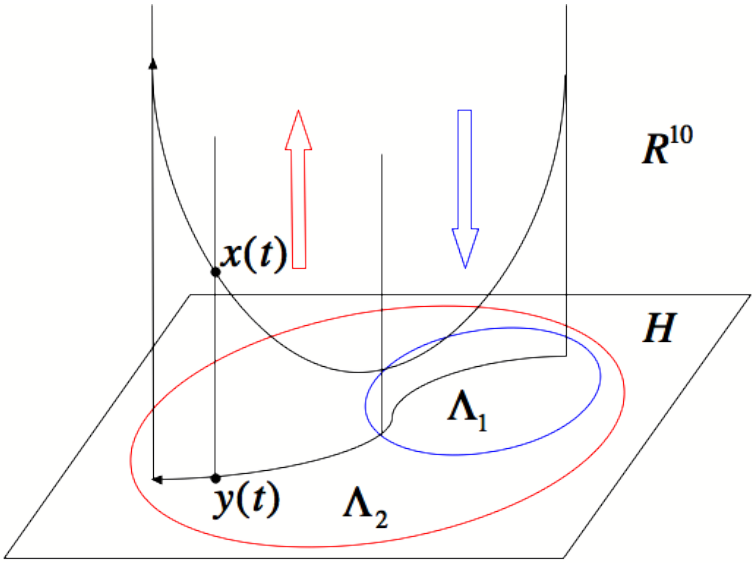


Figure 2.2: Conceptual scheme of chaotic itinerancy in 10-dimensional GCM. When the projected point $y(t)$ of the orbit $x(t)$ moves inside of the attractor ruin Λ_1, transversal stability of the invariant subspace H is stable and $x(t)$ moves closer to H. When the projected point $y(t)$ reaches the exit of Λ_1 and moves inside of Λ_2, transversal stability of H becomes unstable and $x(t)$ moves away from H. Based on [1].

2.5 Chaotic Itinerancy as Crisis-Induced Intermittency

Besides the change of stability of periodic point, there exists another class of bifurcation called **global bifurcation**. Generally, the global expansion of an attractor by touching to unstable periodic orbit is called **crisis**. The crisis is further classified into **interior crisis** and **boundary crisis** according to whether the orbit is constrained inside of a finite domain. The mechanism of global bifurcation is not yet well understood in high-dimensional map.

Chaotic itinerancy observed in globally coupled map (GCM) is characterized in terms of **crisis-induced intermittency**. Interior crisis of attractors produces plural attractor ruins similar to the destabilized Milner attractor, and the orbit shows intermittent dynamics between them. The mechanism of chaotic itinerancy in GCM is partially characterized as the crisis-induced intermittency on invariant subspaces and the inversion of stability in its orthogonal directions. Fig. 2.5 shows the qualitative stability structure of attractor ruin in GCM examined in [1].

2.6 Lyapunov Exponent

Chaotic dynamics is originated from the coexistence of stability and unstability of the orbits, which is expressed as the stretching and folding by the map. To evaluate the existence of chaos, we need to detect orbital unstability. Let us consider the stability of an orbit with 1-dimensional map $x(t+1) = f(x(t))$. We add a small perturbation $\delta x(t)$ to $x(t)$ and measure its time development $\delta x(t+1)$:

$$\begin{aligned} \delta x(t+1) &= f(x(t)+\delta x(t)) - f(x(t)) \\ &= \frac{df(x(t))}{dx}\delta x(t) + O(2) \qquad (2.16) \\ &\approx f'(x(t))\delta x(t). \qquad (2.17) \end{aligned}$$

Where (2.16) uses Taylor expansion at $x(t)$ and $O(2)$ is more than 2nd order terms of $\delta x(t)$. The last approximation suppose that $\delta x(t)$ is sufficiently small. By repeating this linear approximation

for N time steps, we obtain

$$\delta x(t+N) \approx \prod_{k=l}^{N} f'(x(t+k-1))\delta x(t). \tag{2.18}$$

This means that if the inclination of the map $f'(x(t+k-1))$ is larger than 1, the perturbation of the orbit is amplified at time $t+k$, and reduced if less than 1. The exponential amplification of the distance between adjacent two points is the definition of chaos. Therefore, the long-term mean of the exponents of $f'(x(t+k-1))$ gives global information on the stability of the orbit. This is called **Lyapunov exponent** and defined as follows:

$$\lambda = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \log |f'x(t+k)| . \tag{2.19}$$

If $\lambda > 0$, the distance between adjacent orbits increases temporally in an exponential scale. Such property is described as **unstable system**.

For the ergodic maps with irregular orbits that are non-uniformly distributed in the domain of x, we can calculate the Lyapunov exponent by using the spacial mean weighted with the density of invariant measure $\rho(x)$:

$$\lambda = \int \log |f'x| \, \rho(x) dx. \tag{2.20}$$

Lyapunov exponent is also defined in high-dimensional dynamical system, and generally we obtain n exponents for n-dimensional variable. In most cases, if the maximum Lyapunov exponent is positive, we call the system is chaotic.

Chapter 3

Preliminaries of Information Geometry

Abstract

In this chapter, we review the preliminary knowledge of information geometry used in this thesis based on [16] and [2].

Keywords: Statistical manifold, Riemannian metric, α-connection, dual flat coordinates, Kullback-Leibler divergence, extended Pythagorean theorem

3.1 Introduction

Information geometry treats an ensemble of probabilistic distributions as a Riemannian space with a natural geometrical structure compatible to various fields concerning probability, such as statistical, information, and systems theories. The transversal property of information geometry allows us to examine the existing methods from different viewpoints or revealing the relation between different methods. In this sense, information geometry has a property of mutual language for the fields relating to probability.

Information geometry is based on differential geometry of Riemannian space following the nature of probability distributions. In Euclidian space, the concept of "straight" and "flat" is almost trivial and do not need specification. However, in general space (such as Riemannian) these need rigorous mathematical definition.

3.2 Space of Probability Distributions

The start point of information geometry is the probabilistic distribution model $f(X;\xi)$ with stochastic variable X and n-dimensional parameters $\xi = (\xi^1, \cdots, \xi^n)$. If we take the parameters ξ as coordinates, the whole ensemble of this model can be considered as a smooth space, in geometrical term **manifold**, and each probability distribution is a point in this space.

- Ex. 1. Discrete distribution. Consider X is discrete variable taking $\{x_0, x_1, \cdots, x_n\}$, and set $Prob(X = x_i) = q_i (> 0)$. Since $\sum_{i=0}^{n} q_i = 1$, the number of independent parameters is n, and if we take for example $q_1, \cdots, q_n$, it becomes a n-dimensional parameter space.

- Ex. 2. Normal distribution. Define X as 1-dimensional real value whose probabilistic density is $f(X;\mu,\sigma^2) = \exp(-(x-\mu)^2/(2\sigma^2))/\sqrt{2\pi\sigma^2}$. Then this model forms a 2-dimensional space with parameters μ and σ as coordinates.

As were shown in above examples, generally parameters are not defined in the whole range of real value but with its subsets (such as $q_i > 0$, $\sigma > 0$).

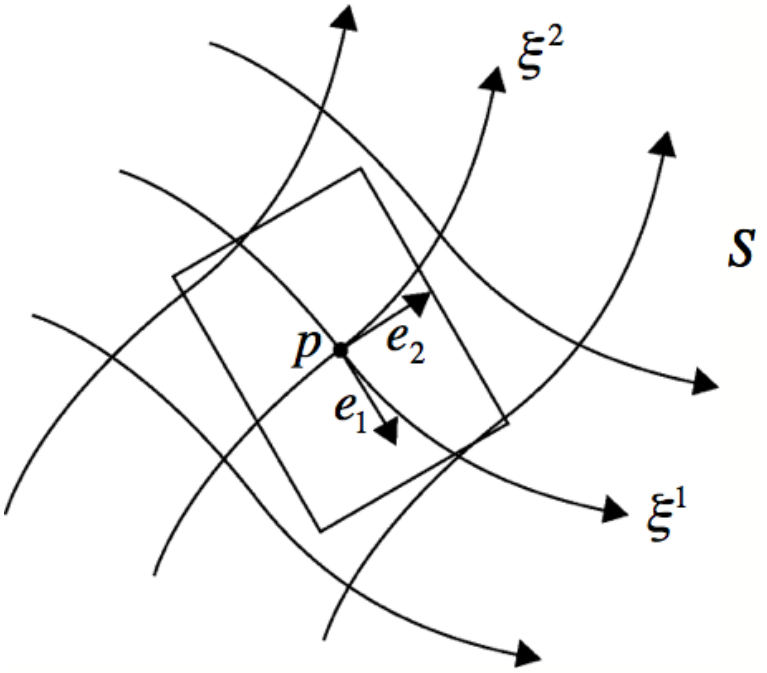


Figure 3.1: Curved space is locally linear if sufficiently magnified. Based on [2].

3.3 Neighborhood of a Point: Euclidian Space

We introduce geometrical structure in the space of statistical model S. The global scenario is first to approximate the neighborhood of each point with Euclidian space and define the local structure with a matrix called **metric**. Next, we define the connectivity between the adjacent neighborhoods with a tensor called **connection**, which leads to define the global structure of S.

Let us explain these concepts by moving a point p on S in a "straight" manner. We describe the coordinates ξ of the point $p \in S$ as $\xi(p)$. Whatever sharp the curvature is, if the space is smooth, sufficiently small neighbor of p can be approximated by the well-known Euclidian space as in Fig. 3.3. Setting p as original point, let us call this local linear space as T_p. In Euclidian space, to move straight a point is simply to follow rectilinear direction.

Though, this notion of moving straight is only valid in the neighborhood of p, and mathematically speaking, the point can only move infinitesimally small distance. Therefore, the straight line direction we considered in this local Euclidian space only specifies the tangent direction associated with the orbit of global straight movement. We call this tangent direction as **tangent vector**, and since T_p is an ensemble of tangent vectors at p, we call is as **tangent space**.

To further pursuit the "straight" movement beyond the linearized neighborhood, we need to use the concept of **connection**, which will be introduced in the next section. In this section, we further focus on the local structure in the tangent space.

We describe the basis of T_p as $e_1, \cdots, e_n$, each corresponding to the direction of coordinates $\xi_1, \cdots, \xi_n$ in S. Then a point on T_p can be specified by the linear sum $\sum_{i=1}^{n} a_i e_i$. To decide the structure of T_p so that to calculate geometrical quantities such as angle and length, we need to define the following inner product between e_i and e_j:

$$g_{ij}(\xi) = \langle e_i, e_j \rangle. \tag{3.1}$$

We call the product $g_{ij}(\xi)$ as **(Riemannian) metric** and list in the matrix $G = (g_{ij}(\xi))$. $g_{ij}(\xi)$ can be defined arbitrary, even be dependent on ξ, according to the nature of geometry we want to construct, as long as G satisfies to be a positive definite symmetric matrix.

In information geometry, we use the following **Fisher information matrix** as Riemannian metric:

$$g_{ij}(\xi) = E_\xi[(\partial_i l)(\partial_j l)]. \tag{3.2}$$

For simplicity, we set $\partial_i = \frac{\partial}{\partial \xi^i}$ and $l = \log f(X;\xi)$. $E_\xi[\ ]$ represents the expected value of $f(X;\xi)$:

$$E_\xi[g(x)] = \int f(X;\xi) g(x) dx. \tag{3.3}$$

There are several important reasons to choose Fisher information matrix among possible Riemannian metrics for information geometry. Besides the properties of Fisher information, we introduce here the relation with **Cramer-Rao inequality** that is a fundamental in statistical estimation. We first represent the estimated parameters with $\hat{\xi}$ that are estimated with some estimation method from N independent sample data. Then the $\hat{\xi}$ stochastically fluctuates according to the finiteness of sampling. If the estimated parameters $\hat{\xi}$ coincide with the true ones ξ^*, the variance of $\hat{\xi}$ satisfies the following with G as Fisher information matrix:

$$Var[\hat{\xi}] \geq \frac{1}{N} G^{-1}. \tag{3.4}$$

We call this relation as **information inequality** or Cramer-Rao inequality. For "good" estimators such as maximum likelihood estimators, the equality of this relation asymptotically holds with respect to the sample number N. Therefore, Fisher information can be considered as the inverse of estimator's variance, so that it is natural to consider it as a distance measure between distributions. For example, if a pair of adjacent distributions have high $Var[\hat{\xi}]$, the distinction between these two distributions are "difficult" in terms of estimation. Therefore, the distance should be defined "close" between them, which is the case of using Fisher information giving small value as $Var[\hat{\xi}]$ converges high.

- Ex. 3. In case of normal distribution. Taking $(\xi^1, \xi^2) = (\mu, \sigma)$ as coordinates, we have $\log f(X;\xi) = \frac{(X-\mu)^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2}$. Then the Fisher information is calculated as the following:

$$G = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \tag{3.5}$$

  Using this, the length of a linear element in T_p by infinitesimally changing (μ, σ) to the direction of $(d\mu, d\sigma)$ is obtained as $\frac{\sqrt{d\mu^2 + 2d\sigma^2}}{\sigma}$. This reflects the fact that when σ is small, tiny change of parameters causes relatively big change of distribution form, while large σ keep the distribution form similar with respect to parameters change. For example, with small σ giving sharp normal distribution, change of mean value μ causes larger deviation of $f(X;\xi)$ than in case of wide distribution with high σ.

If we take another coordinates θ on S, whatever the nonlinear transformation between θ and ξ is, we can always approximate linearly for the neighborhood of a point p. More concretely, it suffices to linearly transform with a Jacobian matrix B taking i, j element as $\frac{\partial \theta^i}{\partial \xi^i}$ at p. Then the point in T_p can be simply transformed from ξ to θ coordinates by multiplying the basis e_i and its coefficient a_i with B. In the same way, Riemannian metric can be also transformed into the expression on θ using B. This means that the concept of tangent space and Riemannian metric is essentially independent or invariant from the way of taking coordinate systems. This concept of **invariance** is one of the principal tools in modern geometry.

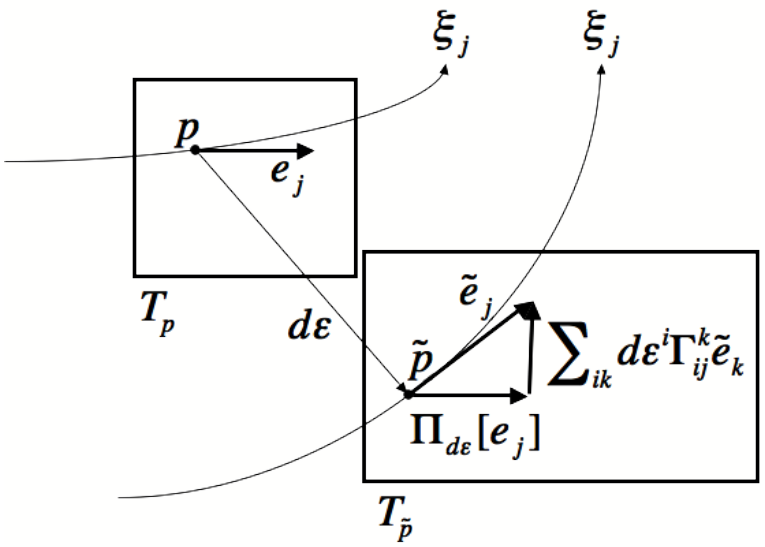


Figure 3.2: Connection coefficient decides the parallel translation between adjacent tangent spaces. Based on [2].

3.4 Connecting Euclidian space

In previous section, we could only move a point p on F for an infinitesimal distance $d\xi$ by considering the tangent space T_p. In this section, we further extend the movement to a global scale. The coordinates of the new point $\tilde{p}$ is given by $\xi(\tilde{p}) = \xi(p) + d\xi$. At $\tilde{p}$, we need to reconstruct the corresponding new tangent space $T_{\tilde{p}}$ so that to define the infinitesimal vector $d\xi'$ that is supposed to be in the "same direction" as $d\xi$. By repeating infinitely this procedure, we can continue to move a point locally and mathematically obtain the global movement by integration.

Let us formulate this process in a general form. Consider a point p and move it to another point $\tilde{p}$ for an infinitesimal distance $d\epsilon = (d\epsilon^1, \cdots, d\epsilon^n)$. Then the vector $d\xi$ on T_p moves on $T_{\tilde{p}}$, which we describe as $\Pi_{d\epsilon}[d\xi]$. We call the movement from $d\xi$ to $\Pi_{d\epsilon}[d\xi]$ as **parallel translation** (Fig. 3.4). Since $d\epsilon$ is infinitesimal, this movement can be approximated with linear transformation. More concretely, we describe the parallel transformation of a basis e_j on T_p as follows:

$$\Pi_{d\epsilon}[e_j] = \tilde{e}_j - \sum_{i,k} d\epsilon^i \Gamma^k_{ij} \tilde{e}_k. \tag{3.6}$$

We call Γ^k_{ij} as **connection (coefficient)**. Intuitively speaking, a tangent vector changes its direction proportionally to the moving distance with the rate of connection coefficient. General tangent vector $d\xi = \sum_{j=1}^n a_j e_j$ then moves to $\sum_{j=1}^n a_j \Pi_{d\epsilon}[e_j]$.

The global "straight" movement of a point can then be defined by continuing local parallel translation of the tangent vector as $d\xi' = \Pi_{d\epsilon}[d\xi]$. The obtained trajectory defines the "straight" line with respect to the defined Riemannian metric and connection coefficient and called as **geodesic**. Note a geodesic does not necessary look straight for an arbitrary coordinates ξ.

3.5 α-connection

How can we decide the connection coefficient ? Natural geometrical requisition is that the parallel translation of two tangent vectors $d\xi_1$ and $d\xi_2$ should not change their geometric relation. More concretely, the inner product should remain invariant before and after the movement. Under this constraint, the connection coefficient is dependent on Riemannian metric g_{ij} and unique. This is called as **Riemannian connection** or **Levi-Civita connection** and used in normal differential geometry where the geometrical structure of space is defined only by Riemannian metric.

However, from a viewpoint of statistical theory, the invariance of inner product is not a primal requisition. The concept of statistical invariance is rather the main issue. The invariance with respect to the transformation of parameters ξ and stochastic variable X are more natural and appropriate to consider. For this reason, in information geometry, we use the following α**-connection** $\Gamma^{(\alpha)}_{ij,k}$ with a free parameter α:

$$\Gamma^{(\alpha)}_{ij,k} = E_\xi \left[\left(\partial_i \partial_j + \frac{1-\alpha}{2} \partial_i l \partial_j l \right) \partial_k l \right], \tag{3.7}$$

where

$$\Gamma_{ij,k} = \sum_h \Gamma^h_{ij} g_{hk}. \tag{3.8}$$

It contains the Riemannian connection since it coincides in case of $\alpha = 0$. In information geometry, the cases of $\alpha = \pm 1$ are particularly important, which will be explained in the following section.

3.6 Flat Space

The connection coefficient represents the difference of alignment between infinitesimally separated tangent spaces. Taking a coordinate system ξ, if all connection coefficient of α-connection is 0, the spacial difference is also 0. Generally such coordinates do not necessary exist, but if exist, we call these as α**-(affine) coordinates** and the space is α**-flat**.

In α-flat space, a geodesic is represented as a straight line on α-coordinates, namely α**-geodesic**. This structure is similar to the straight lines in Euclidian space, although the metric may be different at each point. For this and other utilities, α-flat space is often used for applications in engineering.

- Ex. 4. **Exponential family** is defined as follows, and is 1-flat with respect to θ as affine coordinates:

$$f(X;\theta) = \exp\left(\sum_{i=1}^n \theta^i F_i(X) - \psi(\theta) + C(X) \right). \tag{3.9}$$

  This family of distribution plays principal role in information geometry of statistics, and we particularly call 1-connection, 1-flat as e-connection and e-flat, respectively, taking after e of *exponential*.

  For example, normal distribution belongs to exponential family by replacing $F_1(X) = X$, $F_2(X) = X^2$, and its e-coordinates are $\theta^1 = \frac{\mu}{\sigma^2}$, $\theta^2 = \frac{-1}{2\sigma^2}$.

- Ex. 5. **Mixture family** is defined as a linear sum of probability distribution $F_i(X)$, and is -1-flat with respect to the affine coordinates θ:

$$f(X;\theta) = \sum_{i=1}^n \theta^i F_i(X) \left(1 - \sum_{i=1}^n \theta^i \right) F_0(X). \tag{3.10}$$

  We particularly call -1-connection, -1-flat as m-connection and m-flat, respectively, taking after m of *mixture*.

- Ex. 6. Consider more generally α**-family** with parameter $\alpha \neq 1$ as follows:

$$f(X;\theta) \propto \left(\sum_{i=1}^{n} \theta^i F_i(X) \right)^{\frac{2}{1-\alpha}}. \tag{3.11}$$

If renormalized as $\int_X f(X;\theta)dX = 1$, this family is not generally α-flat except the case $\alpha = -1$ that coincides with mixture family.

As were seen with above examples, only the cases of $\alpha = \pm 1$ are important when considering generally probability distribution. Therefore, in application of information geometry, in most cases we focus on ± 1-connection, namely e- and m-connections.

3.7 Dual Coordinates

The pair of α- and $-\alpha$- connections has interesting properties in mathematical sense. The most fundamental one is the fact that if a space is α-flat, then it is also $-\alpha$-flat, namely **dual flat** structure. Though generally, the corresponding affine coordinates are different. Let us represent the α-coordinates and $-\alpha$-coordinates of a dual flat space S with $\theta = (\theta^1, \cdots, \theta^n)$ and $\eta = (\eta_1, \cdots, \eta_n)$, respectively. These coordinates are interchangeable with the following Legendre transformation with potential functions $\psi(\theta)$ and $\phi(\eta)$:

$$\psi(\theta) + \phi(\eta) - \sum_{i=1}^{n} \theta^i \eta_i = 0, \tag{3.12}$$

$$\frac{\partial \psi(\theta)}{\partial \theta} = \eta, \ \frac{\partial \phi(\eta)}{\partial \eta} = \theta. \tag{3.13}$$

Note that if we describe the Riemannian metrics of θ coordinates and η coordinates with g_{ij} and g^{ij}, respectively, the followings hold:

$$\frac{\partial \eta_i}{\partial \theta^j} = g_{ij}, \ \frac{\partial \theta_i}{\partial \eta^j} = g^{ij}, \tag{3.14}$$

$$(g_{ij}) = (g^{ij})^{-1}. \tag{3.15}$$

Therefore, g_{ij} and g^{ij} are at the same time metrics and Jacobian matrices of local coordinates transformation.

For the basis vectors e_i of α-coordinates and e^j of $-\alpha$-coordinates on a tangent space T_p, there exists the following dual orthogonal relation:

$$\langle e_i, e^j \rangle = \delta_i^j. \tag{3.16}$$

This gives the basis of orthogonal projection explained in the next section. Orthogonality in $\pm\alpha$-flat space appears between the dual coordinates.

- Ex. 7. From dual flat structure, 1-flat (e-flat) exponential family is also -1-flat (m-flat). The corresponding m-coordinates are $\eta_i = E_{\theta[F_i(X)]}$ that forms a space of sufficient statistics. Therefore, given a series of observation data, one can specify the e-coordinates of the point on S by calculating sufficient statistics.

In case of normal distribution (Ex. 4), for instance, $\eta_1 = E[X] = \mu$, $\eta_2 = E[X^2] = \mu^2 + \sigma^2$, and observation data can be expressed as $\eta = (\hat{\mu}, \hat{\mu}^2 + \hat{\sigma}^2)$ using sample mean $\hat{\mu}$ and sample variance $\hat{\sigma}$. The potential function $\psi(\theta)$ is analytically derived from (3.9), and $\psi(\eta)$ can be calculated from (3.12).

In case of mixture family, since it is -1-flat (m-flat), from dual flat structure it is also 1-flat (e-flat). Though, the corresponding e-coordinates are not as simple as those of exponential family. Therefore, although both incorporate dual flat structure, exponential family is easier to relate with statistical estimation than mixture family.

3.8 Subspace and Projection

In information geometry, we generally consider a dual flat statistical manifold S (such as exponential family) including both observed data as a point of empirical distribution and the model as a subspace. In application to machine learning, the learning algorithm is geometrically interpreted as to find the projection of observed data to a model subspace. In this section, we introduce the property of subspaces in S and projection.

In Euclidian space, a projection to a flat subspace is easier to consider than to a curved one. The concept of flat subspace is also important in information geometry. Given a dual flat space S, the flat subspace with α-coordinates M (actually a linear subspace) is called **α-flat subspace**. It is however not assured whether α-flat subspace is $-\alpha$-flat, since the transformation between $\pm\alpha$-coordinates are generally non-linear and can not be separated into different independent sets of sub-coordinates.

The concept of divergence is important to consider a projection to a subspace. α-divergence between two points p, q in a dual flat space is defined as follows, similar to the Legendre transformation (3.12):

$$D^{\alpha}[p:q] = \psi(\theta(p)) + \phi(\eta(q)) - \sum_{i=1}^{n} \theta^i(p)\eta_i(q). \tag{3.17}$$

This quantity represents the discrepancy between the two points, but does not satisfy the axioms of mathematical "distance" in terms of symmetric law and triangle inequality. Nevertheless, we dare consider α-divergence since it is compatible to the affine coordinates and still possesses important property of distance. More concretely, $D^{\alpha}[p:q] \geq 0$ and the equality holds for $p = q$. When p is sufficiently close to q, it converges to the conventional distance. The dual $-\alpha$-divergence is $D^{-\alpha}[p:q] = D^{\alpha}[q:p]$.

Particularly for exponential family, the e-divergence with $\alpha = 1$ coincides with Kullback-Leibler divergence between two distributions $f(X)$ and $g(X)$:

$$D^{1}[f:g] = \int f(X)[\log f(X) - \log g(X)]dX. \tag{3.18}$$

The dual m-divergence with $\alpha = -1$ is $D^{-1}[f:g] = D^{1}[g:f]$.

The projection in Euclidian space is defined in relation to Pythagorean theorem. In this case, the distance between a point and a subspace is decomposed between orthogonal and parallel elements. In information geometry, we utilize the following extended Pythagorean theorem in dual flat space.

> **Extended Pythagorean theorem.**
> For the points p, q, r on a dual flat space S, connect p and q with α-geodesic, and q and r with $-\alpha$-geodesic. When the tangent vectors of these geodesics cross orthogonally at q, the following holds:

$$D^{\alpha}[p:r] = D^{\alpha}[p:q] + D^{\alpha}[q:r]. \tag{3.19}$$

Here, if the α-geodesic between the point p on S and the subspace M crosses orthogonally with M at q, we call this projection as **α-projection**. From the extended Pythagorean theorem, we derive the following relation between the α-projection to a subspace and α-divergence.

> **Projection theorem.**
> The α-projection q on M from a point p on dual flat space S is a stationary point of α-divergence $D^{\alpha}[p:q]$. If M is a $-\alpha$-flat subspace, the projection is unique and take the minimum value of $D^{\alpha}[p:q]$.

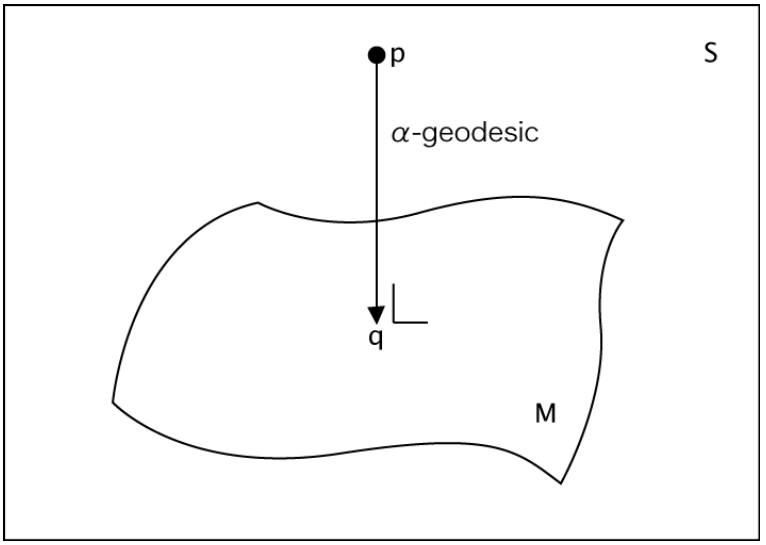


Figure 3.3: Projection is a stationery point of divergence. Based on [2].

This theorem is schematically represented in Fig. 3.8.

Since S is dual flat, the extended Pythagorean theorem and the projection theorem also hold with the permutation of α and $-\alpha$. From the projection theorem, if M is a $-\alpha$-flat subspace, it is natural to take α-projection. In such case, if we take separately $-\alpha$-coordinates inside of M and α-coordinates outside of M, the geometrical situation becomes similar to that of Euclidian space.

Suppose M is a k-dimensional $-\alpha$-flat subspace, and describe separately the first k coordinates from the rest $n-k$ ones as (θ^I, θ^{II}) and (η_I, η_{II}). Using appropriate linear transformation, M can be defined as a linear subspace $\eta_{II} = \hat{\eta}_{II} = Const.$ as depicted in Fig. 3.8. Then consider newly another coordinates (θ^I, η_{II}), namely **mixture coordinates**. Still, any point on S can be uniquely specified with this system. Using the mixture coordinates, the α-projection from (θ^I, η_{II}) to M is simply represented with the difference of latter $n-k$ elements as $(\theta^I, \hat{\eta}_{II})$, giving a concrete expression of the projected point.

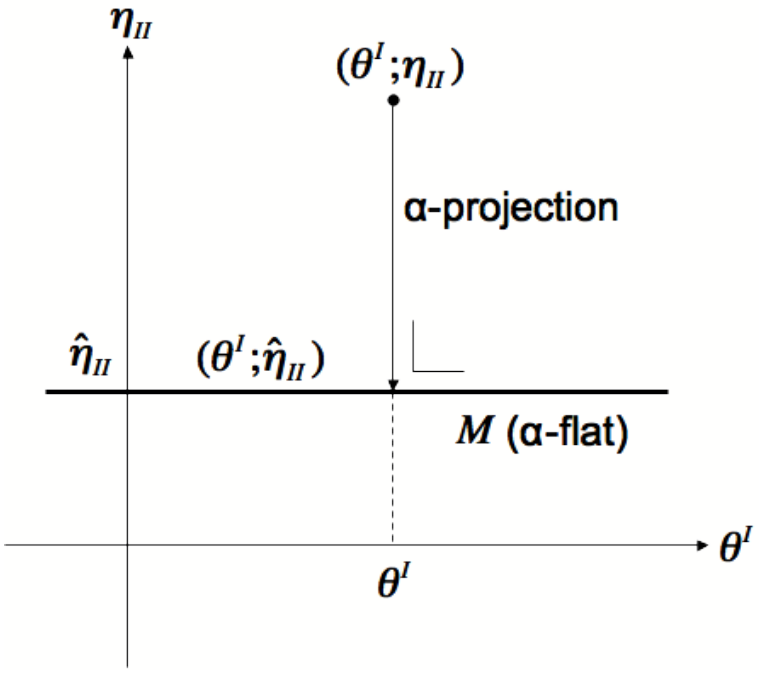


Figure 3.4: Representation with mixture coordinates looks like Euclidian space. Based on [2].

Part II

Synthetic Modeling with Chaotic Neural Network

Chapter 4

Synthetic Modeling of Autonomous Learning with a Chaotic Neural Network

Abstract

We investigate the possible role of intermittent chaotic dynamics called chaotic itinerancy, in interaction with non-supervised learnings that reinforce and weaken the neural connection depending on the dynamics itself. We first performed hierarchical stability analysis of the chaotic neural network model (CNN) according to the structure of invariant subspaces. Irregular transition between two attractor ruins with positive maximum Lyapunov exponent was triggered by the blowout bifurcation of the attractor spaces, and was associated with riddled basins structure. We secondly modeled two autonomous learnings, Hebbian learning and spike-timing-dependent plasticity (STDP) rule, and simulated the effect on the chaotic itinerancy state of CNN. Hebbian learning increased the residence time on attractor ruins, and produced novel attractors in the minimum higher dimensional subspace. It also augmented the neuronal synchrony and established the uniform modularity in chaotic itinerancy. STDP rule reduced the residence time on attractor ruins, and brought a wide range of periodicity in emerged attractors, possibly including strange attractors. Both learning rules selectively destroyed and preserved the specific invariant subspaces, depending on the neuron synchrony of the subspace where the orbits are situated.

Keywords: Chaotic itinerancy, invariant subspace, blowout bifurcation, riddled basins, Hebbian learning, STDP learning.

Methodology: Modeling of autonomous learnings with CNN as a dynamical system → Linear stability analysis of CNN as a dynamical system → Analysis of emerged neural correlation with information geometry

4.1 Introduction

Recent development of neuroscience has been largely promoted by the interaction between biological experiment and mathematical modeling. Although experimental setting is limited to a part of entire phenomenon situated in the ultimate interest, different measurement methods varying in both temporal and spacial scales enable us to acquire experimental evidences ranging from single neuron to cortical and cerebral activities, and integrate them into a simulated model with certain assumption. This interaction between experimental measurement and mathematical modeling helps us not only to reproduce the observed brain dynamics itself, but to discover appropriate levels of description that are accessible to mathematical analysis and eventually help hypothesis testing and further hypothesis forming in experiment.

Skarda and Freeman investigated the chaotic dynamics in olfactory bulb of rabbit with such experiment-modeling interaction, and discovered the function of chaos as catalyst of learning in macroscopic level [17]. Here, the chaotic wandering state among memorized patterns is shown to additionally create another attractor corresponding to the novel oder.

Later Tsuda showed in a modeling study of cortical chaos that such intermittent chaos associated with temporal laminar phase realizes the acquisition of externally exposed patterns without loosing the former memory structure [18]. This dynamical linking of memory with chaotic transition is further investigated in a more physiologically detailed settings, and are conceptualized with the name of *chaotic itinerancy* [19] [20].

Chaotic itinerancy is based on the concept of attractor ruins, where plural attractors localized in low-dimensional states loses their transverse stability. Possible candidate of attractor ruin is considered as destabilized Milnor attractor that went through the crisis [21][22][23]. Chaotic itinerancy is also observed in a chaotic neural network model (CNN), with the chaotic neuron based on the physiological property of the squid giant axon [24] [25].

Besides the learning of the exposed patterns, there has recently been studies on the self-organizing change of the network' s connectivity without any supervisor signals in neural network model [26] and in relevance to neural network [27]. These spontaneous changes of connectivity are based on the local learning rule such as Hebbian learning and the spike-time-dependent plasticity (STDP) learning. Indeed, such autonomous learning depending on the dynamics of the network itself can be considered to happen widely and permanently as long as the brain manifests spontaneous activity and possesses plasticity. It is of further interest whether and how such internal dynamics can lead to structure formation of neural network and realize adaptive function.

In this article, we investigate the possible role of chaotic itinerancy associated with autonomous learning without supervisor signal. For this purpose, we combine the chaotic neural network model with Hebbian and STDP learning rules and investigate the dynamics.

4.2 Dynamics of Chaotic Neural Network

We first define the chaotic neural network model (CNN) with two periodic attractors in different invariant subspaces, and investigate the dynamics without any learning rule.

4.2.1 Definition of Chaotic Neural Network

The architecture of CNN is defined as in Fig. 4.1. CNN is a discrete time system that consists of two layers of neuron model: Context and Output. Each neuron of the Context layer is connected to all neurons in the Output layer. The output of the Context layer at time t is identical to that of the Output layer at time $t-1$. The chaotic neuron used in the Output layer of CNN are defined as follows [24]:

$$\begin{aligned} x_i(t+1) &= f(\sum_{j=1}^{n} w_{ij} \sum_{d=0}^{t} k_f^d x_j(t-d) \\ &\quad -\alpha \sum_{d=0}^{t} k_r^d x_i(t-d) + \theta_{out}), \end{aligned} \tag{4.1}$$

where at time t , $x_i(t)$ is the output of the ith chaotic neuron, and θ_i^{out} is its threshold. The parameters $\alpha \geq 0$ and k_r $(0 \leq k_r \leq 1)$ control the refractoriness of neuron. By augmenting these parameters, the orbital stability changes, and can induce chaotic dynamics. Hence, the situation $\alpha = k_r = 0$ corresponds to simple analog neuron model. The exponentially decreasing influences of past outputs (outputs of the Context layer) are controlled by k_f $(0 \leq k_f \leq 1)$. n is the number of chaotic neurons. In this study, eight chaotic neurons with fixed parameter $k_f = 0.1$ were used in the Output layer of CNN. θ_{out} will be defined as a function of k_r in the following section.

The sigmoid function $f(\cdot)$ is defined as follows, with the increment parameter β.

$$f(x) = \frac{1}{1+\exp(-\beta x)}. \tag{4.2}$$

Table 4.1: **Definition of the four patterns A, B, C, D to be memorized in the Output layer of CNN.**

Pattern symbol	Pattern vector
A	$(1,1,1,1,0,0,0,0)^T$
B	$(0,0,0,0,1,1,1,1)^T$
C	$(1,1,0,0,0,0,1,1)^T$
D	$(0,0,1,1,1,1,0,0)^T$

We used $\beta = 5.0$ in this study.

The connection matrix $W = (w_{ij})$ was set with Hebbian rule to memorize two two-periodic patterns A⇌B and C⇌D defined in Tab. 4.1. Let $A = (a_1, \cdots, a_8)^T$, $B = (b_1, \cdots, b_8)^T$, $C = (c_1, \cdots, c_8)^T$, $D = (d_1, \cdots, d_8)^T$, where T is the transpose of vector. Then $W = (w_{ij})$ $(1 \leq i, j \leq 8)$ was defined as follows:

$$w_{ij} = \frac{1}{4}\{(2a_i - 1)(2b_j - 1) + (2b_i - 1)(2a_j - 1) + (2c_i - 1)(2d_j - 1) + (2d_i - 1)(2c_j - 1)\}. \quad (4.3)$$

Consequently,

$$W = \begin{pmatrix} -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \end{pmatrix}, \quad (4.4)$$

was obtained. Note that the number of neurons and the patterns of the attractors was chosen to realize chaotic itinerancy between attractor ruins situated in different invariant subspaces with the simplest network, which is described in the following section.

We judge the retrieval of a pattern at time t by interpreting the output $\mathbf{x}(t) = (x_1(t), \cdots, x_8(t))^T$ into binary value $\mathbf{h}(t) = (h_1(t), \cdots, h_8(t))^T$ as follows:

$$h_i(t) = \begin{cases} 0 & if \quad x_i(t) < 0.5 \\ 1 & else \end{cases}. \quad (4.5)$$

We state that the network retrieved a pattern when the value of $\mathbf{h}(t)$ coincides with any of the patterns A,B,C,D. By setting the refractoriness parameters $k_r = \alpha = 0$, the network retrieved either of two-periodic patterns A⇌B or C⇌D depending on the initial condition. Note almost all initial conditions converge either of these patterns, unless specifically taken on the invariant subspaces where (11335577) is not included (See next section).

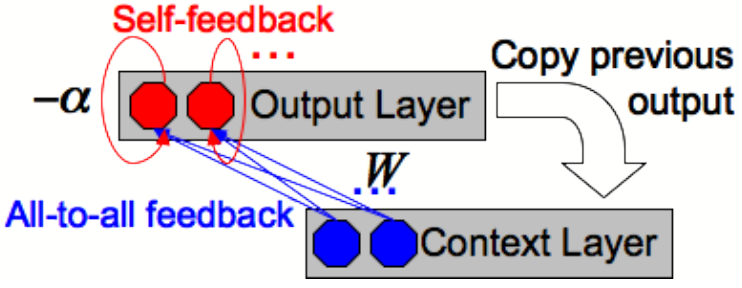


Figure 4.1: **Architecture of CNN.**

4.2.2 Invariant Subspaces

In discrete time neural networks, considering certain symmetry of the model, it is possible to decompose its dynamics into subspaces of the whole state space. The decomposition of the dynamics into lower dimensions has a possibility to characterize the network dynamics more precisely. The decomposed subspaces in which dynamics can fall and remains permanently are called invariant subspaces. In this section, we derive the invariant subspaces of CNN based on [28].

First of all, let the CNN be a differentiable map

$$\Phi = (\Phi_f, \Phi_r) : R^n \times R^n \to R^n \times R^n, \tag{4.6}$$

defined by

$$\begin{cases} \Phi_f(\eta, \zeta) = k_f \eta + W\mathbf{x} \\ \Phi_r(\eta, \zeta) = k_r \zeta + \theta_r - \alpha\mathbf{x} \end{cases} . \tag{4.7}$$

Where

$$\mathbf{x} = \mathbf{x}(\eta, \zeta) = (f(\eta_i + \zeta_i))_{i=1}^n, \tag{4.8}$$

and

$$\theta_r = \theta_{out}(1 - k_r). \tag{4.9}$$

$\theta_r = 0.0$ was used for the following simulation.

Now, let S_n be a symmetric group with degrees of n , and $C(W)$ be a subgroup of S_n. For $\sigma \in S_n$, we define a linear transformation $P_\sigma : R^n \to R^n$ by

$$P_\sigma : (u_1, \cdots, u_n)^T \mapsto (u_{\sigma(1)}, \cdots, u_{\sigma(n)})^T. \tag{4.10}$$

For $\sigma \in S_n$, we define that σ belongs to $C(W)$ if and only if

$$P_\sigma W = W P_\sigma. \tag{4.11}$$

This condition corresponds to the condition $w_{ij} = w_{\sigma(i)\sigma(j)}$ of the connection matrix between the Output and Context layer. Using this symmetry of the connection matrix, we define the invariant subspaces of CNN as follows. Let $\sigma \in C(W)$. Then $\Phi = (\Phi_f, \Phi_r)$ of a CNN is $P_\sigma \times P_\sigma$ -invariant;

$$(P_\sigma \times P_\sigma)\Phi = \Phi(P_\sigma \times P_\sigma). \tag{4.12}$$

We define a linear subspace $H_1(\sigma)$ of R^n by

$$H_1(\sigma) = \{(u_1, \cdots, u_n)^T \in R^n \mid u_i = u_{\sigma(i)}, 1 \le i \le n\}. \tag{4.13}$$

We also define the invariant subspace $H(\sigma)$ of R^{2n} by

$$H(\sigma) = H_1(\sigma) \times H_1(\sigma) = \{(\eta_1, \cdots, \eta_N; \zeta_1, \cdots, \zeta_N)^T \in (\eta_i, \zeta_i) = (\eta_{\sigma(i)}, \zeta_{\sigma(i)}), 1 \le i \le n\}. \tag{4.14}$$

For each $\sigma \in S_n$, $H(\sigma) \subset R^{2n}$ is Φ -invariant;

$$\Phi(H(\sigma)) \subset H(\sigma). \tag{4.15}$$

Hence, this holds for all time steps of CNN. Consequently, it means that if the initial states of CNN were taken inside of $H(\sigma)$, the i th neuron and the $\sigma(i)$ th neuron of the Output layer are always synchronized regardless of transversal stability, and output the same value at each time step.

Here, we give an example of the invariant subspace. If

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 4 & 3 & 6 & 5 & 8 & 7 \end{pmatrix} \in C(W), \tag{4.16}$$

then

$$H(\sigma) = \{(\eta_1, \eta_1, \eta_3, \eta_3, \eta_5, \eta_5, \eta_7, \eta_7; \zeta_1, \zeta_1, \zeta_3, \zeta_3, \zeta_5, \zeta_5, \zeta_7, \zeta_7)^T\}. \tag{4.17}$$

For simplicity, let us denote

$$H(\sigma) = (11335577), \tag{4.18}$$

representing $H(\sigma)$ with the minimum subscripts of vector η and ζ that are identical by P_σ . We take this notation rule for the following sections.

The defined patterns A and B are situated inside of (11115555), while C and D are inside of (11333311). These two invariant subspaces are isomorphic from the symmetry of the model, and therefore it is sufficient to investigate only one subspace to know the dynamics in both.

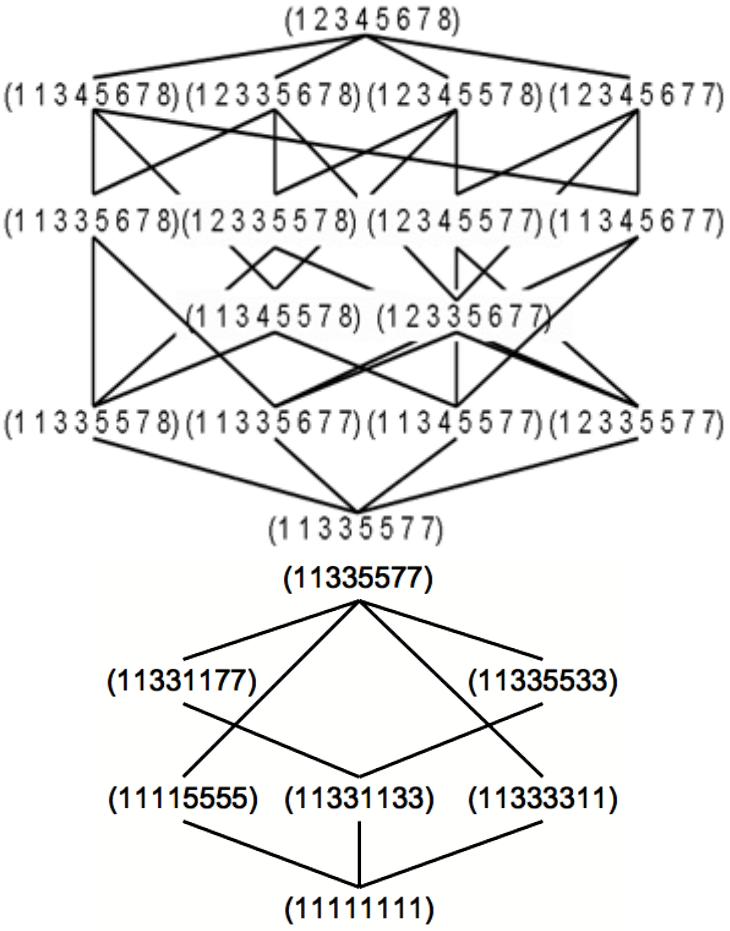


Figure 4.2: **Invariant subspaces of CNN in the defined setting.** Top: Above (11335577). Bottom: Below (11335577). The lines denote inclusion relation. Although there exist actually 76 invariant subspaces, for simplicity only those in the above/below hierarchy of (11335577) are shown.

4.2.3 Periodicity Analysis

In chaotic itinerancy state of CNN, the orbit is reported to visit irregularly among memorized patterns [25]. We investigated the periodicity of the temporal dynamics to search for irregular orbits where chaotic itinerancy may occur, by changing empirically the refractoriness parameters k_r and α. The result is shown in Fig. 4.3.

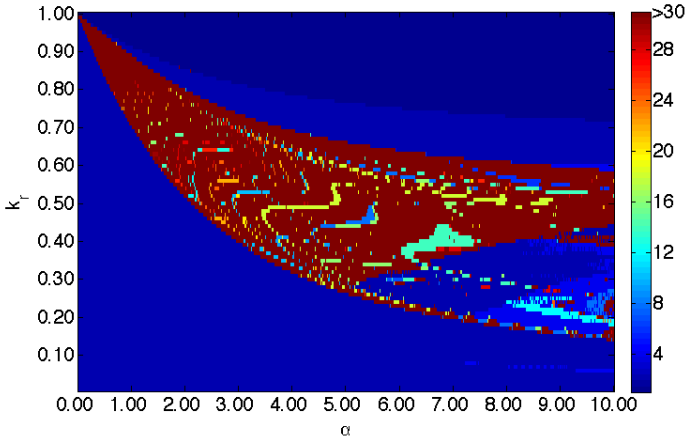


Figure 4.3: **Periodicity of CNN.** Horizontal axis: α. Vertical axis: k_r. After cutting 5000 transient, periodicity was judged with the tolerance of 1.0e-6 for each neuron output. More than 30 periodic orbit is depicted in brown. After cutting 5000 transient, 10000 steps were used to calculate the periodicity at each 0.01 step of k_r and α. Initial conditions were taken randomly.

4.2.4 Deviation Rate from Attractor Ruins

Chaotic itinerancy is characterized not only by the lack of periodicity, but the intermittent visit to the attractor ruins. To find out such parameter region in $k_r - \alpha$ plane, we calculated the deviation rate $DR(k_r, \alpha)$ from the patterns A,B,C,D, which is defined as follows:

$$\begin{aligned} DR(k_r,\alpha) &= \frac{1}{N}\lim_{N\to\infty}\sum_{t=1}^{N}[1-\prod_{i=1}^{n}\{\delta(h_i(t),a_i)\}-\prod_{i=1}^{n}\{\delta(h_i(t),b_i)\} \\ &\quad -\prod_{i=1}^{n}\{\delta(h_i(t),c_i)\}-\prod_{i=1}^{n}\{\delta(h_i(t),d_i)\}], \end{aligned} \tag{4.19}$$

where $\delta(\cdot,\cdot)$ is the delta function. $N = 10000$ was used for the calculation. The result is shown in Fig. 4.4. In case of $DR(k_r, \alpha) = 0$, the orbit stays either of the periodic cycle A⇌B or C⇌D. While in case $DR(k_r, \alpha) = 1$, the orbit never visit any of the memorized patterns. Chaotic itinerancy can occur in $0 < DR(k_r, \alpha) < 1$.

We also investigated the wandering range among the patterns A,B,C,D in $k_r - \alpha$ plane to find out the parameter region where the orbits visit both pair of patterns A⇌B and C⇌D. In such case, both periodic cycle are partially destabilized and become the attractor ruins that allow the orbits to escape and be attracted intermittently. The result is shown in Fig. 4.5.

The irregular transition between two attractor ruins is therefore occurring in the parameter regions where there is no obvious periodicity (inside of brown regions in Fig. 4.3), more than 0 but less than 1 deviation rate in Fig.4.4, and wandering orbit ranging among all patterns (red regions in Fig. 4.5).

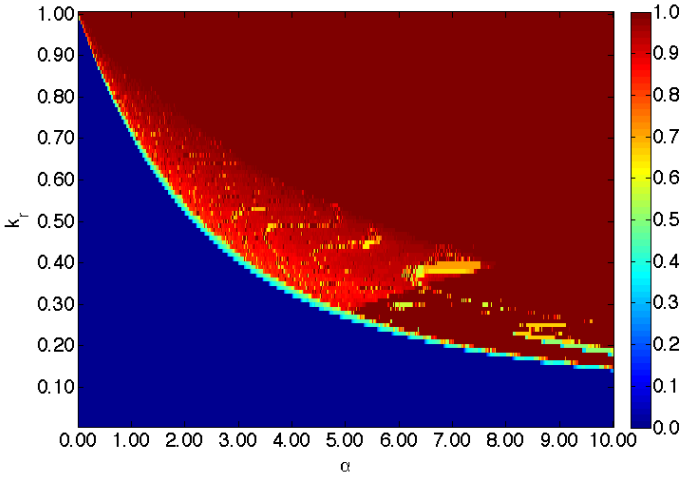


Figure 4.4: **Deviation rate from memorized patterns of CNN.** Horizontal axis: α. Vertical axis: k_r. After cutting 5000 transient, 10000 steps were used to calculate the deviation rate at each 0.01 step of k_r and α. Initial conditions were taken randomly.

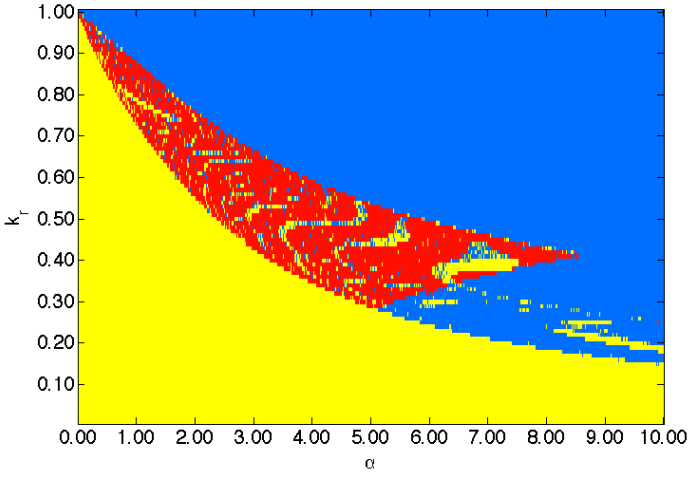


Figure 4.5: **Wandering range among memorized patterns of CNN.** Horizontal axis: α. Vertical axis: k_r. The blue region is where the dynamics visit none of the patterns A, B, C, D. The yellow region is where the dynamics visit only either pair of patterns A⇌B or C⇌D. The red region is where the dynamics visit both pairs of patterns A⇌B and C⇌D. After cutting 5000 transient, 10000 steps were used to calculate the wandering range at each 0.01 step of k_r and α. Initial conditions were taken randomly.

4.2.5 Linear Stability Analysis

Chaotic itinerancy state is assumed to follow chaotic orbit in the transition process between attractor ruins. Therefore, the orbital stability should be investigated to prove the existence of chaos. Since possible attractor ruins A⇌B and C⇌D are situated inside of the invariant subspaces (11115555) and (11333311) respectively, chaotic wandering orbits between these patterns are situated in the upper hierarchy of invariant subspaces. Indeed, in most of the parameter region, the dynamics is situated inside of (11335577). To properly examine the stability of each invariant subspace, it is important to consider the effect of such hierarchical structure on the dynamics . Classical method to calculate Lyapunov spectrum such as Gram-Schmidt orthonormalization simply calculate the Lyapunov exponents in the descending order, and does not necessary consider such symmetry of the model. We derive here the method to decompose the Lyapunov spectrum according to a series of chosen invariant subspaces.

First, we formulate the Jacobian matrix $D\Phi(\eta,\zeta)$ of CNN as follows:

$$D\Phi(\eta,\zeta)=\begin{pmatrix} D_\eta\Phi_f(\eta,\zeta) & D_\zeta\Phi_f(\eta,\zeta) \\ D_\eta\Phi_r(\eta,\zeta) & D_\zeta\Phi_r(\eta,\zeta) \end{pmatrix}, \tag{4.20}$$

where

$$D_\eta\Phi_f(\eta,\zeta)=\begin{pmatrix} k_f & & 0 \\ & \ddots & \\ 0 & & k_f \end{pmatrix}+W\begin{pmatrix} D_\eta f(\eta_1+\zeta_1) & & 0 \\ & \ddots & \\ 0 & & D_\eta f(\eta_n+\zeta_n) \end{pmatrix}, \tag{4.21}$$

$$D_\zeta\Phi_f(\eta,\zeta)=W\begin{pmatrix} D_\zeta f(\eta_1+\zeta_1) & & 0 \\ & \ddots & \\ 0 & & D_\zeta f(\eta_n+\zeta_n) \end{pmatrix}, \tag{4.22}$$

$$D_\eta\Phi_r(\eta,\zeta)=-\alpha\begin{pmatrix} D_\zeta f(\eta_1+\zeta_1) & & 0 \\ & \ddots & \\ 0 & & D_\zeta f(\eta_n+\zeta_n) \end{pmatrix}, \tag{4.23}$$

$$D_\zeta\Phi_r(\eta,\zeta)=\begin{pmatrix} k_r & & 0 \\ & \ddots & \\ 0 & & k_r \end{pmatrix}-\alpha\begin{pmatrix} D_\zeta f(\eta_1+\zeta_1) & & 0 \\ & \ddots & \\ 0 & & D_\zeta f(\eta_n+\zeta_n) \end{pmatrix}. \tag{4.24}$$

D_η and D_ζ denote the differential operator with respect to η and ζ, respectively.

We consider the perturbation in the state apace of the chaotic neuron $(\eta,\zeta)\in R^{2n}$ to the directions inside and outside of the invariant subspaces (11115555) and (11335577), where chaotic itinerancy takes place. For simplicity, we define the transpose σ_2 and σ_4 as follows, under which the orbits inside of two-dimensional (11115555) and four-dimensional (11335577) are invariant, respectively.

$$\sigma_2=\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 1 & 2 & 3 & 8 & 5 & 6 & 7 \end{pmatrix}\in C(W), \tag{4.25}$$

$$\sigma_4=\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 4 & 3 & 6 & 5 & 8 & 7 \end{pmatrix}\in C(W). \tag{4.26}$$

We then define the perturbation vectors $p_1^{2in},\cdots,p_4^{2in}$ to the directions inside of (11115555), and $p_1^{2out},\cdots,p_4^{2out}$ to the directions outside of (11115555) but inside of (11335577) as follows:

$$(p_1^{2in},\cdots,p_4^{2in}) = \delta \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (p_1^{2out},\cdots,p_4^{2out}) = \delta \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (4.27)$$

where δ represents infinitesimal quantity. Hence the eight vectors $p_1^{2in},\cdots,p_4^{2in},p_1^{2out},\cdots,p_4^{2out}$ form the linearly independent basis of the state space $(\eta,\zeta) \in (11335577)$.

In the same way, we also define the perturbations $p_1^{4in},\cdots,p_8^{4in}$ to the direction inside of (11335577), and $p_1^{4out},\cdots,p_8^{4out}$ to the direction outside of (11335577) as follows:

$$(p_1^{4in},\cdots,p_8^{4in}) = \delta \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (4.28)$$

$$(p_1^{4out},\cdots,p_8^{4out}) = \delta \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \quad (4.29)$$

The sixteen vectors $p_1^{4in}, \cdots, p_8^{4in}, p_1^{4out}, \cdots, p_8^{4out}$ form the linearly independent basis of the state space $(\eta, \zeta) \in (12345678)$.

We then consider the development of these perturbations with time development. Using the numerical symmetry of Jacobian matrix $D\Phi(\eta, \zeta)$, the linear approximation of the perturbation development for one time step $D\Phi(\eta, \zeta)p$ has the following relation.

$$D\Phi(\eta,\zeta)p \in (11115555) \quad if \quad p \in \{p_1^{2in}, \cdots, p_4^{2in}\},\ (\eta,\zeta) \in (11115555), \tag{4.30}$$
$$D\Phi(\eta,\zeta)p \in (11115555)^{\perp} \quad if \quad p \in \{p_1^{2out}, \cdots, p_4^{2out}\},\ (\eta,\zeta) \in (11115555), \tag{4.31}$$
$$D\Phi(\eta,\zeta)p \in (11335577) \quad if \quad p \in \{p_1^{4in}, \cdots, p_8^{4in}\},\ (\eta,\zeta) \in (11335577), \tag{4.32}$$
$$D\Phi(\eta,\zeta)p \in (11335577)^{\perp} \quad if \quad p \in \{p_1^{4out}, \cdots, p_8^{4out}\},\ (\eta,\zeta) \in (11335577). \tag{4.33}$$

Here, $(11115555)^{\perp}$ is the state space outside of (11115555) but inside of (11335577), while $(11335577)^{\perp}$ is outside of (11335577). For simplicity, we call $(11115555)^{\perp}$ simply as "outside of (11115555)". These relations mean that the perturbation developments are strictly separated according to the hierarchical structure of the invariant subspaces, and it is possible to decompose them into the directions inside and outside of where the orbit are situated.

Since the invariant subspaces are defined on the synchrony of the state space, their geometrical compositions are situated always in the diagonal lines of the coordinates (η, ζ). We are now motivated to choose another coordinate system to separate the independent development of the perturbations. We first define the coordinates $(\mathbf{H}, \mathbf{Z}) = (H_1, \cdots, H_8, Z_1, \cdots, Z_8)^T$ which separate (11335577) and $(11335577)^{\perp}$ as follows.

$$(\mathbf{H}, \mathbf{Z}) = A(\eta, \zeta), \tag{4.34}$$

where

$$A = (a_{ij}) \tag{4.35}$$

$$= \begin{pmatrix} cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & cos(\pi/4) & sin(\pi/4) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & cos(\pi/4) & sin(\pi/4) \\ -cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -cos(\pi/4) & sin(\pi/4) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -cos(\pi/4) & sin(\pi/4) \end{pmatrix} \tag{4.36}$$

The transformation matrix A is chosen so that $H_1, \cdots, H_4$ and $Z_1, \cdots, Z_4$ form the coordinates inside of (11335577), while $H_5, \cdots, H_8$ and $Z_5, \cdots, Z_8$ outside of (11335577).

Then the CNN $\Phi : R^{2n} \to R^{2n}$ can be reformulated on the coordinates $(\mathbf{H}, \mathbf{Z})$ as $\Psi : R^{2n} \to R^{2n}$:

$$\Psi = (\Psi_f, \Psi_r) : R^n \times R^n \to R^n \times R^n, \tag{4.37}$$
$$\begin{cases} \Psi_f(\mathbf{H}(t), \mathbf{Z}(t)) = \mathbf{H}(t+1) \\ \Psi_r(\mathbf{H}(t), \mathbf{Z}(t)) = \mathbf{Z}(t+1) \end{cases}, \tag{4.38}$$
$$\begin{cases} H_i(t+1) = k_f H_i(t) + \sum_k a_{ik} \sum_j w_{kj} f(\sum_l a_{lj}(H_l(t) + Z_l(t))) \\ Z_i(t+1) = k_r Z_i(t) + \sum_k a_{ik}\theta_r - \alpha \sum_k a_{ik} f(\sum_l a_{lk}(H_l(t) + Z_l(t))) \end{cases} (1 \le i \le 8). \tag{4.39}$$

The corresponding Jacobian matrix $D\Psi(\mathbf{H}, \mathbf{Z})$ becomes as follows:

$$D\Psi(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} D_H\Psi_f(\mathbf{H},\mathbf{Z}) & D_Z\Psi_f(\mathbf{H},\mathbf{Z}) \\ D_H\Psi_r(\mathbf{H},\mathbf{Z}) & D_Z\Psi_r(\mathbf{H},\mathbf{Z}) \end{pmatrix}, \tag{4.40}$$

where

$$D_H\Psi_f(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} k_f & & 0 \\ & \ddots & \\ 0 & & k_f \end{pmatrix} + \begin{pmatrix} \sum_k a_{1k}\sum_j w_{kj} D_{H_1} f(\sum_l a_{lj}(H_l+Z_l)) & \cdots & \sum_k a_{1k}\sum_j w_{kj} D_{H_8} f(\sum_l a_{lj}(H_l+Z_l)) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k}\sum_j w_{kj} D_{H_1} f(\sum_l a_{lj}(H_l+Z_l)) & \cdots & \sum_k a_{8k}\sum_j w_{kj} D_{H_8} f(\sum_l a_{lj}(H_l+Z_l)) \end{pmatrix} \quad (4.41)$$

$$D_Z\Psi_f(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} \sum_k a_{1k}\sum_j w_{kj} D_{Z_1} f(\sum_l a_{lj}(H_l+Z_l)) & \cdots & \sum_k a_{1k}\sum_j w_{kj} D_{Z_8} f(\sum_l a_{lj}(H_l+Z_l)) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k}\sum_j w_{kj} D_{Z_1} f(\sum_l a_{lj}(H_l+Z_l)) & \cdots & \sum_k a_{8k}\sum_j w_{kj} D_{Z_8} f(\sum_l a_{lj}(H_l+Z_l)) \end{pmatrix} \quad (4.42)$$

$$D_H\Psi_r(\mathbf{H},\mathbf{Z}) = -\alpha \begin{pmatrix} \sum_k a_{1k} D_{H_1} f(\sum_l a_{lk}(H_l+Z_l)) & \cdots & \sum_k a_{1k} D_{H_8} f(\sum_l a_{lk}(H_l+Z_l)) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k} D_{H_1} f(\sum_l a_{lk}(H_l+Z_l)) & \cdots & \sum_k a_{8k} D_{H_8} f(\sum_l a_{lk}(H_l+Z_l)) \end{pmatrix}, \quad (4.43)$$

$$D_Z\Psi_r(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} k_r & & 0 \\ & \ddots & \\ 0 & & k_r \end{pmatrix} -\alpha \begin{pmatrix} \sum_k a_{1k} D_{Z_1} f(\sum_l a_{lk}(H_l+Z_l)) & \cdots & \sum_k a_{1k} D_{Z_8} f(\sum_l a_{lk}(H_l+Z_l)) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k} D_{Z_1} f(\sum_l a_{lk}(H_l+Z_l)) & \cdots & \sum_k a_{8k} D_{Z_8} f(\sum_l a_{lk}(H_l+Z_l)) \end{pmatrix}. \quad (4.44)$$

For simplicity, we denote the (i,j)-th element of the matrix $D_H\Psi_f(\mathbf{H},\mathbf{Z})$, $D_Z\Psi_f(\mathbf{H},\mathbf{Z})$, $D_H\Psi_r(\mathbf{H},\mathbf{Z})$, $D_Z\Psi_r(\mathbf{H},\mathbf{Z})$ as $D_{H_i}\Psi_{f_j}$, $D_{Z_i}\Psi_{f_j}$, $D_{H_i}\Psi_{r_j}$, $D_{Z_i}\Psi_{r_j}$, respectively.

From the fact (4.32) and (4.33), The Jacobian $D\Psi(\mathbf{H},\mathbf{Z})$ can be decomposed into the following, since the perturbation inside of (11335577) never turns outside and vice versa.

$$D\Psi(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} D_{\mathbf{H}_{in}}\Psi_{f_{in}} & 0 & D_{\mathbf{Z}_{in}}\Psi_{f_{in}} & 0 \\ 0 & D_{\mathbf{H}_{out}}\Psi_{f_{out}} & 0 & D_{\mathbf{Z}_{out}}\Psi_{f_{out}} \\ D_{\mathbf{H}_{in}}\Psi_{r_{in}} & 0 & D_{\mathbf{Z}_{in}}\Psi_{r_{in}} & 0 \\ 0 & D_{\mathbf{H}_{out}}\Psi_{r_{out}} & 0 & D_{\mathbf{Z}_{out}}\Psi_{r_{out}} \end{pmatrix}, \quad (4.45)$$

where

$$D_{\mathbf{H}_{in}}\Psi_{f_{in}} = \begin{pmatrix} D_{H_1}\Psi_{f_1} & \cdots & D_{H_4}\Psi_{f_1} \\ \vdots & \ddots & \vdots \\ D_{H_1}\Psi_{f_4} & \cdots & D_{H_4}\Psi_{f_4} \end{pmatrix}, \tag{4.46}$$

$$D_{\mathbf{Z}_{in}}\Psi_{f_{in}} = \begin{pmatrix} D_{Z_1}\Psi_{f_1} & \cdots & D_{Z_4}\Psi_{f_1} \\ \vdots & \ddots & \vdots \\ D_{Z_1}\Psi_{f_4} & \cdots & D_{Z_4}\Psi_{f_4} \end{pmatrix}, \tag{4.47}$$

$$D_{\mathbf{H}_{out}}\Psi_{f_{out}} = \begin{pmatrix} D_{H_5}\Psi_{f_5} & \cdots & D_{H_8}\Psi_{f_5} \\ \vdots & \ddots & \vdots \\ D_{H_5}\Psi_{f_8} & \cdots & D_{H_8}\Psi_{f_8} \end{pmatrix}, \tag{4.48}$$

$$D_{\mathbf{Z}_{out}}\Psi_{f_{out}} = \begin{pmatrix} D_{Z_5}\Psi_{f_5} & \cdots & D_{Z_8}\Psi_{f_5} \\ \vdots & \ddots & \vdots \\ D_{Z_5}\Psi_{f_8} & \cdots & D_{Z_8}\Psi_{f_8} \end{pmatrix}, \tag{4.49}$$

$$D_{\mathbf{H}_{in}}\Psi_{r_{in}} = \begin{pmatrix} D_{H_1}\Psi_{r_1} & \cdots & D_{H_4}\Psi_{r_1} \\ \vdots & \ddots & \vdots \\ D_{H_1}\Psi_{r_4} & \cdots & D_{H_4}\Psi_{r_4} \end{pmatrix}, \tag{4.50}$$

$$D_{\mathbf{Z}_{in}}\Psi_{r_{in}} = \begin{pmatrix} D_{Z_1}\Psi_{r_1} & \cdots & D_{Z_4}\Psi_{r_1} \\ \vdots & \ddots & \vdots \\ D_{Z_1}\Psi_{r_4} & \cdots & D_{Z_4}\Psi_{r_4} \end{pmatrix}, \tag{4.51}$$

$$D_{\mathbf{H}_{out}}\Psi_{r_{out}} = \begin{pmatrix} D_{H_5}\Psi_{r_5} & \cdots & D_{H_8}\Psi_{r_5} \\ \vdots & \ddots & \vdots \\ D_{H_5}\Psi_{r_8} & \cdots & D_{H_8}\Psi_{r_8} \end{pmatrix}, \tag{4.52}$$

$$D_{\mathbf{Z}_{out}}\Psi_{r_{out}} = \begin{pmatrix} D_{Z_5}\Psi_{r_5} & \cdots & D_{Z_8}\Psi_{r_5} \\ \vdots & \ddots & \vdots \\ D_{Z_5}\Psi_{r_8} & \cdots & D_{Z_8}\Psi_{r_8} \end{pmatrix}. \tag{4.53}$$

Therefore, the Lyapunov spectrum to the direction inside and outside of (11335577) can simply be calculated from the corresponding non-zero Jacobian elements. Let

$$D_{in}\Psi_{in} = \begin{pmatrix} D_{\mathbf{H}_{in}}\Psi_{f_{in}} & D_{\mathbf{Z}_{in}}\Psi_{f_{in}} \\ D_{\mathbf{H}_{in}}\Psi_{r_{in}} & D_{\mathbf{Z}_{in}}\Psi_{r_{in}} \end{pmatrix}, \tag{4.54}$$

and

$$D_{out}\Psi_{out} = \begin{pmatrix} D_{\mathbf{H}_{out}}\Psi_{f_{out}} & D_{\mathbf{Z}_{out}}\Psi_{f_{out}} \\ D_{\mathbf{H}_{out}}\Psi_{r_{out}} & D_{\mathbf{Z}_{out}}\Psi_{r_{out}} \end{pmatrix}. \tag{4.55}$$

Then the Lyapunov spectrum to the directions inside of (11335577) and outside of (11335577) can be obtained separately from the time series of $D_{in}\Psi_{in}$ and $D_{out}\Psi_{out}$, with a classical method such as QR decomposition.

In the same way, we can also decompose the Lyapunov spectrum into the directions inside of (11115555), outside of (11115555) in (11335577), and outside of (11335577). To avoid redundancy, we only give here the definition of the coordinates $(\mathbf{H}', \mathbf{Z}') = (H'_1, \cdots, H'_8, Z'_1, \cdots, Z'_8)^T$ which separate the elements of Jacobian $D\Psi(\mathbf{H}', \mathbf{Z}')$ according to the hierarchy.

$$(\mathbf{H}', \mathbf{Z}') = A'A(\eta, \zeta), \tag{4.56}$$

where

$$A' = (a'_{ij}) \tag{4.57}$$

$$= \begin{pmatrix} cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 \\ -cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -cos(\pi/4) & sin(\pi/4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.58}$$

In this case, $(H'_1, H'_2, Z'_1, Z'_2)^T$ are the coordinates inside of (11115555), $(H'_3, H'_4, Z'_3, Z'_4)^T$ are outside of (11115555), and $(H'_5, \cdots, H'_8, Z'_5, \cdots, Z'_8)^T$ are outside of (11335577). The last coordinates $(H'_5, \cdots, H'_8, Z'_5, \cdots, Z'_8)^T$ coincide with $(H_5, \cdots, H_8, Z_5, \cdots, Z_8)^T$.

The result of calculation of the maximum Lyapunov exponents in the decomposed directions are shown in Figs. 4.6 and 4.7. This enable us to verify the existence of chaos in each hierarchy of the invariant subspace. To investigate the possible role of chaos in this model, chaotic itinerancy between patterns A⇌B and C⇌D should have at least positive maximum Lyapunov exponent inside of (11335577).

4.2.6 Temporal Dynamics of Chaotic Itinerancy State

Among the possible regions of chaotic itinerancy state, we investigate the temporal dynamics on a specific refractory parameters, $k_r = 0.4$ and $\alpha = 5.0$. In this parameters, there exists chaotic attractive set inside of (11115555) and (11333311) including the patterns A,B and C,D, respectively (Fig.4.8), and the maximum Lyapunov exponents to the direction inside and outside of the (11115555) and (11333311) are positive (Fig.4.6). The transversal instability of (11115555) and (11333311) to the direction inside of (11335577) implies these subspaces went through the blowout bifurcation. The temporal dynamics of the maximum moment Lyapunov exponents inside and outside of (11115555) and (11333311) are also dominantly positive (Fig.4.10), and the orbits inside of (11335577) fall neither inside of (11115555) nor (11333311) at least for a quite long period (We tested the case until $t = 1000,000,000$). The orbits starting from almost all points of (12345678) are attracted inside of (11335577), which reflect the transversal stability of (11335577) (Fig.4.7). The orbits inside of (11335577) irregularly visit the patterns A,B,C,D, including partially preserved temporal sequences of two-periodic cycles A⇌B and C⇌D (Fig.4.8), and the maximum Lyapunov exponent and moment Lyapunov exponent are positive (Figs.4.7 and 4.9). This support the existence of chaos in term of the positive maximum Lyapunov exponent in this chaotic itinerancy state.

The loss of stability to the transversal directions of (11115555) and (11333311) can be considered as the blowout bifurcation. It has been widely reported in continuous time systems that the blowout bifurcations of plural attractors associate the riddled structure of the attractors' basins [29][30]. In this CNN, there exist two attractors A⇌B and C⇌D that are destabilized to their transversal directions. To reveal whether there exists similar phenomenon in the chaotic itinerancy state, we investigated the basin structure of CNN inside of (11335577). Since the dynamics seems to settle on neither of the patterns A⇌B nor C⇌D, these attractors are considered to have gone through the transversal crisis, and became the attractor ruins with attractive basins of positive measure. We judged the basins according to the first retrieval of the patterns A,B or C,D, with respect to the values of $h_i(t)$ $(1 \leq i \leq 8)$.

The result shows a fractal boundary structure of two basins (Fig.4.11). The spatially complex mixture strongly supports the occurring mechanism of the intermittent transition between two attractor ruins. Due to the fractal boundary structure and orbital instability, each orbit is expected to show a unique visiting sequence order and residence time to the attractor ruins, which becomes distinguishable from others in a short time period. Note it is of further question whether the blowout bifurcation associated with riddled basins is the general occurring mechanism of chaotic itinerancy.

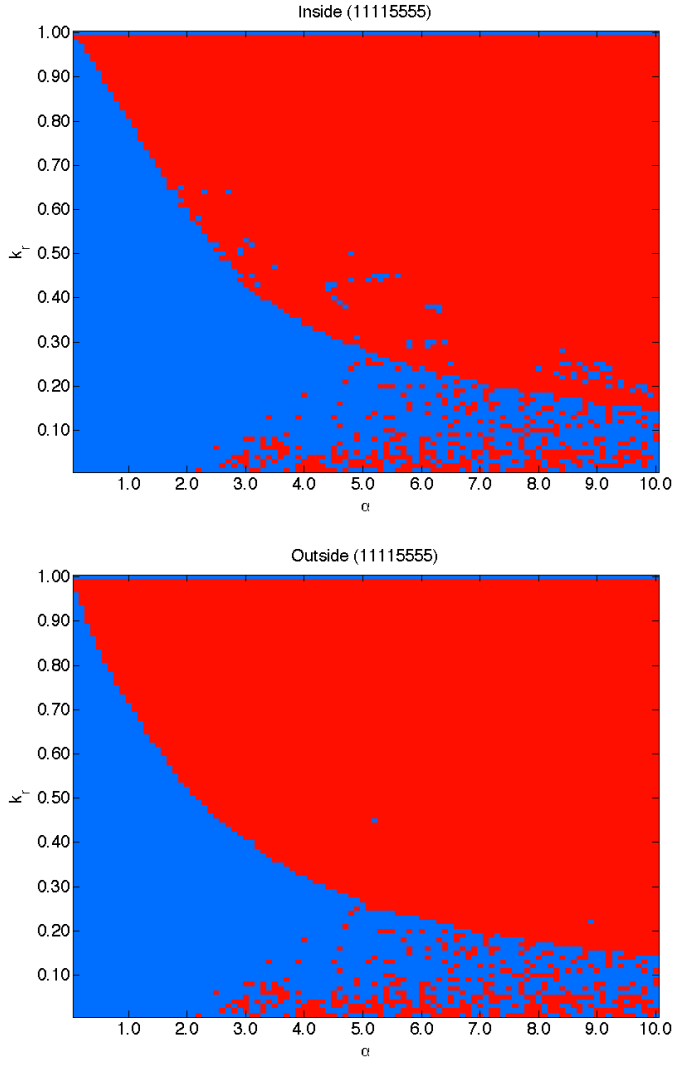


Figure 4.6: **Positive/Negative sign of the maximum Lyapunov exponents in (11115555).** Top: Directions inside of (11115555). Bottom: Outside of (11115555). Vertical axis: k_r. Horizontal axis: α. Initial conditions were taken randomly inside of (11115555). The blue region is where the maximum Lyapunov exponent is negative, while the red region is positive. After cutting 5000 transient, 1000 steps were used for calculation at each k_r and α. The values of k_r and α were taken for each 0.01 and 0.1 step, respectively.

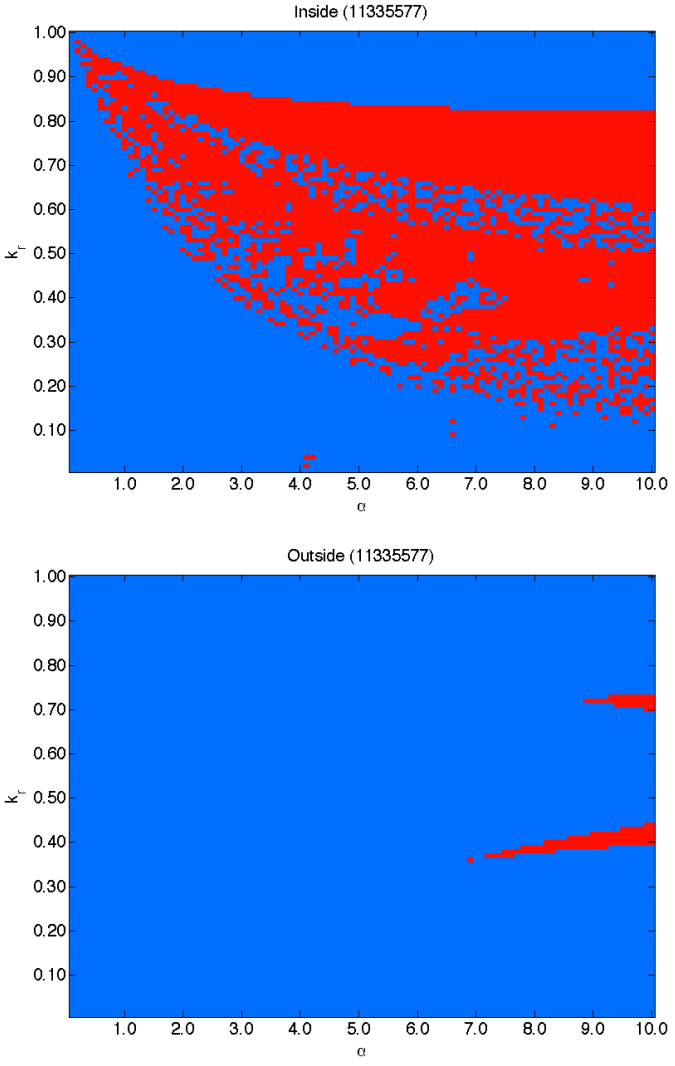


Figure 4.7: **Positive/Negative sign of the maximum Lyapunov exponents in (11335577).** Top: Inside of (11335577). Bottom: Outside of (11335577). Vertical axis: k_r. Horizontal axis: α. Initial conditions were taken randomly inside of (11335577). The blue region is where the maximum Lyapunov exponent is negative, while the red region is positive. After cutting 5000 transient, 1000 steps were used for calculation at each k_r and α. The values of k_r and α were taken for each 0.01 and 0.1 step, respectively.

Similar situation is also investigated in other stochastic and dynamical system, with the interest of defining *"emergence"* with respect to the temporal development of probability distribution of the system variables, namely space-time phase [31] [32]. R.S. Mackay defined the emergence as the existence of plural states of convergence for space-time phase, with weak/strong distinction. The converged space-time phase corresponds to the SRB measure on an attractor with measure positive attractive basin in dynamical system. *"Weak emergence"* is defined as due to the trivial topological reason of the system, such as the existence of simple separatrix between plural attractors in dynamical system. The coexistence of stable attractors in CNN at low refractoriness parameters region corresponds to such case. While *"strong emergence"* is referred to as the sensitivity of converged space-time phase to the initial condition, where tiny perturbation at the beginning can lead to different phase. This is similar to the riddled basin situation at the bifurcation point in CNN. Furthermore, Mackay mentioned to a more complex class of dynamics where space-time phase never settles down (*e.g.* biological evolution, gravitational systems, aggregation and coarsening models). In dynamical system counterpart, chaotic itinerancy in CNN is an example of such class of complex system which goes beyond the definition of strong emergence.

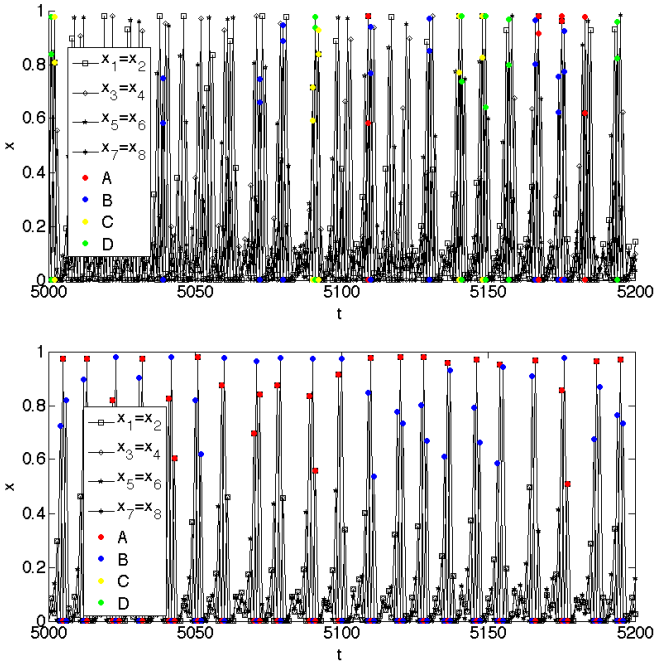


Figure 4.8: **Temporal dynamics of CNN.** Top: Initial condition inside of (11335577). Bottom: Initial condition inside of (11115555). Horizontal axis: Time step t. Vertical axis: CNN output $\mathbf{x}(t)$. The color dots representing the patterns A,B,C,D are superimposed to the value of $\mathbf{x}(t)$ when the network retrieved them. $k_r = 0.4$, $\alpha = 5.0$.

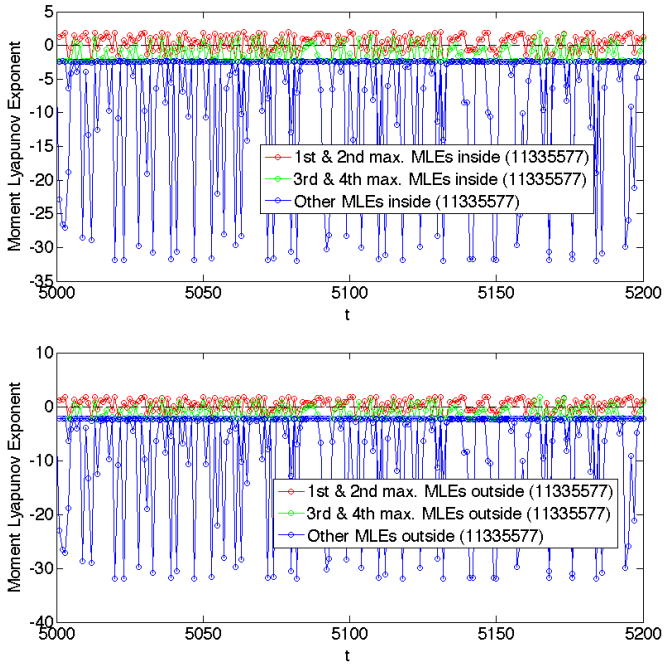


Figure 4.9: **Temporal dynamics of moment Lyapunov exponents with initial condition inside of (11335577).** Top: Moment Lyapunov exponents to the direction inside of (11335577). Bottom: Moment Lyapunov exponents to the direction outside of (11335577). Horizontal axis: Time step t. Vertical axis: Moment Lyapunov exponent. $k_r = 0.4$, $\alpha = 5.0$.

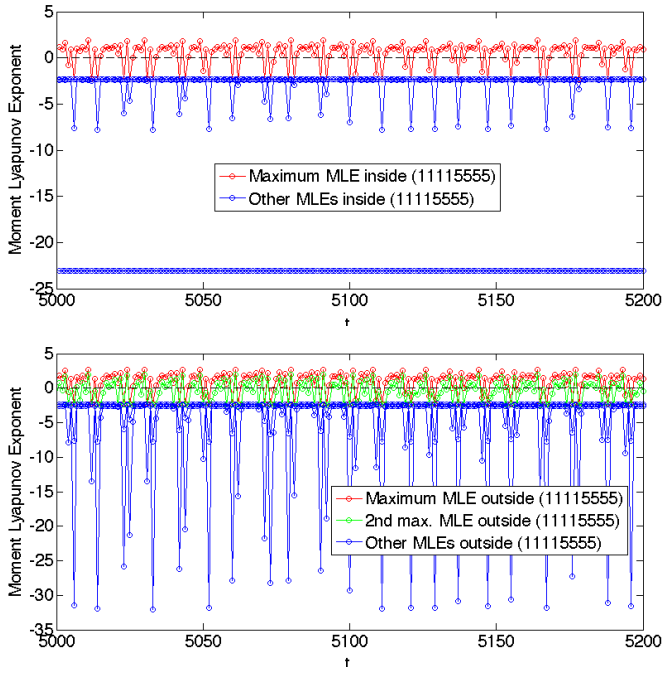


Figure 4.10: **Temporal dynamics of moment Lyapunov exponents with initial condition inside of (11115555).** Top: Moment Lyapunov exponents to the direction inside of (11115555). Bottom: Moment Lyapunov exponents to the direction outside of (11115555). Horizontal axis: Time step t. Vertical axis: Moment Lyapunov exponent. $k_r = 0.4$, $\alpha = 5.0$.

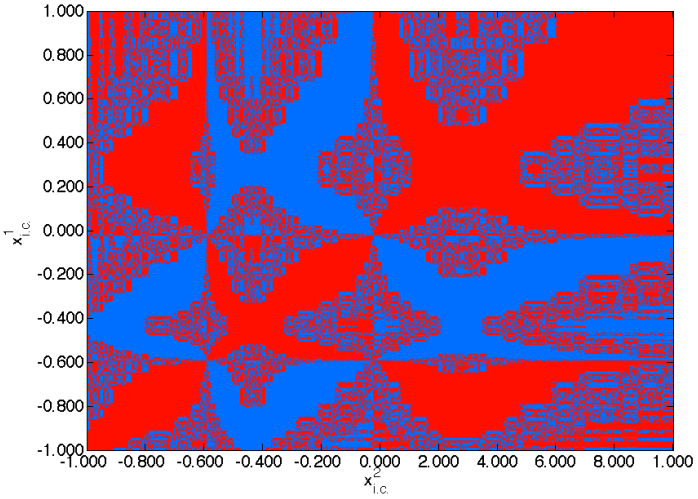


Figure 4.11: **Basins structure of CNN.** Horizontal axis: Value of $x^2_{i.c.}$. Vertical axis: Value of $x^1_{i.c.}$. Initial condition were taken inside (11335577) and close to (11331133) as $(\eta_1, \cdots, \eta_8)^T = (\zeta_1, \cdots, \zeta_8)^T = (x^1_{i.c.} + \delta, x^1_{i.c.} + \delta, x^2_{i.c.} + \delta, x^2_{i.c.} + \delta, x^1_{i.c.} - \delta, x^1_{i.c.} - \delta, x^2_{i.c.} - \delta, x^2_{i.c.} - \delta)^T$ with $\delta = 1.0e-6$. The red region is where the orbits reach to the neighbor of the patterns A$\rightleftharpoons$B first, while the blue region reach to the neighbor of the patterns C$\rightleftharpoons$D first. $x^1_{i.c.}$ and $x^2_{i.c.}$ were taken for each 0.002 step with parameters $k_r = 0.4$ and $\alpha = 5.0$.

4.3 Synthetic Modeling of Autonomous Learning

Based on the chaotic itinerancy state we have investigated in the above section with refractoriness parameters $k_r = 0.4$ and $\alpha = 5.0$, we add and simulate the effect of autonomous learning. We chose two basic ways of autonomous learning that are expected to exist widely and generally in neuronal activity: Hebbian learning and the spike-timing-dependent plasticity (STDP) rule [33] [34]. Of course, actual learning dynamics of *in situ* neurons have their own specificity and variation, and cannot be reduced into simple rules with a small number of parameters. What we investigate here is the possible role of mathematically simplified form of autonomous learning during chaotic itinerancy. This is not necessary the modeling of experimentally observed neural dynamics, but the combination of prototypical learning rules with a biological neuron model expressing candidate dynamics of cortical transitory activity, which is accessible to mathematical analysis. In this sense, we call it as synthetic modeling. In a wide sense, synthetic modeling is a part of constructive modeling in complex system sciences, but rather emphasizes on the generic and qualitative property which appears in the combination of candidate models, than the reconstruction of some observed phenomenon with quantitative reproducibility.

4.3.1 Autonomous Learning with Hebb's Learning Rule

Hebbian learning is the most classical rule to explain the self-organization of neural network, and is based on the "cells that fire together, wire together" principle. We implemented this rule to CNN to reinforce the connection matrix $W = (w_{ij})$ $(1 \leq i, j \leq 8)$ depending on the dynamics itself of the network. The additional Hebbian learning in CNN was defined as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \epsilon(2x_i(t) - 1)(2x_j(t-1) - 1), \tag{4.59}$$

where t is the time step of CNN and ϵ is the learning coefficient. We chose $\epsilon = 0.001$ for the following simulation.

4.3.2 Autonomous Learning with STDP Rule

The STDP rule is known to modify the synaptic conductance depending asymmetrically on the difference of pre-synaptic and post-synaptic firing time. We defined the discrete time version of STDP rule as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t), \tag{4.60}$$

$$\Delta w_{ij}(t) = g_{ij}(t) - g_{ji}(t), \tag{4.61}$$

$$g_{ij}(t) = A \sum_{d=1}^{d_f} k_{STDP}^{d-1} AND(x_i(t), x_j(t-d)), \tag{4.62}$$

where A is the norm of learning coefficient, and k_{STDP} controls the exponential decrease of learning effect with respect to the spike timing difference d. We use $A = 1$ and $k_{STDP} = 0.1$ for the following simulation. To implement the 'nearest neighbor model' of STDP learning between single pair of neuron spikes, d_f is defined at each time step and neuron in relation to the past neuron activity, so that

$$x_j(t-d) < 0.5 \ (1 \leq d < d_f), \tag{4.63}$$

$$x_j(t-d_f) \geq 0.5. \tag{4.64}$$

hold. In actual learning dynamics of CNN with STDP rule, neuron spikes are generally sparse and past influence quickly decay in exponential. For the given parameters, it is approximately equal to set $d_f = \infty$. The $AND(\cdot, \cdot)$ is the extended Boolean expression defined as follows:

$$AND(x_i(t), x_j(t-d)) = h_i(t)h_j(t-d)x_i(t)x_j(t-d). \tag{4.65}$$

This means the function returns the value $x_i(t)x_j(t-d)$ only when there exist spikes $x_i(t) \geq 0.5$ and $x_j(t-d) \geq 0.5$ of interval d. The defined STDP rule is depicted in Fig. 4.12.

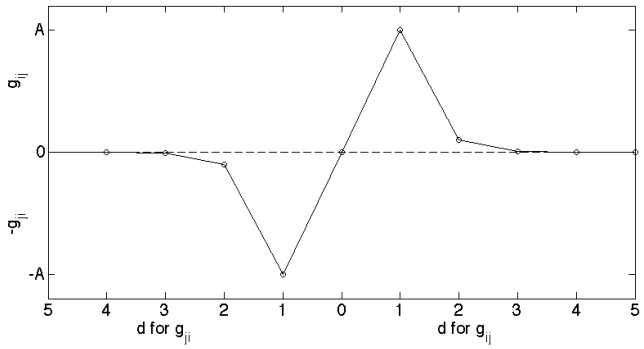


Figure 4.12: **STDP rule for discrete time CNN.** Horizontal axis: Spike time interval d. Vertical axis: Modification to w_{ij} defined in equation (4.62). $k_{STDP} = 0.1$.

4.4 Simulation Result and Discussion

The simulation is performed by adding Hebbian or STDP rule on the chaotic itinerancy state with the parameters $k_r = 0.4$ and $\alpha = 5.0$, over an empirical range of the learning time steps. After the learning, the refractoriness parameters were set to $k_r = \alpha = 0$ to investigate the memory structure. Note that such transition between chaotic wandering state and ordered state is also reported in EEG data [17]. We summarize here the observed modifications of the dynamics.

4.4.1 Periodicity of Emergent Attractors

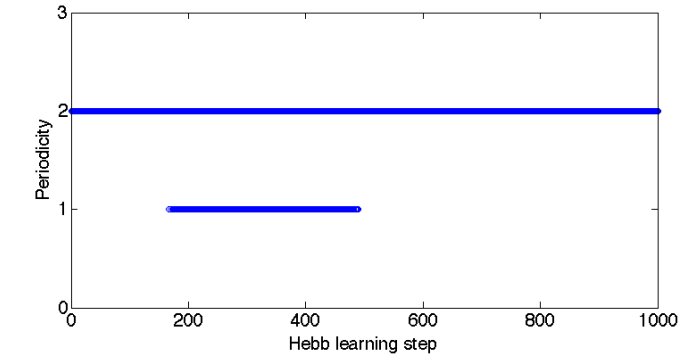
Both Hebbian learning and STDP rules brought the network novel attractors, based on the patterns A,B,C,D and transitional chaotic dynamics. We investigated the periodicity of the emergent attractors, which is shown in Fig. 4.13 Top. Hebbian learning did not change the periodicity and the modified attractors are always two-periodic, except the divergent case where the only attractor is the fix point $\mathbf{x} = (1,1,1,1,1,1,1,1)^T$. As for STDP learning, the periodicity of newly emerged attractors varies in a quite wide range (Fig. 4.13 Middle). There exist orbits even more than ten thousand period, which strongly implies the synthesis of strange attractors. Almost all periodicity between 1 to 100 are observed. The histogram of the obtained periodicity is shown in Fig. 4.13 Bottom.

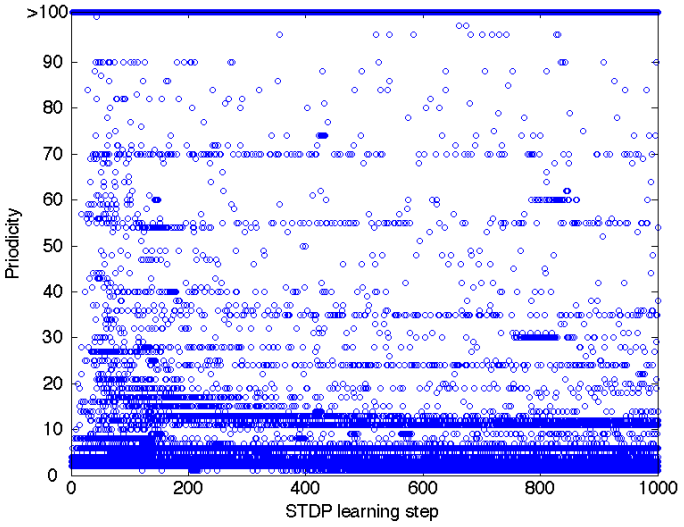
Such difference of results between Hebbian and STDP rules can be considered to relate the temporal symmetry of the learning rules. STDP rule has the temporal asymmetry which is not expressed in Hebbian rule. The effect of the past inputs with more than one time step interval may also bring variation to periodicity.

4.4.2 Spatial Configuration of Emergent Attractors

The two-periodic patterns emerged from Hebbian learning consists of two kinds; the former attractor A⇌B, C⇌D or the novel patterns which are different from any of A, B,C,D. The cases of A⇌B and C⇌D are investigated in the following section. We examine here the spatial configuration of the novel patterns, which are depicted in Fig. 4.14.

The newly emerged patterns are also limited to two kinds, and they are both situated in the equidistance points from A⇌B and C⇌D. Furthermore, these patterns are situated in the invariant subspace (11335577), which is the minimum union of the (11115555) and (11333311) where former attractors were situated. This geometry implies that the novel patterns are in a sense the integration of the attractors A⇌B and C⇌D in the smallest subspace preserving their periodicity. This is interesting as an analogical process of dialectic, where the opposing two theses find the way





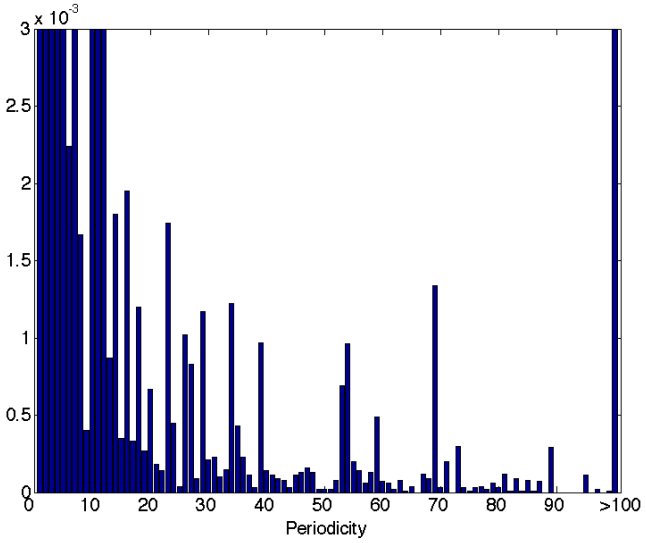


Figure 4.13: **Periodicity of attractor after autonomous learning.** Top: Hebbian learning. Horizontal axis: Time step of Hebb learning. Vertical axis: Periodicity of attractor after the learning. Middle: STDP learning. Horizontal axis: Time step of STDP learning. Vertical axis: Periodicity of attractor after the learning. Bottom: Histogram of periodicity after STDP learning. Horizontal axis: Periodicity of attractor after the learning. Vertical axis: Percentage (Values more than 3.0e-3 are cut off). The periodicity after the learning was judged by resetting $k_r = \alpha = 0$ and cutting 5000 transient. 100 trials are superimposed for each learning step. Initial conditions were taken randomly.

to be sublated not by logically solving the contradiction but by changing the way of asking itself. Here, we used the analogy of chaotic itinerancy state as the contemplation process comparing two symmetric patterns [35] [36], and the result of autonomous learning as the formation of an sublated idea. Although this is a quite simplified realization, this may lead to one of the adaptive computing principles with neural network, where neuronal plasticity plays essential role to produce internal complexity of the system dynamics to cope with external environment. This perspective is further discussed in later section, in relation to the formal logic in topological psychology. On the other hand, computing with dynamical systems such as associative memory generally utilizes the retrieval process to classify the input patterns according to memory structure, and does not consider the modification of the memory itself in interaction with the retrieval dynamics.

Although the possible emergent patterns are limited with Hebbian learning, its reproducibility requires a quite high numerical precision of initial condition down to the order of 1.0e-17 (in case of 350 learning steps). This may be a reflection of the observed riddled basins structure to the learning dynamics and orbital instability, since the global behaviors of the orbits differ depending quite sensitively on the initial conditions.

The emerged attractors from STDP rule are more suppressive, and in most cases express the output patterns $\mathbf{h}(t) = (0,0,0,0,0,0,0,0)^T$ and a few patterns with only a single pair of neurons firing inside (11335577) according to the varying periodicity, such as $\mathbf{h}(t) = (1,1,0,0,0,0,0,0)^T$. The sensitive dependence on initial conditions was also observed numerically in the same order as Hebbian learning.

4.4.3 Modification of Invariant Subspaces

Since the connection matrix W is modified by autonomous learning, the structure of the invariant subspaces reflecting its symmetry also changes temporally. In both Hebbian and STDP leaning, the remaining invariant subspaces in rigorous definition are only those in above hierarchy of (11335577) (Fig. 4.2 Top). This is due to the loss of the other synchronizations in the orbits. Since chaotic

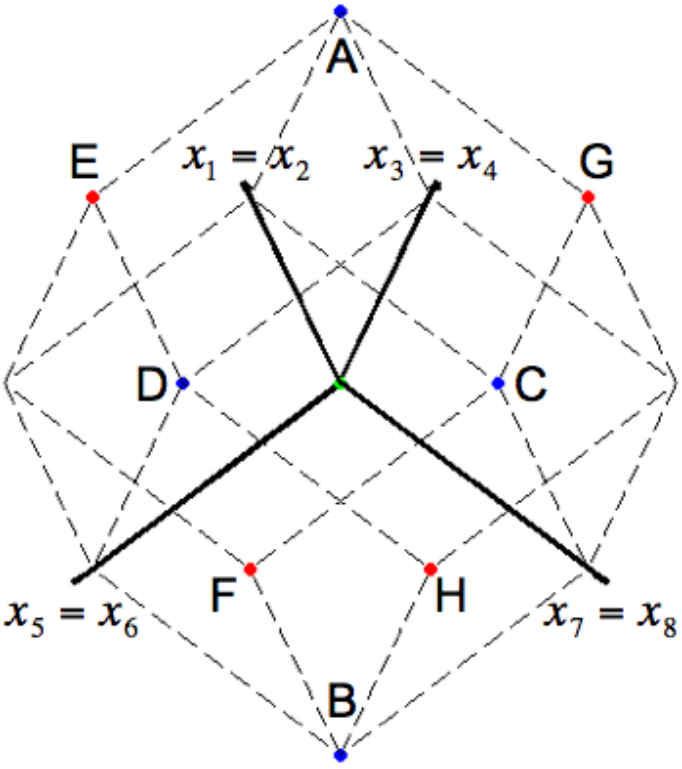


Figure 4.14: **Possible emergent attractors after Hebb learning.** The arrows are four-dimensional coordinates $(x_1 = x_2, x_3 = x_4, x_5 = x_6, x_7 = x_8)^T$ of the CNN output inside of (11335577). The blue points are memorized patterns A, B, C, D. The red points E, F, G, H are possible emergent two-periodic attractors, and are situated at the equidistance points from A⇌B and C⇌D. Only two-periodic patterns E⇌F and G⇌H are possible to emerge. The green point is the divergent case that corresponds to the fix point $(1, 1, 1, 1)^T$.

Table 4.2: **Oscillation of the number of invariant subspaces in case of the asymmetric convergence to one attractor C⇌D.** Invariant subspaces were calculated for each 100 time steps with the allowance of 0.01. $k_r = 0.4$, $\alpha = 4.0$. Only the relative differences with respect to the past time step are listed.

Number of Invariant Subspaces	Time Steps of Hebbian Learning
76	0
49	100
34	200
43	1100
25	10000+

itinerancy takes place in (11335577), four pairs of neurons are always synchronized regardless of its irregularity. Therefore, although all values of connection matrix are being modified, the symmetry of the connection matrix supporting the above hierarchy of (11335577) is completely maintained. While other invariant subspaces are destroyed by the chaotic irregularity.

Since (11115555) and (11333311) are situated below (11335577), the attractors A⇌B and C⇌D are destabilized along with the destruction of the invariant subspaces (Fig. 4.2 Bottom). The invariant subspaces which are neither above nor below (11335577) also possess attractor inside (11111111). Although these are measure zero subspaces with respect to the whole state space, their dimensions vary from one to seven. These can also be considered as a part of memory structure. As the invariant subspaces are trimmed leaving only the structure used in retrieval process, the observed modification can be interpreted as the simplification or refinement of the memory. Hence, the selective destruction and conservation of invariant subspaces are impossible to realize with stochastic noise. Such utility of chaos lies in its completely deterministic property.

Although the rigorous invariant subspaces other than above (11335577) are instantaneously destroyed by the autonomous learning, approximate invariant subspaces shows certain dynamics which reflects the nature of dynamics. We consider the case of Hebbian learning in which the competition between two attractors A⇌B and C⇌D takes place. We define the approximate invariant subspaces by admitting certain tolerance to the condition (4.11), and investigate the asymmetric convergence to one attractor.

Tab. 4.2 shows an example with $k_r = 0.4$ and $\alpha = 4.0$. In this case, the number of the approximate invariant subspaces is oscillating with respect to the time steps of Hebbian learning. The dynamics finally settled to the pattern C⇌D after more than 10000 learning steps.This dynamics implies that certain invariant subspaces are repeatedly destroyed and recovering during the learning. The prevalence of certain invariant subspaces are assumed to relate the localization of orbit between two attractors. The oscillation of the number of invariant subspaces continues until one basin structure becomes completely dominant and fixed. This can be analogically interpreted as the decision-making process, where one is supposed to choose only one side from the two possibilities. The chaotic itinerancy state is again the analogy of contemplation process comparing two patterns. The final settlement to the pattern C⇌D that drove out the pattern A⇌B corresponds to the decision-making.

4.4.4 Change of Residence Time Distribution

The defined autonomous learnings always follow the principle that the more frequently causal temporal structures appear, the more they are reinforced. The way of visiting each attractor ruin during the learning gives us the information how the self-organization of novel attractors proceeds. We investigated the residence time distributions for the typical emergent cases of the novel two-periodic attractor E⇌F (see Fig. 4.14) in Hebbian learning and three-periodic attractor in STDP learning in Fig.4.15.

Residence time distribution of the chaotic itinerancy state with $k_r = 0.4$ and $\alpha = 5.0$ shows unimodal pattern that may be approximated with gamma-distribution. Additional Hebbian learning

triggers the right shift of the distribution and the overall increase of the visiting rate. This means that the more Hebbian learning proceeds, the more frequently and the longer the learning orbit visits on the patterns A⇌B and C⇌D. This can be considered as the self-organized reinforcement of attractiveness to each attractor ruin, and the learning was revealed to augment the neural correlations of memorized patterns during chaotic itinerancy. The increase of correlations may also be supported by the learning of transitional chaotic state with low neural activity. The continuous sequences of low outputs also increase the connection weight in the defined setting. Though, the simple increase of the Hebbian learning step does not lead the network to the divergent pattern.

As for STDP learning, generally the learning orbits become periodic regardless of the refractoriness, and the residence time distribution shrink to the residence time 1. The dynamics do not stay more than 1 step on each pattern during learning, including no residence time on the four patterns. STDP learning on attractor ruins seems to be cancelled out with the periodicity two of the patterns A⇌B and C⇌D, since the reinforced connections at A→B and C→D are weakened by the successive B→A and D→C. On the other hand, the exponential duration of STDP rule seems also to reinforce the firing patterns other than A,B,C,D in transitional chaotic state which are generally sparse. Such isolated patterns are weakened in Hebbian learning. In chaotic dynamics, it is known that there exist countably infinite sets of unstable periodic orbits with arbitrary periodicity. It is of interest whether the STDP rule brings rich variation of periodicity by stabilizing long periodicities from chaotic orbit.

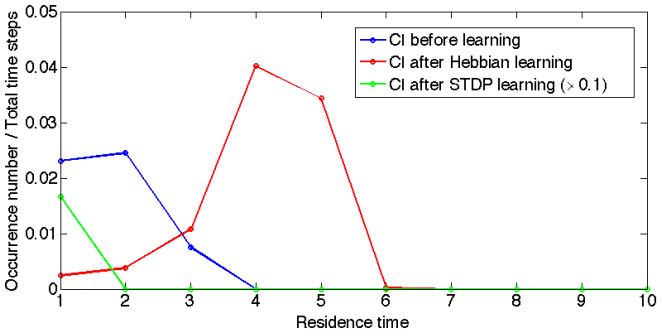


Figure 4.15: **Residence time distribution on memorized patterns before and after autonomous learning during chaotic itinerancy (CI).** Horizontal axis: Residence time step on patterns A, B, C, D. Vertical axis: Occurrence number of each residence time divided by the total time step 10000. The blue line is calculated from the dynamics without any modification to connection matrix W, while the red line is after adding 350 steps of Hebbian learning. The green line is after adding 350 steps of STDP learning. The distribution of STDP learning is multiplied by 0.1. Initial condition was chosen to emerge the patterns E⇌F (see Fig. 4.14) for Hebbian learning, and three-periodic patterns $\mathbf{h}(t) = (1,0,0,0)^T \to (0,0,0,0)^T \to (0,0,0,0)^T \cdots$ for STDP learning in (11335577). $k_r = 0.4, \alpha = 5.0$.

4.4.5 Change of System Decompositionability by Hebbian Learning

The effect of Hebbian learning during chaotic itinerancy was revealed to reinforce simultaneous neural correlation in this model. It is of interest which combinations of the neuronal synchrony arise. The augmentation of coincident firing patterns rate can be interpreted as the formation of modularity, which can lead to the functional differentiation. To investigate the emerged system modularity during Hebbian learning, we consider the time-averaged four-dimensional discrete distribution $P(y_1, y_2, y_3, y_4)$ of binary outputs $\mathbf{h}(t)$ in (11335577) as

$$P(y_1,y_2,y_3,y_4) = \frac{1}{T}\sum_{t=1}^{T}\delta(h_1(t)=h_2(t),y_1)\delta(h_3(t)=h_4(t),y_2) \cdot\delta(h_5(t)=h_6(t),y_3)\delta(h_7(t)=h_8(t),y_4). \quad (4.66)$$

Where $\delta(\cdot,\cdot)$ is the delta function, T is the temporal average window and $y_i(t) \in \{0,1\}$ $(1 \le i \le 4)$.

To measure the statistical modularity expressed as correlation among (y_1, y_2, y_3, y_4), we define the decomposed distribution $P^{dec}(y_1, y_2, y_3, y_4)$ as the product of two joint distributions $P(\mathbf{y}^1)$ and $P(\mathbf{y}^2)$ of the decomposed subsystems $\mathbf{y}^1$ and $\mathbf{y}^2$.

$$P^{dec}(y_1, y_2, y_3, y_4) = P(\mathbf{y}^1)P(\mathbf{y}^2), \tag{4.67}$$

where

$$\mathbf{y}^1 = (y_{i_1}, \cdots, y_{i_k}), \quad \{i_1, \cdots, i_k\} \subset \{1, \cdots, 4\}, \quad k \in \{1, 2, 3\}, \tag{4.68}$$

$$\mathbf{y}^2 = (y_{j_1}, \cdots, y_{i_{4-k}}), \quad \{i_1, \cdots, i_k\} \not\supset \{j_1, \cdots, j_{4-k}\} \subset \{1, \cdots, 4\}. \tag{4.69}$$

Note that the system decomposition is not limited to two parts, and there exist up to the four subsystems in (11335577) which can be easily generalized such as

$$\begin{aligned} P^{dec}(y_1, y_2, y_3, y_4) &= P(\mathbf{y}^1)P(\mathbf{y}^2)P(\mathbf{y}^3), && (4.70)\\ \mathbf{y}^1 &= (y_{i_1}, y_{i_2}), \quad \{i_1, i_2\} \subset \{1, \cdots, 4\}, && (4.71)\\ \mathbf{y}^2 &= (y_j), \quad \{i_1, i_2\} \not\ni \{j\} \in \{1, \cdots, 4\}, && (4.72)\\ \mathbf{y}^3 &= (y_k), \quad \{i_1, i_2, j\} \not\ni \{k\} \in \{1, \cdots, 4\}, && (4.73) \end{aligned}$$

for three subsystems, and

$$\begin{aligned} P^{dec}(y_1, y_2, y_3, y_4) &= P(\mathbf{y}^1)P(\mathbf{y}^2)P(\mathbf{y}^3)P(\mathbf{y}^4) \\ &= P(y_1)P(y_2)P(y_3)P(y_4), && (4.74) \end{aligned}$$

for four subsystems. In each decomposed distribution $P^{dec}(y_1, y_2, y_3, y_4)$, only correlations between the defined subsystems are set to be independent, and the correlations inside of each subsystem are preserved.

We measured the decompositionability of the system $P(y_1, y_2, y_3, y_4)$ into $P^{dec}(y_1, y_2, y_3, y_4)$ with Kullback-Leibler (KL) divergence $D[\cdot : \cdot]$ as follows.

$$\begin{aligned} & D[P(y_1, y_2, y_3, y_4) : P^{dec}(y_1, y_2, y_3, y_4)] \\ = & \sum_{y_1, y_2, y_3, y_4 \in \{0,1\}} P(y_1, y_2, y_3, y_4) \log \frac{P(y_1, y_2, y_3, y_4)}{P^{dec}(y_1, y_2, y_3, y_4)}. \end{aligned} \tag{4.75}$$

This means that if the value of $D[P(y_1, y_2, y_3, y_4) : P^{dec}(y_1, y_2, y_3, y_4)]$ is large, the system loses much information by the decomposition, therefore there exist high modularity on the nullified interactions. Note that the total system decompositionability $D[P : P^{dec} = P(y_1)P(y_2)P(y_3)P(y_4)]$ is referred to as a complexity measure in several studies [31] [37][38].

The calculation result of KL divergence for each system decomposition during chaotic itinerancy after 350 steps of Hebbian learning is shown in Figs. 4.16. The parameters were chosen to realize the emergence of novel two-periodic attractor E⇌F (see Fig. 4.14) after the learning. Before adding the learning, the chaotic itinerancy state showed modularity in $\mathbf{y} = (y_1, y_3)$ and $\mathbf{y} = (y_2, y_4)$ because cutting between them do not lose much information measured with KL divergence. As the Hebbian learning proceeded, the KL divergences strongly augmented, meaning the system increased the corresponding statistical coherence. After 350 steps of learning, chaotic itinerancy state showed three kinds of uniform KL divergence values proportional to the number of decomposed subsystems. Interestingly, the relation

$$\begin{aligned} & \frac{1}{4-1} D[P : P^{dec} = P(y_1)P(y_2)P(y_3)P(y_4)] \\ = & \frac{1}{3-1} D[P : P^{dec} = P(\mathbf{y}^1)P(\mathbf{y}^2)P(\mathbf{y}^3)] \\ = & \frac{1}{2-1} D[P : P^{dec} = P(\mathbf{y}^1)P(\mathbf{y}^2)], \end{aligned} \tag{4.76}$$

holds. This relation stands for the fact that the system loses the same amount of information when isolating any arbitrary subsystem. This means that the neuronal synchrony rate augmented in the way that the degrees of interactions became identical among all elements. In such distribution, there exists no easiest decomposition. In other word, the existed modularity of the dynamics are integrated in the total coherence of the system after the learning.

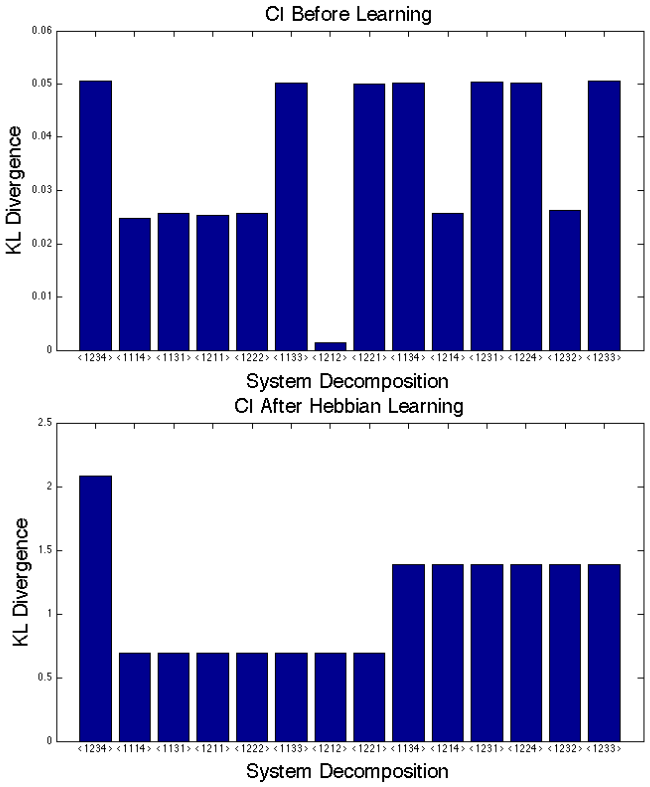


Figure 4.16: **KL divergence of the system decompositions.** Top: Chaotic itinerancy (CI) state before learning. Bottom: Chaotic itinerancy (CI) state after 350 steps of Hebbian learning with learning coefficient $\epsilon = 0.01$. Horizontal axis: $< \cdot \ \cdot \ \cdot \cdot >$ represent system decompositions in which correlations between different numbers of elements are set to be independent. *e.g.* $< 1212 >$ corresponds to $P^{dec}(y_1, y_2, y_3, y_4) = P(\mathbf{y}^1)P(\mathbf{y}^2)$ where $\mathbf{y}^1 = (y_1, y_3)$ and $\mathbf{y}^2 = (y_2, y_4)$. Vertical axis: KL divergence $D[P : P^{dec}]$. $k_r = 0.4$, $\alpha = 5.0$, $T = 10000$ were used for the calculation.

4.4.6 Computational Rationale of Autonomous Learning in Relation to Topological Psychology

We finally revisit the analogy with dialectic of the emerged attractors during the autonomous learning, in relation to a formal logical description of dialectic in topological psychology. We aim here to investigate the computational rationale of autonomous learning in view of realization of a system which incorporate dialectical information processing.

Usually dialectic describes the one's mental process which wander around contradictory experience or thoughts, meanwhile creating a solution which trivializes the cognitive dissonance to some extent. Such dynamical view on mental activity was also one of the connectionist's motivation concerning the versatility and generativity of cognitive system. Neural network model based on parallel distributed information processing was proposed analogous to the physiological structure of neural systems, with expectation to realize certain aspect of our mental activity missing in classicist's way of serial-processing symbolic computation. [39]

Indeed, several authors have proposed connectionist models of dialectical psychology with dynamical system, in both on-going mental process [40][41][42][43][44] and long-term developmental

process [45].

Our model with CNN also find common interest with this stream, but gives distinguished emphasis on chaotic dynamics with neurophysiological basis. The representation of memory structure is also particularly grounded on intermittent chaotic dynamics in CNN called chaotic itinerancy, in which the interior crisis of attractors leads to the dynamical transition between plural attractor ruins. The analogy of chaotic itinerancy with volatile aspect of mental process were assigned by I. Tsuda and insisted to be a core concept of hermeneutics and mathematical model of dynamical brain activity [21] [35] [36].

Acknowledging the background of connectionist's interest on dialectic, we go back to the relation between our model and topological psychology. One of the classical philosophical descriptions of dialectic with particular emphasis on the contradiction process was developed by celebrated philosopher G.W.F. Hegel [46]. It was further adapted in the context of Piagetian theory of cognitive development in dialectical psychology by K. Riegel [47]. Riegel's psychological interpretation of Hegel's dialectic are summarized by C.W. Hoffman as follows, namely Hegel-Riegel laws [48]:

1. The unity and struggle of opposites.
2. The transformation of quantitative into qualitative change.
3. The negation of the negation.

C.W. Hoffman expressed these principles with formal logic by means of set-theoretic operation with symmetric difference structure, and further relation to various mental activities are investigated in the name of topological psychology. We briefly introduce his framework for the following argument [48][49]. The symmetric difference \$ between two cognitive sets S and S' is defined using set-theoretic intersection ($\cap$) and union ($\cup$)as follows:

$$S\$S' = (S \cup S') \setminus (S \cap S'). \tag{4.77}$$

The symmetric difference and the complements $\neg(S \cup S')$ are equivalent to Sheffer stroke in logic, which is the NAND operation in Boolean algebra and the generator of other operators in ordinary first order logic such as AND, OR, NOT, and IMP [50]. Formal operation is thus realizable by the combination of the symmetric difference and NOT ($\neg$), AND ($\cap$) operation. For example, the OR ($\cup$) operation is derived as follows:

$$S \cup S' = (S\$S')\$(S \cap S'). \tag{4.78}$$

Hoffman insists that the symmetric difference represents the first two of the Hegel-Riegel laws: In the definition 4.77, the term $(S \cup S')$ is naturally the unity, from which the commonality $(S \cap S')$ is substituted as the struggle of opposites. The quantitative change between two sets $|S| - |S'|$ is transformed into qualitative one $S\$S'$ by set-theoretic definition of commonality $(S \cap S')$. Here, $|S|$ corresponds to a simple quantitative measure of the set S, such as the element number. The third law, "the negation of the negation" corresponds to a part of the complementary set of symmetric difference $-(S\$S')$. From

$$\neg(S\$S') = (S \cap S') \cup (\neg S \cap \neg S'), \tag{4.79}$$

the complementary set $\neg(S\$S')$ is the union between the Hegel's union of opposite $S \cap S'$, and the intersection of not S and not S', as the negation of the negation. The second term can be considered as the context within the universe of discourse. The commonality $S \cap S'$ is nothing but the synthesis through the struggle between the thesis $S \setminus (S \cap S')$ and anti-thesis $S' \setminus (S \cap S')$. The cognitive sets S and S' are not appropriate in itself to be called thesis and anti-thesis, since the synthesis $S \cap S'$ is supposed to contain a novel dimension which does not exist beforehand. The substitution of $S \cap S'$ in both theses represent the novelty of synthesis, which is at the same time a cognitive commonality between them. Fig. 4.17 shows the Venn diagram of the symmetric difference \$

in relation to ordinary set-theoretic operations. Dialectical pair with symmetric difference is an elemental unit of our symbolic conceptualization, which establishes semantic relation where the meaning is defined by contrastive reference to symmetric counterpart [47].

We utilize the introduced dialectical pair S and S' to formalize symbolically the results of autonomous learnings in CNN in a way compatible to set-theoretic operation. Chaotic itinerancy actually incorporates the dynamics on dialectical pair: The orbit intermittently transit between symmetric attractor ruins, which can be regarded as the symmetric sets without commonality. The presence of orbit in each attractor ruin is mutually exclusive. Therefore, the memory attractors A⇌B and C⇌D corresponds to the thesis T and anti thesis T' defined as a part of the cognitive sets S and S' as follows:

$$T = S \setminus (S \cap S'), \tag{4.80}$$

$$T' = S' \setminus (S \cap S'). \tag{4.81}$$

Where T is A⇌B and T' is C⇌D in binary output space of CNN.

The synthesis $S \cap S'$ corresponds to the emerged attractors such as E⇌F, G⇌H in Hebb learning, and various periodic attractors in STDP learning. Fig.4.18 shows the relation between thesis, antithesis, synthesis, and the underlying cognitive sets in case of Hebb learning. The commonality of symmetric cognitive sets $S \cap S'$ can be expressed as the OR operation of CNN binary output patterns. According to the simulation results, the cognitive sets can be directly defined from the geometrical composition of thesis and anti-thesis, but usually there is no analytical way to derive it: The synthesis is only possible in the OR product space of the memory attractors. In case of STDP learning, the situation becomes more complex since it includes further information of the change in periodicity. Empirical simulation can still incorporate statistically major results of emerged attractors into Boolean logic expression.

One would question what is the use of the autonomous learning, if the emerged results can be simply classified and expressed as Boolean logic. Actually our model does not allow simple symbolic classification of the results with respect to the parameter space: There exist high sensitivity on initial conditions as the property of chaotic dynamics. The group of the initial conditions giving the same result of autonomous learning is supposed to form a Cantor set. As shown in Fig.4.18, we can only represent the results of emerged attractors with parameters specified at the digit limit of computation. Theoretically, it may reach to real value precision. If one try to simulate the same system as CNN with autonomous learning only by a system composed of Boolean logic, the infinite length of the program is needed to perfectly follow the sensitivity to initial conditions. The characterization of autonomous learning with Boolean logic only shows us the computational rationale in classified results with specified parameters, and does not reduce the complexity of chaos.

This raises a deeper consideration on what has been severely argued between classicists and connectionists. With the presence of chaotic dynamics, both approaches encounter the same difficulty in different aspects. The connectionists emphasize biological realism. They try to imitate the structure of the system and expect to reconstruct the mechanism of the input-output relations. The difficulty lies in the observational resolution of a cognitive system, which is extremely complex in different scales. One can not incorporate hierarchical structures including chaos of molecular scale in a macroscopic model. The artificial intelligence, on the other hand, is not necessary attached to biological mechanisms but rather seeks for realizing "intelligent" behaviors in a more metaphysical level. They start from abstracting human mental activity, then try to construct an algorithmic system and implement it in artificial systems such as computer simulation and robots to test the degree of achievement [51].

The connectionist model including chaos such as CNN can not specify the control resolution of initial conditions with respect to the long-term prediction. No matter how small we divide the phase space, miniscule changes quickly propagate to the system size. Therefore, one can not specify a finite set of symbols which would accurately reproduce input-output patterns in a long run. In a more complex case such as the emergence of strange attractors in STDP learning, the condition requires much more resolution in shorter term.

Even without the autonomous learning, the approximation of chaotic itinerancy in CNN with time-series analysis model such as high-order Markov process encounters a similar difficulty from

chaos. Chaotic dynamics usually contain unstable periodic orbits of arbitrary periodicity, which can not be expressed as a deterministic finite state automaton.

Although the CNN is defined with a finite length of algorithm, the reconstruction of the generated sequences without the mechanism becomes impossible with a finite code due to the Cantor set structure of the phase space. The effort from the outside to capture the chaotic dynamics in an algorithmic way with a finite symbolic set is nullified by the infinite riddled structure of Cantor sets.

In other words, the perfect emulation of the autonomous learning in CNN is generally impossible, because the initial condition giving the same results is undecidable in computability theory. A set of natural numbers is called recursive, computable or decidable if there is an algorithm which terminates after a finite amount of time and correctly decides whether or not a given number belongs to the set. Since the Cantor set is defined on the infinite self-recurrence, such algorithm does not exist. Therefore, the emulation of CNN is undecidable in the same way as the halting problem of Turing machines.

The approach of artificial intelligence generally consider the input-output relations from the outside of the system, and propose the algorithm which would replace the black box, the inside of the system. Connectionist model including CNN try to reproduce the input-output patterns by reconstructing the internal mechanism of the system. With this, we suggest to call the top-down approach of artificial intelligence as the "*out-in emulation*", while the bottom-up connectionist model as the "*in-out reconstruction*", considering the difficulty of the same origin that both encounter in chaotic systems. Figs. 4.19 schematically show the difference of these conceptualizations. Besides the conventional top-down and bottom-up classification, we propose to consider the complexity of chaotic input-output relations, which originates from inside of the system.

Autonomous learning is difficult to control and takes much more time to calculate than simple symbolic operation. Though, it is profoundly grounded to the intrinsic mechanism of cognitive systems and seek for the autonomous dynamics of information processing which actually produce novelty in living system. Interest of connectionism in terms of parallel information processing also shares the same perspective.

It is of further interest whether the other basic Boolean operators such as AND and NOT is realizable with autonomous learning. Actually, the emerged various periodic attractors from STDP learning contain the AND patterns of the memorized attractors. The NOT sets of the synthesis E⇌F and G⇌H can also be interpreted as the NAND of the two cognitive sets in Fig. 4.18. Empirical listing of the learning rules and parameters with respect to the modification of memory structure is needed to realize the application in analog computing.

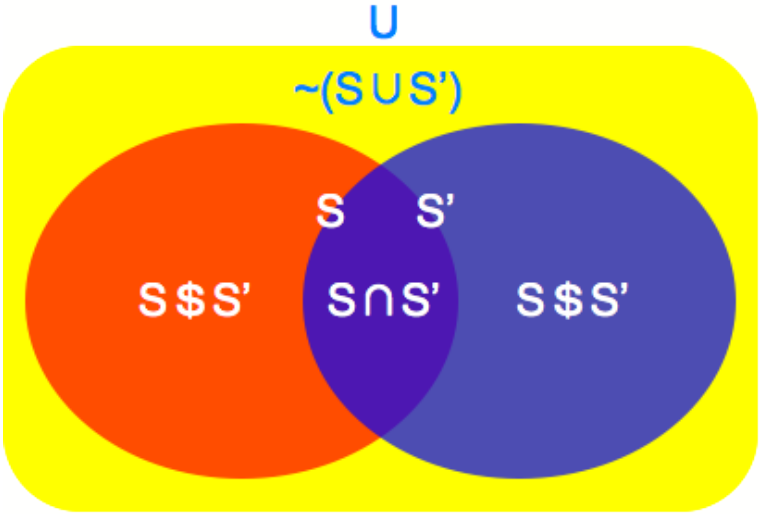


Figure 4.17: **Venn diagram of symmetric difference $.** The exclusive disjunction or XOR of two cognitive sets S and S' coincides with $S\$S'$, within the universe of possible cognitive sets U. Based on [48].

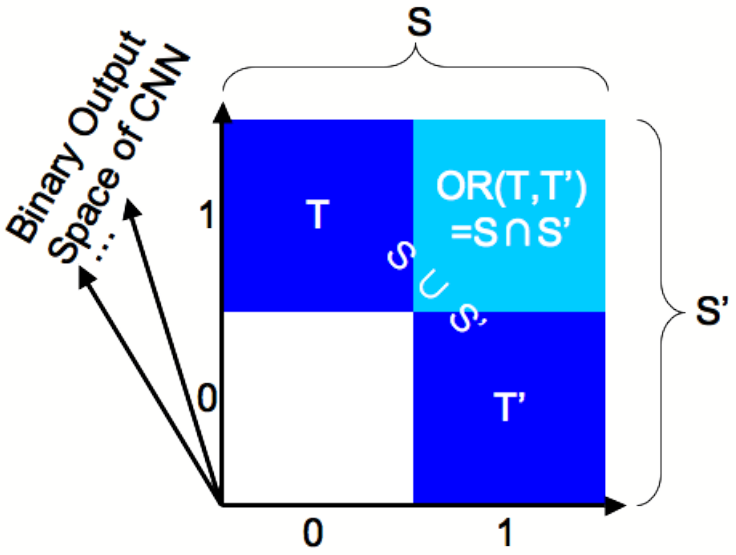


Figure 4.18: **Relation between cognitive sets S and S', including thesis T, antithesis T' respectively, and their OR product OR(T, T') in binary phase space of CNN outputs.** Results of Hebb learning.

4.5 Conclusion

We first analyzed the dynamics of CNN mainly in terms of the periodicity, the deviation rate from attractors, the wandering range, and the linear stability according to the hierarchical structure of invariant subspaces. The chaotic itinerancy state of CNN was revealed to occur with the blowout bifurcation of the attractors, and was associated with the riddled basins structure.

Next, we synthetically investigated in the simulation the possible constructive role of chaotic itinerancy state in interaction with the autonomous learning rules. Hebbian learning was shown to be able to converge the memorized attractors to form novel ones in their equidistance points, conserving the periodicity and augmenting neural correlation. STDP learning rather suppressed average neural activity but derived a rich variation of periodicity in the emerged attractors. The deterministic property of the system allowed both Hebbian and STDP learning to conserve the invariant subspaces superior than where the orbits are situated, which can be interpreted as the selective destruction and preservation of memory structure.

Further experimental verification is needed for the generalization of the results in more physiology-grounded situations.

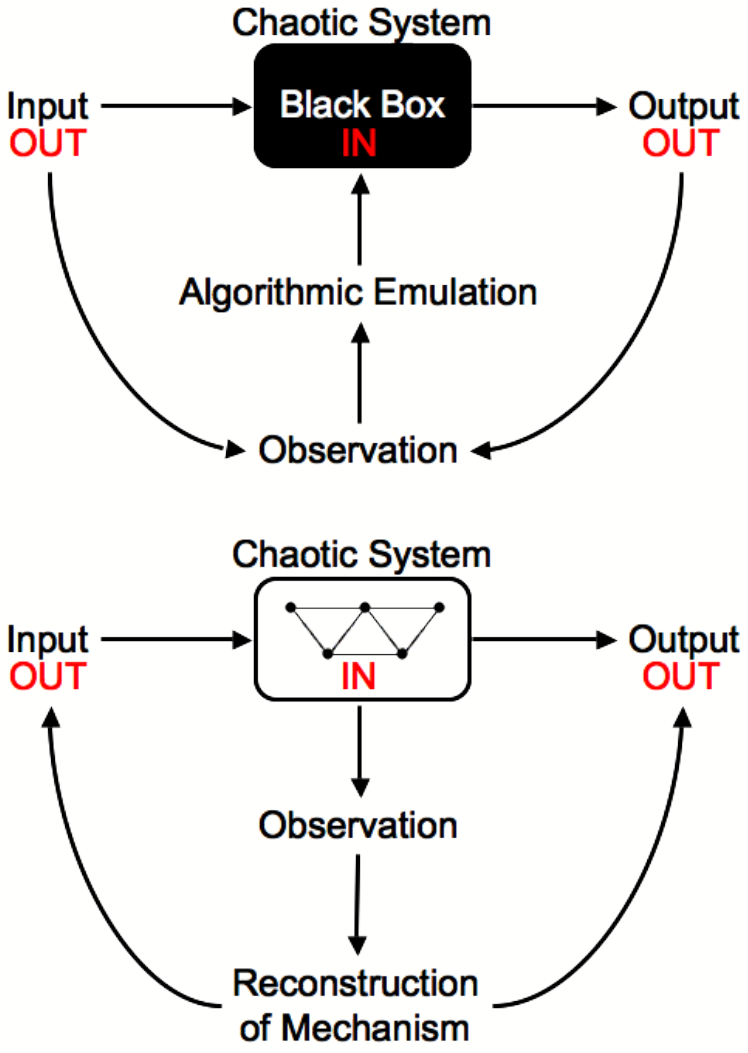


Figure 4.19: **Difference of approaches to chaotic systems in artificial intelligence and connectionism.** Top: Out-in emulation of artificial intelligence. Bottom: In-out reconstruction of connectionism. Artificial intelligence generally try to imitate the complex input-output relation of the system from the outside, by proposing the algorithm which realizes observed patterns. The model is not necessary following the neural basis, and in most cases conceptual. The connectionism starts from knowing the neural structure of cognitive system from the inside, and imitate its mechanism, expecting the reconstructed system would produce similar input-output relation.

Chapter 5

Internal Measurement of Chaotic Itinerancy as Catalyst of Learning

Abstract

We review and extend the measurement-oriented theory proposed by Y.-P. Gunji to interpret the emergent process of novel attractors in CNN investigated in the previous chapter. The results will be interpreted as an extension of celebrated concept of "*chaos as catalyst of learning*", which also finds the interpretation in terms of internal measurement. The novel attractors can be explained as a process toward the realization of pseudo-solution, generated by the autonomous learning based on the intrinsic mechanism of the system, introducing the modification of the interface to avoid chaotic itinerancy state as contradiction.

Keywords: Internal measurement, Measurement-oriented physics, Pseudo-solution

Methodology: Formulation of chaos in dynamical system as an internal observer → Formulation of CNN as an internal observer → Interpretation of the results of autonomous learning with measurement-oriented theory

5.1 Introduction of Measurement-Oriented Physics

Yukio-Pegio Gunji has been insisting the importance of transition from "*state-oriented physics*" to "*measurement-oriented physics*" in complex systems sciences. The measurement-oriented theory is based on the process of the internal observation by which it becomes conceptually accessible to the problem of emergence such as the origin of autonomous system. Such approach is also expected to bring novel explanation on the genesis of some observed laws in state-oriented physics, such as Zipf's law and 1/f fluctuation, to name a few. In this section, we will briefly introduce the perspective of Y.-P. Gunji based mainly on [4] (Ch.15) and [52], for further application on CNN dynamics observed in the previous chapter.

5.1.1 Cartesian Cut

The separation between an observer and an object is called *Cartesian cut* taking after the celebrated philosopher René Descartes. If we observe a system according to the Cartesian cut, we encounter the following problem:

When we try to identify an object, namely X_n, we have to define the observer X_{n+1} that is independent from X_n. Here, the relation between X_n and X_{n+1} defines the context of observation and we call it as *interface*, as depicted in Fig. 5.1. The interface of the observation, however, does not have any foundation since the X_n and X_{n+1} is supposed to be completely independent. The interface was chosen in its own terms, without any comprehensive ground. If we then look for the foundation of the interface between X_n and X_{n+1}, it requires a larger perspective that can be

called X_{n+2}. Then follows the foundation of X_{n+2} as X_{n+3}, whose foundation as X_{n+4}, ... and so forth. The pursuit of the foundation of observation ends up with infinite regression as in Fig. 5.2.

Gunji formalized such infinite regression by representing the interface as F, and symbolized with the following equation [52]:

$$F(X_n) = X_{n+1}. \tag{5.1}$$

Here, the object and observer are in absolutely relative relation, therefore for any $i < j$, the following holds.

$$F(X_i) = X_j. \tag{5.2}$$

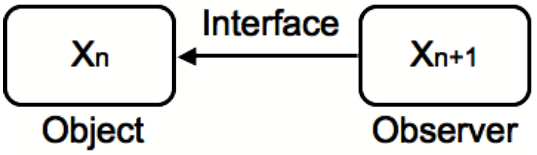


Figure 5.1: **Cartesian cut between object and observer, and connecting interface.** Based on [52].

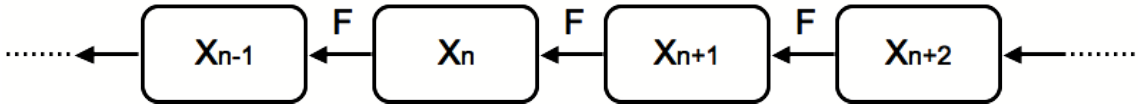


Figure 5.2: **Infinite regression between object and observer.** Based on [52].

5.1.2 Genesis of Contradiction and Pseudo-Solution

From the compatibility with respect to the limit operation of F in equation 5.2, we have the following:

$$F(\lim_i X_i) = \lim_F F(X_i). \tag{5.3}$$

Then with $F(X_i) = X_{i+1}$, we have

$$F(X_\infty) = X_\infty. \tag{5.4}$$

The X_∞ is formally called as *fix point* since it is invariant under the operation F. However, this notation seems to be contradicting the proper definition of F, which is to define the observer from the object. In Cartesian cut, the object and observer always have to be separated.

To solve this contradiction, we recall here that the equation 5.4 is not the definition, and reconsider it as an *equation* that corresponds to establishing a novel problem. We are interested in what brings the framework of measurement-oriented theory, at the limit of infinite regression. Therefore, the X_∞ is something that is at the same time object and observer. In actual situation, the fix point of a dynamical system is at the same time the input (object) and output (observer), if we consider that the system is described by F.

5.1.3 Groundlessness of Interface

In previous section, the pursuit of the foundation of interface between the object and observer led to the contradiction with the definition of F. Gunji introduced a perspective to identify the definition with the equation to solve this contradiction, but one should ask what is the justification of this identity. In fact, Gunji admits that there is no foundation for this rhetoric. The definition and the equation was groundlessly considered as compatible in equation 5.4. Therefore, the pursuit of the foundation of interface was shifted to that of between definition and equation, and the problematics are just postponed in another conceptual framework. In this sense, Gunji called this treatment as *pseudo-solution*, and rather insisted that this paradoxically supports the groundlessness of the interface. Gunji describes the situation as the following [52]:

If the goal of science is to establish a perfect, consistent description of the world by converging all movement and process into geometry, what we finally obtain would only be an absolutely relative interface. (Translation by M. Funabashi)

This indication is not only the criticism of usual sciences, but rather based on the problem consciousness in internal measurement theory, which tries to tackle the problem of groundlessness in a positive manner. Internal measurement theory is an attempt to start from an absolutely indefinite interface to describe the living world such as time, history, evolution, etc.

5.1.4 Dynamical Model of Internal Measurement

Gunji incorporated the groundlessness of interface into dynamical model (See Fig. 5.3). Let us think in discrete time steps t, $t+1$, $t+2$, $\cdots$, which is also a groundless definition called time. The state of the time t develops to that of $t+1$ with function f_t. Then, the contradiction to use infinitely the rule f_t appears to be the fix point in the equation (5.4), whose solution provides a novel rule of time development. This pseudo-solution defines the new function, for example, f_{t+1}.

Continuing this sequence, the time development produces at each time step a novel problem and solves it, which never reaches to a complete settlement but an endless continuation [52]. This is exactly the process of the identification between the definition and equation, and the derivation of its pseudo-solution in equation 5.4.

In this process, at every time step the system regenerates the time development rule, demolishing the previous one. Therefore, as long as we observe such system from exterior, the time development seems not to follow the principle of causality. Gunji concludes that the situation where the pseudo-solution in the form of the fix point is temporally admitted as an solution by the system is the very process of happening in internal measurement. He incorporated this principle in cellular automata and propose a basic model of internal measurement that provided further insights on this subject [52]..

5.1.5 Emergence as Pseudo-Solution of Contradiction

In our subjective cognitive process, when we encounter some contradiction, it is also possible to ignore such dissociation and change the context or way of asking the question, which would bring another way of dealing with than directly tackling it. Such attitude can be abstracted to create another dimension to eliminate the contradiction into it and pretend to have solved the problem in former context (Fig. 5.4). This treatment is abundant in our real life, although it is not permitted in the rhetoric of usual physics or logical operation. Such process is nothing but a pseudo-solution, as previously mentioned, but by declaring it as a temporal solution in former context, the contradiction actually disappears in that level. Gunji insists that this is the very core of the *emergence* [52]:

Emergence is the connection between the former and later logic, which are irrelevant with each other. The later logic appears only when the contradiction in former logic is solved at the limit of infinite regression with it. (Translation by M. Funabashi)

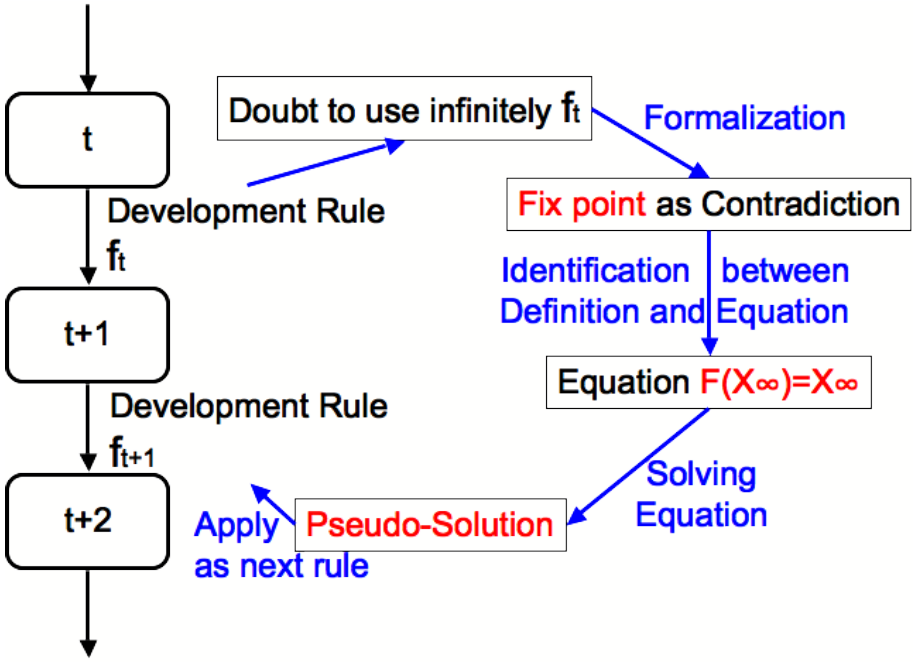


Figure 5.3: **Dynamical Model of Internal Measurement.** The time development rule f_t is reconsidered by the system itself due to the internal contradiction arising from the dynamics itself. The groundlessness of the interface f_t produces a contradiction at the limit of infinite regression, which can find its pseudo-solution by solving the fix point in equation (5.4). Based on [52].

After the emergence, novel logic (later logic) becomes the rule of the system, and produces another contradiction in given level. In any case, the contradiction is not completely solved, but continues to create another pseudo-solution.

5.1.6 Selection from outside of selective region

Let us rethink on the identification of the state in terms of the internal measurement. When we select something in the context of our daily life, one may think that there exist already some choices from which we make decision. Though what we select actually also includes the choices outside of the given ones. This situation is easier to understand in consideration of the "*leap in the dark*" by S.A. Kripke.

Consider two persons having a conversation. When one person speaks to another, the latter interpret the word with his own interpretation, and there exists no fundamental reason to define the way of understanding. What intermediates them in the most profound level is nothing but the groundless interface. Then the hearer can not only have the choice between the speaker's offers, but from the exterior of them. It comes quite often that we over/underestimate what lies behind a person's word. It becomes clearer in the communication between different language. The conversation does not have identical basis between two persons, which corresponds to the *leap in the dark* (Fig. 5.5). The identification of meaning is nothing but the risky leap in the unknown consequence. Such relation is not limited to our natural language communication. Gunji rather insists that similar situation also occurs when an observer identifies the state of an object in physical system [52]:

The identification of a state is in large sense a linguistic act, which means we can never define the exact context or way of detection. The observer should make a decision resisting the ambiguity that he never knows whether the choice is inside or outside of some context. (Translation by M. Funabashi)

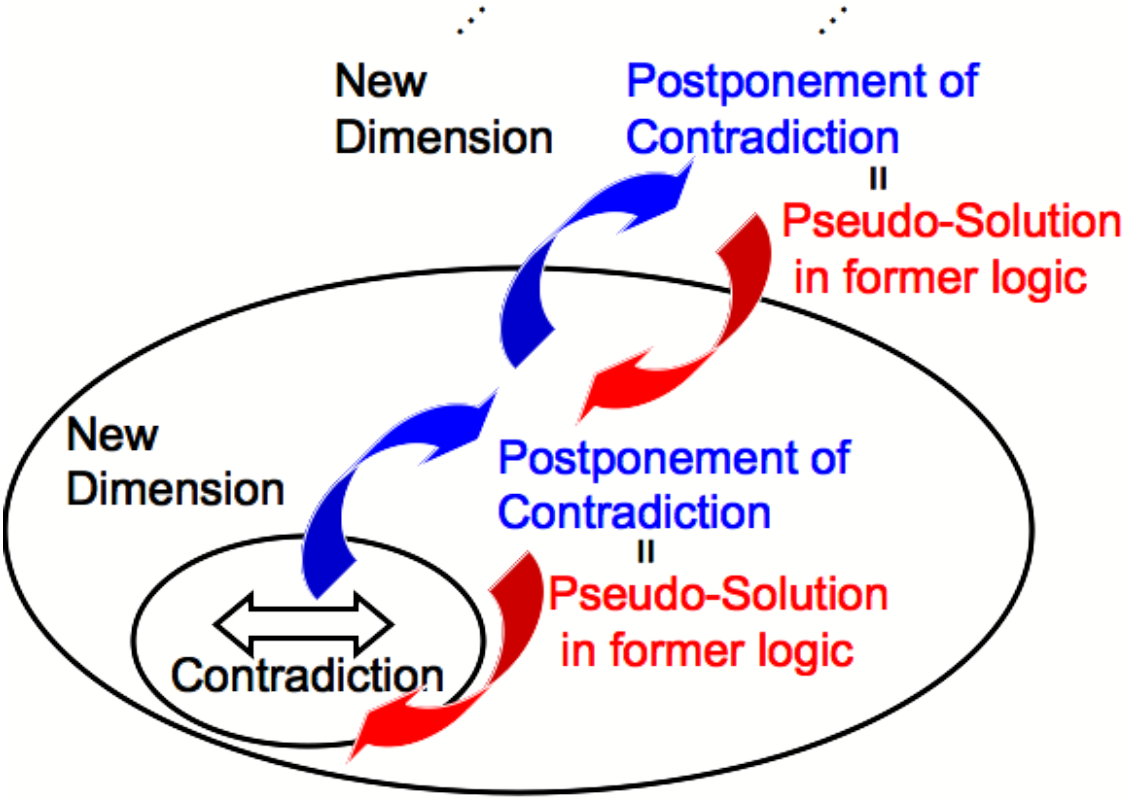


Figure 5.4: **Postponement of contradiction and emergence.** By introducing a new dimension that somehow trivializes the contradiction, the system declares it as the solution in former logic, which is actually a postponement of the problem in the new dimension. The emergence is defined here as the apparition of pseudo-solution in former logic, coupled and supported by the new dimension. Based on [52].

Gunji called the selection from outside of the selective region as emergence. Concrete examples are also abundant in biological systems:

Allergic reactions are one of the most convincing examples. The immune cells attack one's own organisms taking them outside of the proper context of exogenous material. In this sense, allergic reaction is an emergent property. Here, immune cells are observers, and the antigen of allergy is the object. Usually the antigen of allergy is harmless and its selective region is the immune tolerance. In allergic reaction, the identification of the antigen are made from outside of this context, which sensitizes immune cells and causes allergic reaction as the contradiction of logic between observer and object. It is of interest that the allergy can actually be desensitized by introducing other exogenous material such as parasites, which corresponds to add a new dimension to postpone the conflict and lead the former contradiction to pseudo-solution.

The dynamics of embryogenesis is also a fascinating example. Despite the communication between cells being nothing but the leap in the dark due to the large physical/chemical fluctuation, the stability and reproducibility of the process is highly attained. Each cell seems to follow its own logic according to their genetic and epigenetic profile, not knowing completely the state of other cells and even less on the total embryo. In this case, each cell is an internal observer, and their interaction is the emergence.

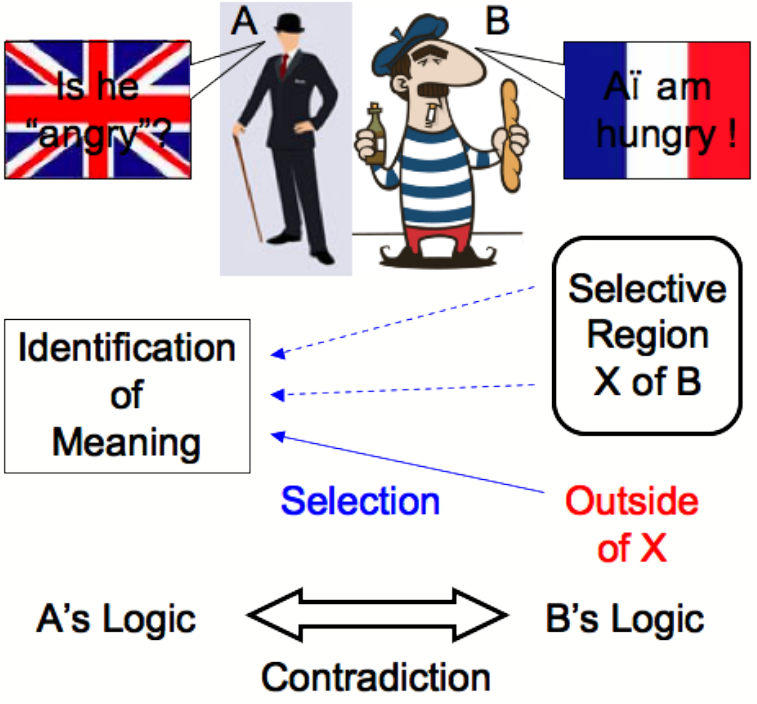


Figure 5.5: **Selected region of object and selection by observer from outside of it.** An object B has its own context and logic which define its selective region X. The observer A identifies the meaning of B's state by "the leap in the dark", not knowing whether the decision is inside or outside of B's selective region. The emergence occurs when the selection is made from outside of the selective region, which arises contradiction between observer and object, leading to pseudo-solution. Based on [52].

5.1.7 Toward Measurement-Oriented Theory

We have reviewed in this section the achievement and the possibility of measurement-oriented theory by Gunji. So far, most of the study in complex systems sciences are indeed in state-oriented theory. This is largely due to the accessibility to mathematical modeling and analysis developed in the same perspective. It is of further question whether we can develop a more convincing framework

of modeling in measurement-oriented theory, in the level to overcome the missing aspect in the state-oriented one.

In the following sections, we apply the measurement-oriented theory to interpret the results of previous chapter in terms of emergence. We first consider the celebrated concept of "*chaos as catalyst of learning*" in this framework, which naturally leads to the interpretation of autonomous learning during chaotic itinerancy in CNN.

5.2 Internal Measurement of Chaos as Catalyst of Learning

5.2.1 Chaos as Catalyst of Learning

We revisit the celebrated concept of "*chaos as catalyst of learning*" that motivated the synthetic modeling with CNN in the previous chapter. The concept was established based on the learning dynamics of novel odor in rabbit's olfactory bulb [17]. Exposure of a rabbit to already known odor stimulus results in the quick convergence of EEG dynamics in the corresponding attractor region, which is associated to the recognition of the odor. Novel odor, however, does not show convergence of EEG dynamics but weakly chaotic itinerant sequence among memory attractors of known ones. Chaos plays here the role of novelty filter. During this chaotic "I don't know" state, synaptic modification occurs and a novel attractor corresponding to the given odor emerges, without losing other memories. It was shown from modeling study that such organized acquisition of novel memory is essentially supported by the chaotic dynamics. Therefore, chaos can be functionally a catalyst of learning.

5.2.2 Internal Measurement Formalization

Let us reconsider the experiment of Skarda and Freeman with the view of the measurement-oriented theory. For simplicity, we formalized the qualitative change of attractor structure in olfactory bulb with symbolic characters in Fig. 5.6. We assume that the subject only knows the odor A' and B', which corresponds to the attractor A and B in its olfactory bulb. The olfactory input of novel odor C' evokes chaotic dynamics, and the initial state X_0 of EEG dynamics wanders between A and B. Olfactory bulb performs autonomous dynamics with external input, and with the presence of chaos, it can be considered as an internal observer: In autonomous dynamics, the system observes itself recurrently but there exists fundamental limit of observation. The system is not capable to observe itself globally In real-value precision. Single neurons act according to the limited information they obtain, without knowing the global state. The interface produces the state of the next time step with the limited part of information of previous one. In chaos, only slight perturbation will affect the global state in short time period. The autonomous dynamics of olfactory bulb with symbolical time t is compatible to the following internal measurement with the interface of observation F:

$$X_{t+1} = F(X_t). \tag{5.5}$$

The recognition of known oder is therefore expressed as fix point, such as

$$A = F(A), \tag{5.6}$$

and

$$B = F(B). \tag{5.7}$$

While with odor C', X_t wanders between A and B.

The wandering chaotic state is the contradiction to recognize the odor, so-called "I don't know" state. Further exposure to the novel odor C' produces synaptic changes, and create the novel attractor C that recognizes the odor C'. This process exactly corresponds to solve the fix point equation (5.4):

To solve the contradiction of "I don't know state", the olfactory bulb seeks for some attractor X such that

$$X = F(X). \tag{5.8}$$

Since we only have the solution of (5.8) as the attractor A and B, which corresponds to the odor A' and B', respectively, the dynamics always follows $X_t \neq F(X_t)$ under exposure of odor C'.

If we define the EEG pattern of olfactory bulb after n time step as $X^n = F(X_0)$, the pseudo-solution of the fix point equation (5.4) corresponds to the attractor C, recognizing the odor C':

$$F^{\infty}(X_0) = F(X_{\infty}) = X_{\infty} = C. \tag{5.9}$$

Where the interface F represents the mechanism of olfactory bulb with the interaction between synaptic plasticity and chaos. In this case, chaos triggers the "doubt to use the same interface", and synaptic modification solves the fix point equation (5.4) in Fig. 5.3.

In actual experiment, for enough large n reflecting the learning period, the following becomes the measurement-oriented formalization of chaos as catalyst of learning:

$$C = X_n = F(X_{n-1}) = F^n(X_0). \tag{5.10}$$

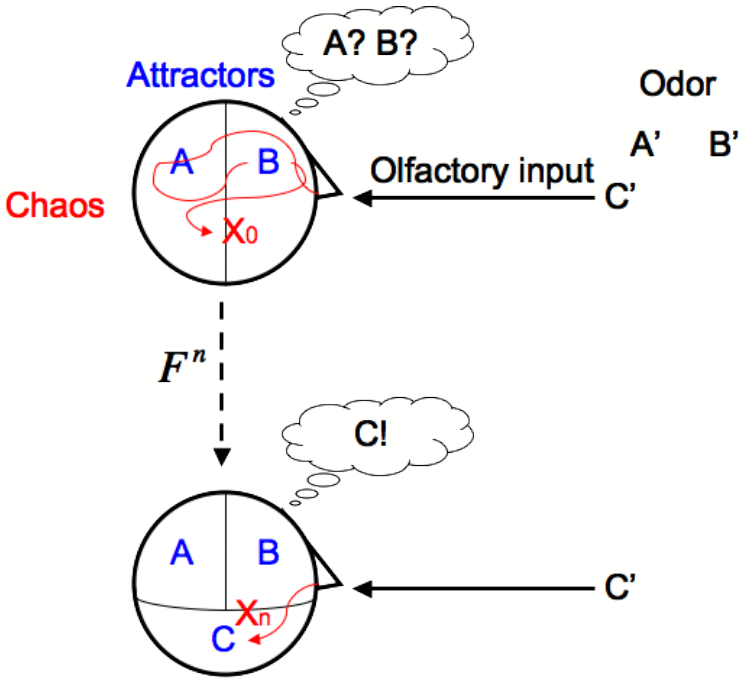


Figure 5.6: **Schematic representation of chaos as catalyst of learning with measurement-oriented theory perspective.** Known odor (A' and B') are quickly recognized by converging to the corresponding attractors (A and B). While novel odor (C') triggers chaotic transition between memory attractors, at the same time affects synaptic connection and modifies memory structure. In a word, chaos catalyzes the genesis of novel attractor (C) of given stimulus (C').

5.3 Internal Measurement of Chaotic Itinerancy as Catalyst of Learning

5.3.1 Internal Measurement of Chaotic Itinerancy in CNN

We further develop the measurement-oriented formalization and apply it to interpret the result of previous chapter. First of all, the simulated CNN is an internal observer in chaotic itinerancy state, and the time development Φ is an internal measurement: A CNN of n neurons without external input is basically a deterministic system of $2n$ degrees of freedom, and all of the past history is maintained with exponential decrease by self-recurrence. In theoretical formalism, the system's state is completely defined with the real-value precision, but the simulation can only realize dynamics of limited digits with rounding error. Since chaotic dynamics is sensitive to initial condition and so is chaotic itinerancy at least to the level of computational capacity $(1.0e-16)$, this rounding error is unignorable and causes nothing but "the leap in the dark". Each neuron model selects its future state out of the context of complete determinism. The rounding error is also a deterministic noise, but we do not know about its sequence until we actually calculate it. Therefore, the time development of the internal state $\mathbf{X}_t = (\eta(\mathbf{t}), \zeta(\mathbf{t}))$ of CNN at time t is the internal measurement with interface Φ, which uses essentially only a part of the information of $\mathbf{X}_t$:

$$\mathbf{X}_{t+1} = \Phi(\mathbf{X}_t). \tag{5.11}$$

Chaotic itinerancy with memorized patterns A⇌B and C⇌D is therefore formally expressed as internal measurement as follows:

$$\begin{aligned} A \rightleftharpoons B &= \Phi^k(C \rightleftharpoons D), \\ C \rightleftharpoons D &= \Phi^k(A \rightleftharpoons B). \end{aligned} \tag{5.12}$$

Where k stands for arbitrary iteration number according to the statistics of transition dynamics between attractor ruins.

Considering the periodic cycles A⇌B and C⇌D each as a symbol, we have the interpretation of chaotic itinerancy as a "contradiction" state, analogous to the "I don't know" state of chaotic dynamics in olfactory bulb.

5.3.2 Internal Measurement Formalization of Autonomous Learning

We consider the result of autonomous learning in the same formalism. Fig. 5.7 shows the emerging process of novel attractors in CNN analyzed in previous chapter, and its schematic composition corresponds to Fig. 5.3 in the context of internal measurement theory.

The increase of refractoriness parameters k_r and α changes the dynamics of CNN into the chaotic itinerancy state, which corresponds to the "doubt to use the same interface" in Fig. 5.3. In CNN, the chaotic deviation from the memory attractors is the doubt to use permanently the same memory structure. In olfactory bulb, such transition was triggered by the novelty of odor stimulus.

Chaotic itinerancy can be formalized as the contradiction in eq. (5.12), which assumes the fix point at the temporal limit as pseudo-solution.

Autonomous learning is the synaptic weight modification of the system by itself, which corresponds to modify Φ at each step of t and leads to a settlement $\Phi_\infty(X_\infty) = X_\infty$. Such X_∞ is known as the result of the divergence or convergence of connection matrix by learning rules.

In autonomous learning, we set the learning period n during chaotic itinerancy state. After n steps of learning, the refractoriness parameters k_r and α are reset to 0, and novel attractor emerges as investigated in previous chapter. (Such escape from chaotic learning dynamics is realized automatically in olfactory bulb, as a function of chaos as novelty filter.) The dynamics after the autonomous learning in CNN can be formally summarized as follows:

$$E \rightleftharpoons F = \Phi^2(E \rightleftharpoons F), \tag{5.13}$$
$$G \rightleftharpoons H = \Phi^2(G \rightleftharpoons H), \tag{5.14}$$
$$P_{STDP} = \Phi^k(P_{STDP}). \tag{5.15}$$

Where P_{STDP} represents various periodic attractors of period k emerged after STDP learning.

These results are the n step convergence of autonomous learning, and are expressed as fix points in Fig. 5.7. Therefore, the patterns $E \rightleftharpoons F$, $G \rightleftharpoons H$, and P_{STDP} are ones of the pseudo-solutions that reflect the dynamics of chaotic itinerancy as contradiction. These pseudo-solutions are consistent in that they contain equidistant patterns from $A \rightleftharpoons B$ and $C \rightleftharpoons D$, and that they exist within the union space of those patterns in terms of effective dimension. In short, the emerged patterns do not coincide with memorized ones, but possess intermediate feature in terms of geometrical composition in the state space. These properties also fit to the characteristics of the pseudo-solution that introduces another dimension to solve the contradiction into the postponement, in this case as a geometrical compromise.

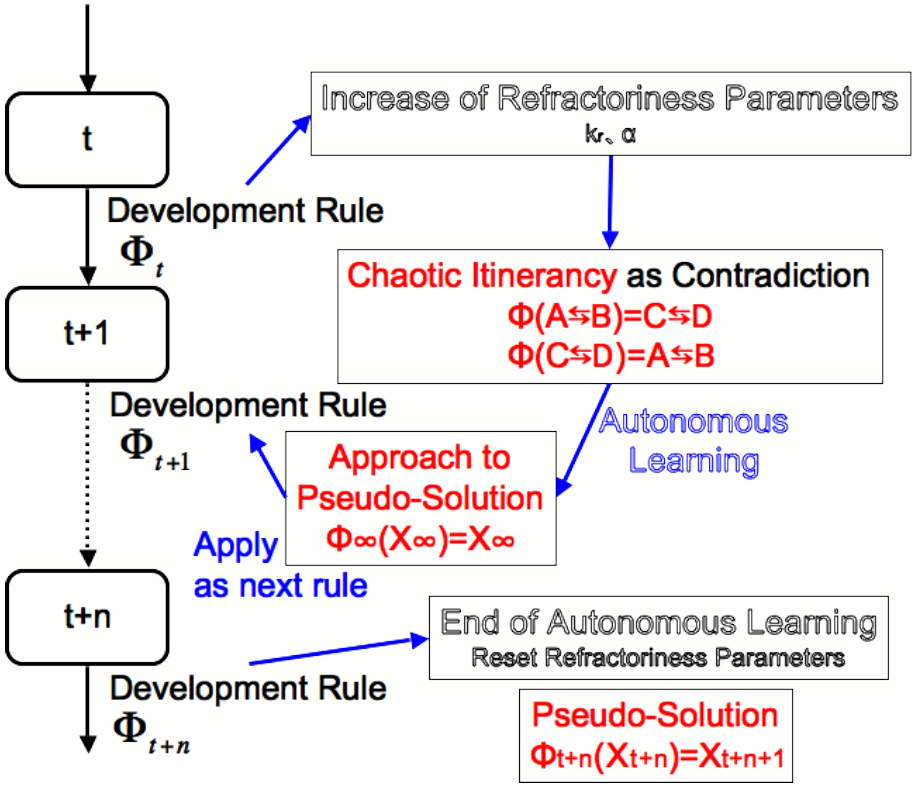


Figure 5.7: **Dynamical model of autonomous learning as internal measurement.** The increase of refractoriness parameters causes CNN chaotic dynamics that can be formalized as the contradiction between memorized patterns. The autonomous learning modifies the interface Φ by changing the connection matrix W according to the dynamics of the system itself, and produces novel memory structure as fix points after n step of learning. For simplicity, we avoid the notation of k in Φ^k.

5.4 Conclusion

We reviewed the measurement-oriented theory proposed by Gunji, and applied to interpret the "*chaos as catalyst of learning*" in olfactory bulb, as well as the emergent patterns of the autonomous learning during chaotic itinerancy state of CNN investigated in previous chapter. Dynamical system

manifesting chaos is shown to be an internal observer. Internal measurement theory provided a way to conceptually define the structure of emergence in CNN catalyzed by chaotic dynamics.

Part III

Applications of Chaotic Neural Network in Emergent Robotics

Chapter 6

Introduction of Neurorobotics

Abstract

We introduce the concept of "*neurorobotics*" as a new source of invention from interdisciplinary domain between neuroscience and robotics. Context of work of this part is summarized.

Keywords: neuroscience, robotics, embodiment

6.1 Introduction: Context of work

The following chapters in this part were realized in cooperation with M. Shyunsuke Aoki, who was a master student at the college of information science and technology, Donghua university in Shanghai, china, and also the founder of a venture company in robotics, YUKAI inc. . Our interests found coincidence in the application of mathematical model of neuroscience to robotics, which we named "*neurorobotics*". In this chapter, we introduce the concept of neurorobotics and the motivation of our works in this part.

6.2 Neurorobotics as the Interdisciplinary Domain between Neuroscience and Robotics

Science and technology has always been evolving depending on each other. The dead rocks for one have often been overcome with the innovation of the other. Recently, the newly evolving neuroscience has been greatly promoted not only by the progress of experimental equipment but the innovation of computer technology. Simulation of dynamical systems that is widely used in neuroscience has become numerically possible, which enable us to investigate the spatiotemporal activity of neural circuit in its dynamics. Although mathematical modeling can only encapsulate a part of the huge complexity in neural dynamics, it helps us analyze rigorously the dynamical response in the defined setting. Typical feature of actual dynamics can synthetically be understood by combining these models. The modeling is also valid not only to test hypothesis by experiment, but to extend it purely on the model basis. Logical autonomy of the model is also a source of theoretical innovation.

Among the existing approach in neuroscience, there is an insistence that to create the brain artificially is a way to understand it. This constructive approach to neuroscience also has its exit in robotics. In this field, the main issue is to find the top-down correspondence between theoretically developed algorithms to solve given tasks in engineering and existing mechanisms of the brain [53]. For example, relations between reinforcement learning and basal ganglia, meta-learning and neuromodulators, the origin of reward system, etc., have been investigated by creating the analogical robots [54][55][56][57][58] .

On the basis of such constructive approach in robotics, we focus on creating the robot with the use of the obtained knowledge on the brain, and investigate its utility in the engineering point of

view seeking for real-world application. Contrary to the mainstream of artificial intelligence, this approach does not ignore the biological foundation of the brain. Our aim is to construct possible application of biologically observed mechanism as an invention in an artificial setting to better contrast the task solvability.

We call this approach as *neurorobotics*, starting from bottom-up models in neuroscience, and by implementing them to the robots, evaluate the efficiency in terms of the top-down algorithm.

This is also a part of the embodiment approach that is often discussed when investigating the function of the brain in the entity, including the environment and the body [59][60][61]. In this setting, the embodiment is investigated in biological context, which is subjective to the emergence of function in actual creatures, since the body defines the characteristic of the interface that play essential role to connect the internal and external information of the system.

In neurorobotics, we do not necessary pursuit the effect of embodiment by imitating actual living organism. Instead, we consider the invention of task-based embodiment to connect the bottom-up models in neuroscience and top-down algorithms in robotics (Fig.6.1). This approach is at the same time oriented to the engineering interest that utilizes brain dynamics out of biological context to create a new value, and discover the hidden efficiency of the dynamics that reversely clarifies the rationality/specialization of the biological mechanisms in generalized function space.

The use of embodiment in robotics context is also investigated in artificial life, to reveal and experiment the abstract top-down condition of life [62]. Here, the reconstruction of life in an artificial setting is the primal objective. Chaotic itinerancy is assumed to be a key function in this approach.

To extend the embodiment context of brain model out of actual biological configuration also allows us to consider the dependence of the brain on actual embodiment designed by evolution, which is related to the design of brain-machine interface.

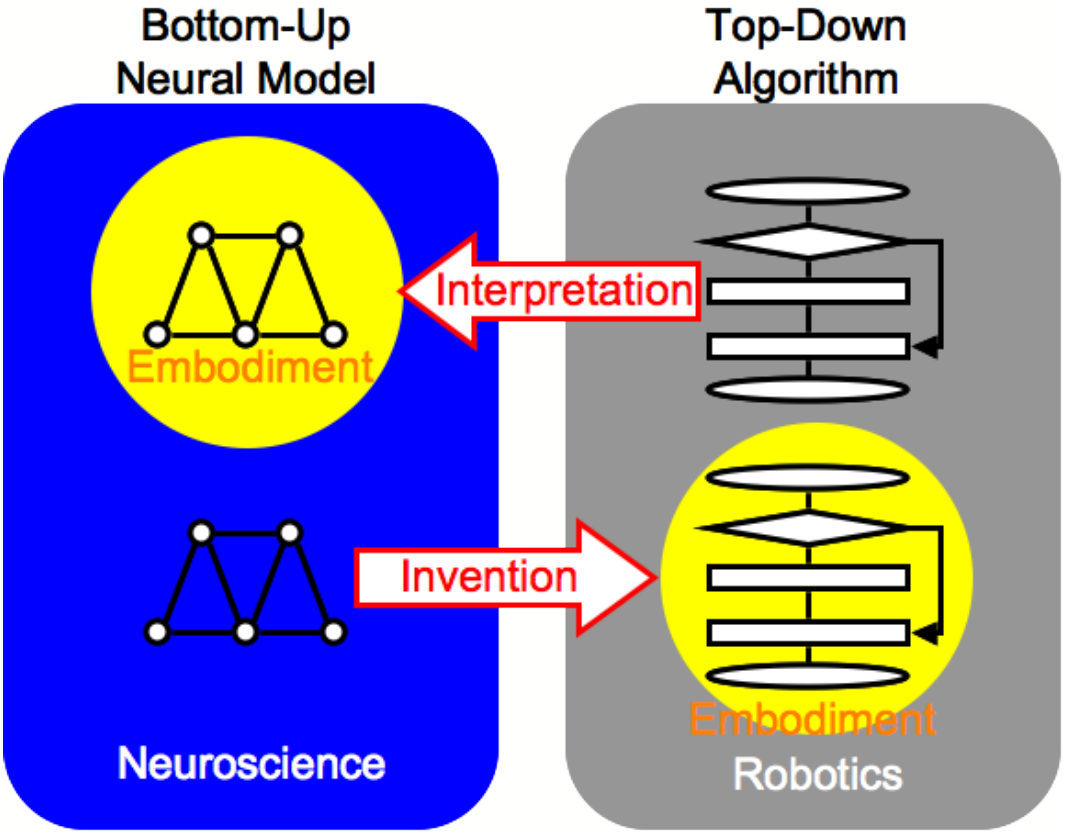


Figure 6.1: **Motivation of neurorobotics in contrast to conventional embodiment in robotics.** Top arrow: Conventional embodiment in robotics with biological context. Bottom arrow: Newly proposed approach using embodiment as a factor of invention in neurorobotics.

Chapter 7

Balancing Autonomy and Environmental Response with Hierarchical Chaotic Dynamics

Abstract

Hierarchical structure of deterministic chaos in a chaotic neural network model (CNN) is investigated in the view of application in robotics. The result shows a rich capacity of CNN in selectively controlling the synchronization of neuron outputs, and sensitively responding to external sensory inputs, both being based on the intrinsic mechanism of the dynamics called chaotic itinerancy. Choosing appropriate parameters, the simple designed robot realized a chaotic search to the hierarchically selected directions. The macroscopic drift preserving chaotic fluctuation was also derived by simply adding weak external inputs to an intended direction. Obstacle avoidance was simulated with the use of these properties.

Keywords: chaotic neural network, invariant subspaces, chaotic itinerancy

Methodology: Designing chaotic roving robot with the use of hierarchical dynamics of CNN as a completely deterministic dynamical system → probabilistic analysis of the system's behavior and response

7.1 Introduction

Recent development of cognitive sciences has brought increasing number of application in robotics. Especially, the model-based approach in neuroscience has a wide range of direct application in designing sensory-motor interaction of robot [62] [63]. Learning from the mechanism of actual brain has a great possibility to extend the information processing ability of artificial intelligence.

Among the accumulating results in neuroscience, a class of dynamics called *chaotic itinerancy* has been insisted as a generic mechanism supporting brain function [21] [23] [22] [20]. The chaotic itinerancy is phenomenologically described as the coupling of locally destabilized plural attractor ruins. The dynamics shows irregular transition among these attractor ruins, at the same time completely following the deterministic development rule. This non-uniform localization of stability and instability bring the network the properties of both destroying and retaining input information.

For example, the property of sensitive response to attractor fragment input in the chaotic neural network model (CNN) [24] has a possibility to provide a robot with quick and efficient way to react to sensor input [64]. We applied this property to quickly coordinating sensor input with motor output.

On the other hand, In networks with symmetry such as GCM (globally coupled map) and CNN, there exist geometrical division called invariant subspace, which is a remarkable difference between

chaotic itinerancy and other stochastic systems, both capable of realizing irregular transition among attractors [28] [1].

We applied the structure of invariant subspace to control robot motion in a hierarchical way, and propose a basic mechanism of actively interacting with environment using the both stability and instability of chaotic dynamics.

7.2 System Description

7.2.1 Chaotic Neural Network

We define the chaotic neural network model(CNN) [24] that we used as the control system of a robot in this paper. The architecture of CNN is defined as in Fig. 16.2. CNN is a discrete time system that consists of three layers of neuron model: Input, Context and Output. Each neuron of the Context layer is connected to all neurons in the Output layer. The output of the Context layer at time t is identical to that of the Output layer at time $t-1$. The Input layer gives the external inputs in real number to the Output layer, which will be explained in the Simulation Result section. The chaotic neurons are used in the Output layer of CNN, that are defined as follows [24]:

$$\begin{aligned} x_i(t+1) &= f(\sum_{d=0}^{t} k_e^d \iota_i(t-d) \\ &\quad + \sum_{j=1}^{n} w_{ij} \sum_{d=0}^{t} k_f^d x_j(t-d) \\ &\quad - \alpha \sum_{d=0}^{t} k_r^d x_i(t-d) + \theta_i^{out}), \end{aligned} \tag{7.1}$$

where at time t , $x_i(t)$ is the output of the ith chaotic neuron, and θ_i^{out} is its threshold. $\iota_i(t)$ is the external input to the ith chaotic neuron. The parameters α and k_r $(0 \leq k_r \leq 1)$ control the orbital stability and can induce chaotic dynamics. The exponentially decreasing influences of past external inputs (outputs of the Input layer) and past outputs (outputs of the Context layer) are controlled by k_e and k_f $(0 \leq k_e, k_f \leq 1)$, respectively. n is the number of chaotic neurons. In this study, eight chaotic neurons with fixed parameters $k_f = 0.2$ and $\theta_i^{out} = \frac{0.2}{1-k_r}$ were used in the Output layer of CNN. Note that the number of neurons and the patterns of the attractors was chosen to realize a non-trivial invariant subspace structure with the most simple network, which is described in the following section.

The function $f(\cdot)$ is defined as follows, with the increment parameter β.

$$f(x) = \frac{1}{1+\exp(-\beta x)}. \tag{7.2}$$

We used $\beta = 20.0$ in this study.

Next, we define the connection matrix W to memorize three pairs of reverse patterns, in total six patterns, each one as fixed point attractor. The six orthogonal patterns $\mathbf{P^1} = (p_1^1, \ldots, p_8^1)^T, \cdots,$ $\mathbf{P^6} = (p_1^6, \ldots, p_8^6)^T$ are defined as column vectors in Tab.7.1. For notational simplicity, we put each two corresponding reverse patterns together, and call them as "a pair of patterns", since they are both situated in the same invariant subspace (see the Invariant Subspace section). To realize these patterns as stable autoassociative memories in CNN, we defined the connection matrix $W = (w_{ij})$ $(1 \leq i, j \leq n)$ as follows, based on the Hebb's learning rule.

$$w_{ij} = \frac{1}{6} \sum_{k=1}^{6} \{(2p_i^k - 1)(2p_j^k - 1)\}. \tag{7.3}$$

Setting the parameters $\alpha = k_r = 0$, after some short transient dynamics, CNN continuously retrieved one of the six patterns depending on the initial value. For the following analysis, we

Table 7.1: **Definition of the six patterns $\mathbf{P^1}$, $\cdots$, $\mathbf{P^6}$, to be memorized in the Output layer of CNN.**

Pattern symbol	Pattern vector
$\mathbf{P^1}$	$(1,1,1,1,0,0,0,0)^T$
$\mathbf{P^2}$	$(0,0,0,0,1,1,1,1)^T$
$\mathbf{P^3}$	$(1,1,0,0,1,1,0,0)^T$
$\mathbf{P^4}$	$(0,0,1,1,0,0,1,1)^T$
$\mathbf{P^5}$	$(1,0,1,0,1,0,1,0)^T$
$\mathbf{P^6}$	$(0,1,0,1,0,1,0,1)^T$

judged the retrieval of the six patterns $\mathbf{P^1}$,$\cdots$, $\mathbf{P^6}$, through the following transformation $h_i(t)$ of the outputs $x_i(t)$ of the Output layer:

$$h_i(t) = \begin{cases} 1 & \text{if} \quad x_i(t) > 0.5 \\ 0 & \text{if} \quad x_i(t) \leq 0.5 \end{cases} . \tag{7.4}$$

For example, if $h_i(t) = p_i^k$ $(1 \leq i \leq 8)$ holds at time t , we state that the network retrieved the pattern $\mathbf{P^k}$.

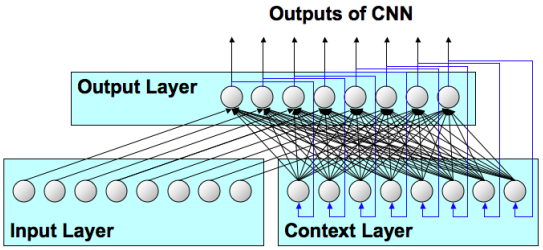


Figure 7.1: **Architecture of the chaotic neural network.**

7.2.2 Projection of Phase Space to Motor Outputs

In this paper, we designed a simple robot capable of moving to three directions in two-dimensional space by three directional motors. The main purpose here is to clarify the effect in real world movement of hierarchical control and environmental response realizable with CNN dynamics, so that this design is a primitive one, focused on the convenience of analysis.

Fig. 7.9 and Fig. 7.10 show the schematic architecture of the robot. each of three motors is located at each corner's bottom of triangular robot body, and control the forward (no backward) movement to its corresponding direction. Only one motor is chosen to give output at each time step t, according to the retrieval pattern of CNN outputs $\mathbf{x}(t)$. That is, the patterns $\mathbf{P^1}$ and $\mathbf{P^2}$ correspond to the motor 1, the patterns $\mathbf{P^3}$ and $\mathbf{P^4}$ correspond to the motor 2, and the patterns $\mathbf{P^5}$ and $\mathbf{P^6}$ correspond to the motor 3, respectively. If none of these six patterns are retrieved, we do not move the robot at this step.

The root-squared norm of CNN outputs $\mathbf{x}(t)$ is then calculated to give the strength of the chosen motor output. In real robot, for example, this value will corresponds to the duty ratio of PWM wave offered to the motor. The robot therefore moves to the motor direction proportionally to its output norm.

Note it is easy to have extension of motor outputs in this architecture by augmenting the number of neurons. By simply applying eq.(7.3), CNN with 2^n neurons can store n pairs of orthogonal patterns defined in the same manner as in Tab.7.1, along with the hierarchical structure of invariant subspaces explained in later sections.

Hence, the embodiment of CNN dynamics is not restricted to this single manner, nor even to physical movement, since this is a neural network model at the beginning. One can think of a robot using the dynamics of CNN to trigger intelligent events of any symbolic form with similar framework.

7.3 Dynamics of Chaotic Neural Network

The CNN with refractoriness parameters $k_r = 0$ and $\alpha = 0$ is identical to the simple recurrent neural network with analog neuron model [24]. The stability of six autoassociative patterns $\mathbf{P^1}$,$\cdots$, $\mathbf{P^6}$ as fix point attractors are therefore theoretically assured by Hebb learning and confirmed numerically. Though, by augmenting the parameters k_r and α, CNN can increase the effect of refractory variables and cause partial instability to these attractors. The state where the attractors undergo boundary crisis and the dynamics of system expands to high-dimensional chaos is phenomenologically described as *chaotic itinerancy* or *chaotic wandering* [23] [24]. In this stage, irregular transition among attractor ruins is commonly observed, where the orbits stay more or less frequently near the destabilized patterns. Since the dynamics holds both the stability towards the memorized patterns and also the instability escaping from them, this state can be considered as a dynamical associative memory and is useful for information processing [62] [63] [24].

In this section, we numerically analyze the nature of chaotic itinerancy occurred in our CNN.

7.3.1 Deviation Rate from the Memorized Attractors

First, we calculated the deviation rate $DR(k_r, \alpha)$ of CNN for an empirical range of the refractoriness parameters k_r and α ($0 \leq k_r \leq 1$ and $0 \leq \alpha \leq 10$). The deviation rate is the rate of outputs' binary interpretation $h_i(t)$ ($1 \leq i \leq n$) that does not match to any of the six stored patterns $\mathbf{P^1}$,$\cdots$, $\mathbf{P^6}$ during the retrieval. This measure therefore gives us a simple notion of instability of stored patterns, and refers to the mean visiting rate to any of them.

The deviation rate $DR(k_r, \alpha)$ for each k_r and α was defined as follows,

$$DR(k_r, \alpha) = \frac{1}{N} \lim_{N\to\infty} \sum_{t=1}^{N} (1 - \sum_{k=1}^{6} \prod_{i=1}^{n} \delta(h_i(t), p_i^k)) \tag{7.5}$$

where $\delta(\cdot, \cdot)$ is the delta function. $N = 10000$ was used for the calculation.

Fig. 7.2 shows the result. In the blue region ($DR = 0$), k_r and α are relatively low, and the stored patterns exist as stable fix point attractors. With the augmentation of these parameters, there exist regions where the deviation rate takes value between 0 and 1, suggesting the memorized patterns are destabilized and dynamics move in and out of some of these patterns. If there exists irregularity in periodicity, or more precisely the instability in terms of the maximum Lyapunov exponent, the dynamics can be judged to be in chaotic itinerancy state. The brown region ($DR = 1$) is where the dynamics retrieved none of the stored patterns during calculation.

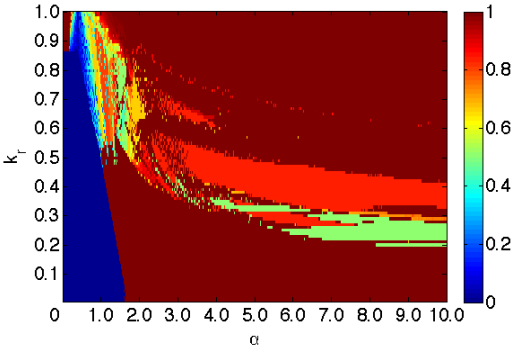


Figure 7.2: **Deviation rate of retrieval dynamics from memorized patterns.** The deviation rate was calculated after cutting 5000 steps of transient dynamics, with 10000 steps of retrieval at each k_r and α. Initial values were taken randomly.

7.3.2 Invariant Subspace

In discrete time neural networks, considering the symmetry of connection matrix, it is possible to decompose its dynamics into subspaces of the whole state space. The decomposition of the dynam-

ics into lower dimensions has a possibility to characterize the network dynamics more precisely. This is also what distinguishes between random and chaotic process. Despite the strong irregularity, chaotic dynamics in CNN can be decomposed into subspaces, according to its deterministic development rule. The decomposed subspaces in which dynamics can fall and remains permanently are called invariant subspaces [28].

In short, the invariance under a permutation of the connection matrix W can be interpreted to the invariance under the same permutation of the dynamics $\mathbf{x}(t)$. Therefore, it is possible to detect the partial synchronous mode of dynamics, whether it is stable or not, by just observing the symmetry of the connection matrix.

Since we do not have enough space to explain the derivation in detail, we simply note the minimum descriptive requirement and the calculation result.

Let S_n be a symmetric group with degrees of n, and σ be an element of S_n. If $w_{ij} = w_{\sigma(i)\sigma(j)}$ holds on the connection matrix W, we can define the invariant subspace $H(\sigma)$ of CNN Output layer $\mathbf{x}(t)$ as follows:

$$\begin{aligned} H(\sigma) &= \{(x_1(t), \cdots, x_n(t))^T \in \mathbf{R}^n \\ &\quad \mid x_i(t) = x_{\sigma(i)}(t), 1 \leq i \leq n\}. \end{aligned} \tag{7.6}$$

Hence, this holds for all time steps t of CNN. Consequently, it means that if the initial states of CNN were taken inside of $H(\sigma)$, the i th neuron and the $\sigma(i)$ th neuron of the Output layer are always synchronized and output the same value at each time step.

If, for example, $n = 8$ and take the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 4 & 3 & 6 & 5 & 8 & 7 \end{pmatrix} \in S_8, \tag{7.7}$$

which satisfies the condition $w_{ij} = w_{\sigma(i)\sigma(j)}$, then

$$\begin{aligned} H(\sigma) &= \{(x_1(t), x_2(t), x_3(t), x_4(t), \\ &\quad x_5(t), x_6(t), x_7(t), x_8(t))^T\}, \end{aligned} \tag{7.8}$$

where

$$\begin{aligned} &x_1(t) = x_2(t), x_3(t) = x_4(t), \\ &x_5(t) = x_6(t), x_7(t) = x_8(t). \end{aligned} \tag{7.9}$$

Note that the actual dimension of dynamics in this $H(\sigma)$ is reduced into four dimensions. Therefore, we conventionally call it as a four-dimensional invariant subspace.

For simplicity, let us denote

$$H(\sigma) = (11335577), \tag{7.10}$$

representing $H(\sigma)$ with the minimum subscripts of vector $\mathbf{x}(t)$ that are identical by the permutation σ . We take this notation rule for the following sections.

Based on the definition of $H(\sigma)$, we detected the invariant subspaces of CNN. The Fig. 16.3 shows the result. Dynamics on invariant subspaces that have the same type of clustered synchronization are phase conjugate, and these are grouped by the same color in Fig. 16.3. Hence, each pair of patterns are situated in the same two-dimensional invariant subspace, i.e., the patterns $\mathbf{P^1}$ and $\mathbf{P^2}$ are in (11115555), the patterns $\mathbf{P^3}$ and $\mathbf{P^4}$ are in (11331133), and the patterns $\mathbf{P^5}$ and $\mathbf{P^6}$ are in (12121212), respectively.

Among the detected invariant subspaces, there exists an important hierarchical structure according to the inclusion relation. Hence, we designed these six patterns $\mathbf{P^1}$,$\cdots$, $\mathbf{P^6}$ intending to have the useful inclusion hierarchy for controlling the orbit of CNN. According to the Fig. 16.3 , each pair of two-dimensional invariant subspaces (yellow) can find their union space with one-to-one relation in four-dimensional invariant subspaces (blue). For example, the set of two-dimensional invariant subspaces (11115555) and (11331133) has its unique corresponding union space (11335577), that is different from the union of any other pair of two-dimensional subspaces. Additionally, the

union of any two four-dimensional invariant subspaces (blue) consists of the whole state space (12345678) (red).

From the definition of the invariant subspace, this hierarchy means that by choosing the initial condition, we can easily restrict the retrieval dynamics of CNN inside of any subspaces shown in Fig. 16.3. This is especially meaningful when CNN performs an itinerant dynamics among all of the six memorized patterns, because we can hierarchically choose any combination of pattern pairs, and eliminate other patterns' output without changing system parameters.

For example, if the dynamics shows chaotic wandering process among six memorized patterns, only one step of operation is sufficient to restrict the orbit in one of the invariant subspaces:

Putting the following condition at time t,

$$\begin{aligned} x_1(t) = x_2(t) = x_3(t) = x_4(t), \\ x_5(t) = x_6(t) = x_7(t) = x_8(t), \end{aligned} \tag{7.11}$$

CNN will keep the future orbit inside of (11115555), and the dynamics only retrieves the patterns $\mathbf{P^1}$ and $\mathbf{P^2}$ intermittently, unless some external inputs destroy this synchronization.

If we put the condition (7.9), the orbit stays inside of (11335577), and the dynamics shows intermittent retrieval among the patterns $\mathbf{P^1}$, $\mathbf{P^2}$, $\mathbf{P^3}$, and $\mathbf{P^4}$.

On the other hand, these restrictions on the orbit can immediately recover its full degree of freedom by adding slight asynchronous noise from external inputs, due to the instability of each invariant subspace in its transverse direction.

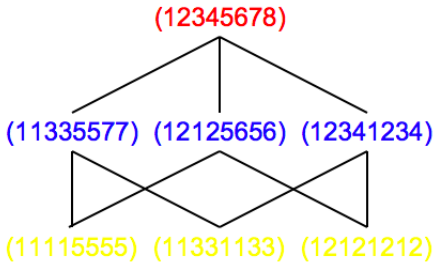


Figure 7.3: **Hierarchical structure of invariant subspaces in CNN.** The spaces of the same color are in the isomorphic relation. The lines denote the inclusion relation regarding the degree of freedom. Among the thirty two possible invariant subspaces detected under the defined W, only seven of them that are important for the interpretation of CNN dynamics are shown.

7.3.3 Wandering range of retrieval dynamics and corresponding hierarchy of invariant subspace

Taking the hierarchical structures in previous section into account, in the same k_r-α plane as the deviation rate, we investigated the wandering range of retrieval dynamics among memorized patterns. This is to judge roughly the localization of the high-dimensional chaos in the hierarchy of invariant subspaces, and refers to their transverse stability/instability.

According to Fig. 7.4, the light green region is where three of the two-dimensional invariant subspace (yellow in Fig. 16.3) are stable in its transverse direction, therefore the orbits fall and stay inside of them. The dynamics can be either stable or chaotic, according to the parameters k_r and α, and can be rigorously judged by the Lyapunov spectrum inside of these subspaces. Note that choosing the appropriate synchronous initial condition, the dynamics can all the way stay inside of these subspaces even if it is not on this light green region, regardless of its transverse stability.

The yellow region in Fig. 7.4 is where the orbits visit two pairs of stored patterns, and not all three of them. Therefore, the three four-dimensional invariant subspace (blue in Fig. 16.3) are assumed to be stable in their transverse direction, competitively absorbing the orbits inside according to the initial condition. This fact was numerically observed as the complete synchronization into

one of these four-dimensional subspaces in this k_r, α region. Since the dynamics is not restricted to any of a single pair of patterns, the three two-dimensional subspaces (yellow in Fig. 16.3) are transversally unstable. The orbits show itinerant dynamics between limited two pairs of patterns.

Finally, the orange region in Fig. 7.4 is where the orbits visit all of the six stored patterns during the retrieval. in this region, both four-dimensional and two-dimensional invariant subspaces (blue and yellow, respectively, in Fig. 16.3) are assumed to be unstable in their transverse directions. Consequently, the dynamics stay in the eight-dimensional whole state space (red in Fig. 16.3), and shows itinerant dynamics among all of the six patterns. Since we can realize the restriction of CNN dynamics into desired invariant subspace by simply changing the initial value, regardless of the parameters k_r and α, this orange region gives us an example of the parameter region where the hierarchical control is fully effective. For the following section, we will focus on the specific set of k_r and α in this region to further investigate the nature of the dynamics and its possible application.

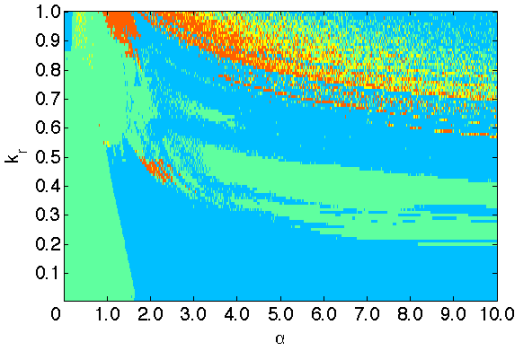


Figure 7.4: **Wandering range of retrieval dynamics among memorized patterns.** The light blue region is where the orbits never retrieve any of six patterns. In the light green region, the orbits visit any of a single pair of patterns. In the yellow region, the orbits visit any of two pairs of patterns. In the orange region, the orbits visit all three pairs of patterns. The wandering range was judged after cutting 5000 steps of transient dynamics, with 10000 steps of retrieval at each k_r and α. Initial values were taken randomly.

7.3.4 Effective Dimension of Dynamics

In this section, we focus on the geometrical localization of orbits with respect to the hierarchical structure of invariant subspaces during chaotic itinerancy. For this purpose, we fix the refractoriness parameters as $k_r = 0.98$ and $\alpha = 1.2$ to situate the system inside of the orange region in Fig. 7.4, and observe both temporal and spatial structure in the phase space. With these fixed parameters, CNN shows chaotic itinerancy among all six stored patterns if the initial value was taken randomly without any synchronization.

To investigate the spatiotemporal structure of retrieval dynamics, we measured the effective dimension ED of a single orbit of CNN outputs $\mathbf{x}(t)$, which is defined as follows [1],

$$ED(\mathbf{x}(t), \delta) = \min\{\dim H_\sigma | H_\sigma \cap B_\delta(\mathbf{x}(t)) \neq \phi\}, \tag{7.12}$$

where $B_\delta(\mathbf{x}(t))$ is the open sphere around $\mathbf{x}(t)$ with radius parameter δ. The effective dimension means the minimum dimension of invariant subspaces that exist in the δ-neighbor of the $\mathbf{x}(t)$, and reflects the geometrical location of the orbit in the hierarchy.

Figs.7.5 and 7.6 show the temporal development of the effective dimension.The initial values were taken asynchronously in Fig. 7.5 and inside of (11335577) in Fig. 7.6. Although the distinctions among the invariant subspaces with the same dimension are lost in this figure, the temporal dynamics shows irregular transition among the three different orders of neighboring subspaces in asynchronous case, that corresponds exactly to the hierarchy in Fig. 16.3. Since the residence rate to the all three levels of effective dimension are dynamically well balanced, chaotic itinerancy in

this parameters seems to be naturally affected by the hierarchical structure in Fig. 16.3 along with their stability, at the same time maintaining the maximum degree of freedom eight. The dynamics with the initial condition in (11335577) shows itinerancy among two-dimensional subspaces inside of (11335577).

To put it more clearly, Figs.7.7 and 7.8 express the same dynamics as in Figs.7.5 and 7.6, respectively, this time as Markov sources with distinction of each invariant subspace as depicted in Fig. 16.3. In Fig. 7.5, the effective dimension eight corresponds to the neighbor of (12345678) in Fig. 7.7, four to one of (11335577), (12125656), (12341234), and two to one of (11115555), (11331133), (12121212). While in In Fig. 7.6, the effective dimension four corresponds to the neighbor of (11335577) in Fig. 7.8, and two to either of (11115555) or (11331133).

With the asynchronous initial condition, the dynamics performs the intermittent search dynamics among all hierarchical structure of invariant subspaces.

When the initial condition is limited in (11335577), The dynamics shows similar searching process but strictly inside of this subspace, although its transverse direction is unstable in (12345678). This result fits to the analysis with the deviation rate and the wandering range in the previous sections. Note that the existence of such dynamics limited under other four-dimensional subspaces are also assured by the symmetry of the system.

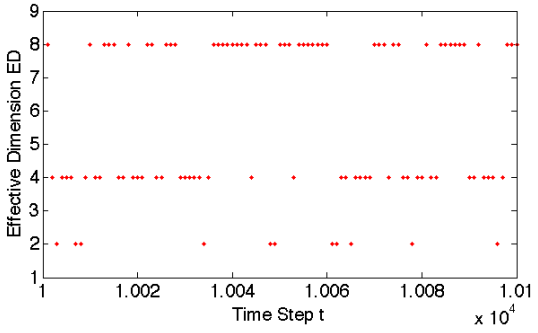


Figure 7.5: **Dynamics of effective dimension.** The initial value was taken inside of (12345678).

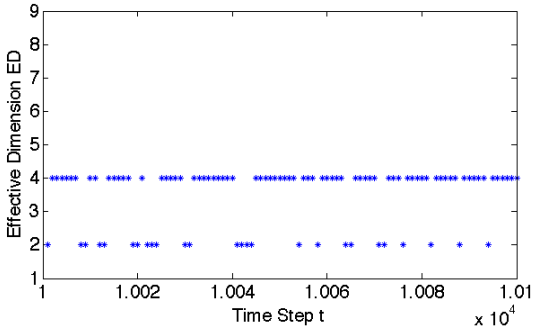


Figure 7.6: **Dynamics of effective dimension.** The initial value was taken inside of (11335577).

7.4 Simulation Result

7.4.1 Single Robot Movement

We simulated the actual movement of the robot in two-dimensional space. Typical autonomously wandering movement is shown in Fig. 7.9. Depending on the restriction imposed on the initial condition, the robot demonstrated three distinctive behaviors. These three types of behaviors are the direct consequence of hierarchical dynamics in CNN. When the initial condition is inside of

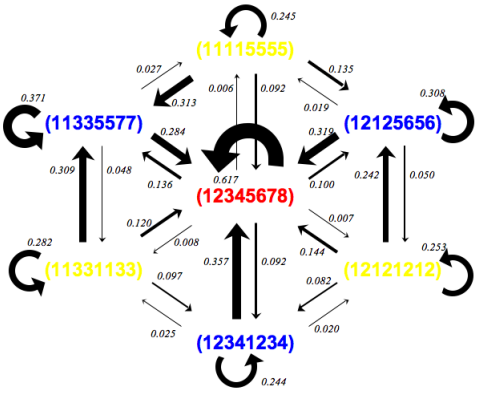


Figure 7.7: **Markov source expression of the effective dimension dynamics.** The initial value was taken inside of (12345678). The arrow width reflects the value of transition probability, which are also displayed nearby.

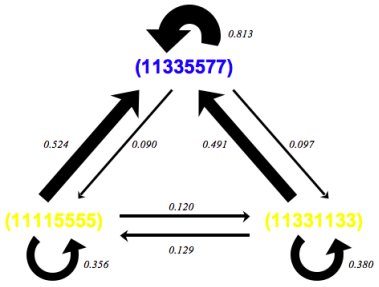


Figure 7.8: **Markov source expression of the effective dimension dynamics.** The initial value was taken inside of (11335577). The arrow width reflects the value of transition probability, which are also displayed nearby.

two-dimensional invariant subspaces, the CNN produces intermittent outputs of one corresponding motor, and the robot can only move to this direction. In the same way, when the initial condition is inside of four-dimensional invariant subspaces, the robot can move to the corresponding one third of the surrounding area. When the initial condition was asynchronous, the robot began to rove irregularly in all of the three directions, similar to random walk.

Hence the fluctuation was produced without any stochastic variable. This randomness caused by the instability of deterministic chaos can be considered as the source of autonomy [62] [63]. In our robot, we realized a way of harnessing chaotic itinerancy with the hierarchical structure of invariant subspaces. Since this structure gives a fundamental mechanism in generating chaotic itinerancy [1], the control is simple and suits well to the nature of the dynamics. We do not need to change any parameters of the system, and the delicate chaotic dynamics is never suppressed nor disturbed externally.

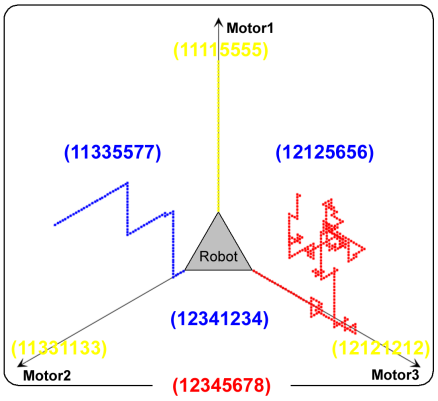


Figure 7.9: **The trajectory of the robot in two-dimensional space.** The coordinates represent the driving direction of the three motors. Possible movement areas are partitioned according to the invariant subspaces where initial conditions of CNN are situated. Three trials with different initial conditions are superimposed. The initial condition was taken inside of (11115555) for the yellow trajectory, inside of (11335577) for the blue trajectory, and inside of (12345678) for the red trajectory. $k_r = 0.98$ and $\alpha = 1.2$.

7.4.2 Response to Environmental Stimulus

Although CNN possesses the orbital instability in chaotic itinerancy state, the coexisting stability toward attractor ruins also plays some essential role. Due to this partial stability, the network possesses spatial correlations in temporal dynamics. This enables the system to retrieve the stored patterns quickly with only a weak input of the patterns fragment [64]. Since this was proved to be faster than a simple random search, the sensitive response can be considered to accompany with some information processing peculiar to chaotic itinerancy state.

To see the utility of this property in the designed robot, we added the following periodic external inputs $\iota_i(t)$ $(1 \leq i \leq n)$ via the Input layer of CNN, as defined in the System Description section.

$$\iota_i(t) = \begin{cases} 0.1 & if\ t = 0 \pmod 4\ and\ 1 \leq i \leq 4 \\ -0.1 & if\ t = 0 \pmod 4\ and\ 5 \leq i \leq 8 \\ 0 & else \end{cases} . \tag{7.13}$$

This means that the network is periodically stimulated with the weak pattern $\mathbf{P^1}$, that corresponds to the motor 1. Here, the sensor input is limited to one direction to investigate the basic response property of the system, but will be generalized in the following section.

As a result, we obtained the chaotic random walk with significant drift in its stimulated attractor direction (see Fig. 7.10). The balance between chaotic fluctuation and the macroscopic directional drift is relevant to the strength of external inputs $\iota(t)$, the period of the stimulation, and the exponential decay parameter of Input layer k_e. Hence we can also qualitatively control

the sensitivity to the inputs by choosing the synchrony of initial condition. The more empirical result of the retrieval ratio of the target pattern with respect to the modulus of input period is shown in Fig. 7.11. Interestingly, the retrieval ratio is not monotonically decreasing with respect to the modulus, and there exist certain sensitive peaks with high retrievals. The existence of such nonlinear response can be considered to relate to the spatiotemporal localization of the orbit in the neighbor of each invariant subspace. Especially the modulus 11 of (11335577) shows in average 97.94 percent of $\mathbf{P^1}$ or $\mathbf{P^2}$ retrievals among the retrieved patterns, which assures almost deterministic response to the intended direction (Only $\mathbf{P^1}$ retrieval ratio is shown in Fig. 7.11).

These results give us the basis of controlling chaotic search of this robot by adding external sensor input to find the intended target in situation with obstacles. This response may also be expected to be robust and efficient in searching with sporadic information [65].

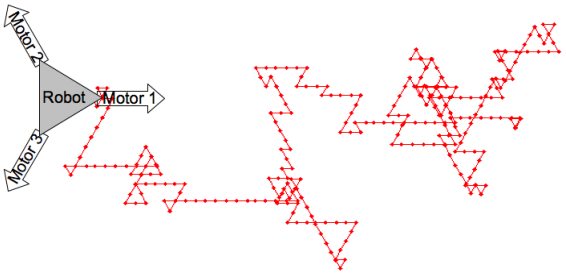


Figure 7.10: **The simulated environmental response with external inputs to Input layer.** The future movement of the robot in two-dimensional space is shown with red trajectory. The motor 1 direction is headed horizontally to the right. The initial condition was taken inside of (12345678). The macroscopic drift is also realizable in other hierarchies of invariant subspaces. $k_e = 0.2$, $k_r = 0.98$ and $\alpha = 1.2$ were chosen.

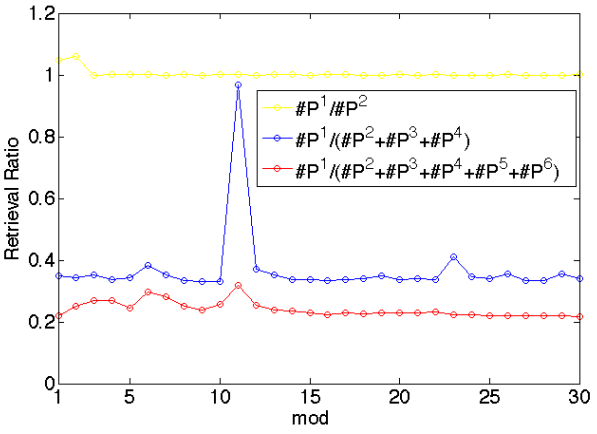


Figure 7.11: **Retrieval ratio of the target pattern $\mathbf{P^1}$ with respect to the other patterns in each invariant subspace.** Horizontal axis: modulus of the input period in equation (7.13). Vertical axis: retrieval ratio of the target pattern $\mathbf{P^1}$ with respect to the other possible retrieval patterns. The red line is inside of (12345678), the blue line is inside of (11335577), and the yellow line is inside of (11115555). After cutting 5000 steps, 100000 steps were used for the calculation. 30 samples were averaged in each modulus.

7.4.3 Simulation with Microsoft Robotics Studio

We have also developed a simulator of the robot with Microsoft Robotics Studio software. The above mentioned characteristics of the dynamics were also realized in the simulated condition with physical laws, such as the existence of inertia and friction.

Furthermore, a simple combination of the investigated dynamics in the different invariant subspaces realized obstacle-avoiding motions. The dynamics was basically synchronized to one of the four-dimensional invariant subspaces that covers the direction to the target, and was temporarily released to the eight-dimensional (12345678) when encountered the obstacle. For example, if the target is situated in the direction between motor 1 and motor 2, then the CNN is synchronized inside of (11335577) (see Fig. 7.9). The robot received constantly the periodic external inputs corresponding to the nearest motor from the target, except after bumping the obstacle, the motor pattern stimulus farthest from the obstacle was chosen for a while.

Examples of the simulated avoiding motions are depicted in Figs. 7.12 and 7.13. Video movies are available at our web-page: http://kappanoid.com/robio2008/

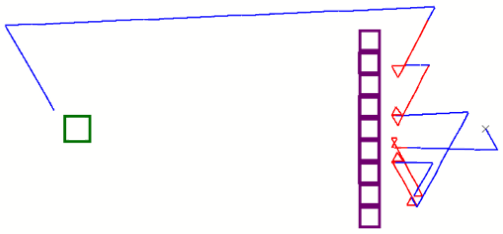


Figure 7.12: **Example of obstacle avoidance.** The cross is the start point of the robot, the violet squares are obstacles, and the green square is the target. The red trajectory is where the CNN orbit is controlled inside of eight-dimensional (12345678), while the blue trajectory is inside of the four-dimensional invariant subspace whose corresponding moving range covers the target direction. Modulus 4 is used for external input period.

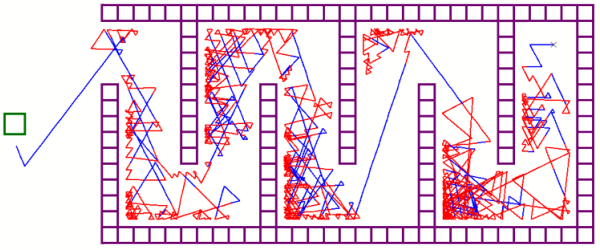


Figure 7.13: **Example of maze task.** The cross is the start point of the robot, the violet squares are obstacles, and the green square is the target. The red trajectory is where the CNN orbit is controlled inside of eight-dimensional (12345678), while the blue trajectory is inside of the four-dimensional invariant subspace whose corresponding moving range covers the target direction. Modulus 4 is used for external input period.

7.5 Further Consideration and Conclusion

7.5.1 Further Consideration

Possible embodiment of the hierarchical chaotic dynamics should be further investigated in other designs of robot.

The hierarchical restriction can also be useful considering collective robot movement. Chaotic itinerancy may also play essential role in maximizing collective information, since similar dynamics is reported to be abundant in actual brain dynamics [21].

The sensitive response to the external input implies useful application in infotaxis [65].

The implementation of an intrinsic learning that reinforces the temporal correlations in the dynamics would be an interesting challenge to derive plasticity in this model [66].

The residence time distribution on each attractor ruin of chaotic itinerancy is reported to follow exponential distribution [67]. Since intermittent foraging behavior observed widely in animals is known to be optimized with exponential law in stochastic model, the CNN may also have the capacity to optimize such saltatory search [68].

7.5.2 Conclusions

We have investigated the spatiotemporal dynamics of CNN, and developed a general framework of controlling chaotic itinerancy by making use of the existing invariant subspaces.

A possibility of implementation to robot was simulated. Obstacle avoidance was realized by balancing the sensitive response to external signals with the autonomy of the system, and by tactical synchronization into purposely-designed invariant subspaces.

Chapter 8

Optimal Intermittent Search with Chaotic Roving Robot and Infotaxis by Collective Interaction

Abstract

We consider the application of chaotic roving robot with CNN designed in previous chapter for collective infotaxis. 2-phase sporadic information search based on animal saltatory behavior is expanded to 2-dimensional surface. Statistical description of chaotic itinerancy is then used to fit the formalization of 2-phase search. The contact graph structure between robots is investigated as the basis of information sharing. The efficiency of collective infotaxis is evaluated on a simulator basis.

Keywords: 2-phase sporadic information search, infotaxis, collective interaction, information sharing

Methodology: Design of a single robot 2-phase search as a completely deterministic dynamical system based on the probabilistic analysis of the previous chapter → Analysis of the optimal search parameter on simulator basis with probabilistic formulation → Design of collective infotaxis as an ensemble of completely deterministic dynamical systems and its simulation → Development of a search evaluation theory with information geometry

8.1 Introduction

The hidden target search has been one of the essential task for animals to survive in natural environment. Experimental observations revealed that the evolution invented an intermittent search strategy with two phases: the active search phase and the fast ballistic motion [69][70][71]. The 2-phase search in several species can be experimentally classified only in two classes depending on the inequality of the duration between two phases. This implies that there exists an global and universal optimal solution for the sparse target search. Therefore, the model of the saltatory search behavior is expected to be useful for the searching robot in a similar condition.

We briefly summarize here the one-dimensional 2-phase search model for further application in searching robot [72]. The model of saltatory search behavior proposed by Bénichou et al. consists of two states. In the *search phase*, the searcher explores the targets in its vicinity with a relatively slow diffusive movement. This process is approximated as a continuous random walk with diffusion coefficient D. In the *motion phase*, the searcher moves relatively fast and do not detect any target. This is a repositioning move and modeled with a ballistic motion of constant velocity v. The two phases stochastically change with each other with given parameters f_S and f_M. Here, f_S represents the changing probability per unit time from the search phase to the motion phase, and f_M is that of from the motion phase to the search phase.

Assuming the uniform target density $1/L$ in one-dimensional search space, modeling of the 2-phase search with the backward Chapman-Kolmogorov differential equation gives the following solution of the average search time $< t >$ in the low density limit $L >> \frac{v}{f_M}, \sqrt{\frac{D}{f_S}}, \sqrt{\frac{f_M D}{f_S v}}$:

$$< t >= \frac{L}{2D}\left(\frac{1}{f_S}+\frac{1}{f_M}\right)\frac{\tau f_M^2 + 2f_S}{\sqrt{\tau f_M^2 + 4f_S}}. \tag{8.1}$$

Where $\tau = \frac{D}{v^2}$. The linear dependence of $< t >$ on the typical intertarget distance L is known to be more efficient than purely diffusive strategy without the moving phase.

The average search time $< t >$ becomes a single minimum with respect to the following constraint between f_S and f_M,

$$f_S = f_S^{max}, \tag{8.2}$$

$$f_M^5 + \frac{6}{\tau}f_M^3 f_S - \frac{8}{\tau^2}f_S^3 = 0, \tag{8.3}$$

where f_S^{max} is the upper bound of f_S defined as the minimum time required for information processing by sensory organs.

The minimum depends on the asymptotic between f_S^{max} and $\frac{1}{\tau}$ and give two different forms: In case $f_S^{max} << \frac{1}{\tau}$, the optimal frequencies are

$$f_S = f_S^{max}, \tag{8.4}$$

$$f_M = \left(\frac{4}{3\tau}\right)^{\frac{1}{3}} f_S^{\frac{2}{3}}. \tag{8.5}$$

This solution is called the regime $\mathcal{S}$ (for search) since $f_S < f_M$ holds, meaning the searcher spends more time searching than moving.

In case $f_S^{max} >> \frac{1}{\tau}$, the optimal frequencies are

$$f_S = f_S^{max}, \tag{8.6}$$

$$f_M = \left(\frac{2\sqrt{2}}{\tau}\right)^{\frac{1}{3}} f_S^{\frac{3}{5}}. \tag{8.7}$$

This solution is called the regime $\mathcal{M}$ (for move) since $f_S > f_M$, and the searcher spends more time moving than searching.

8.2 2-phase Search with Chaotic Roving Robot

In this section, we develop a framework to realize the 2-phase intermittent search robot with the use of a chaotic neural network (CNN). The above described 2-phase search model is a stochastic model with exponential residence time distribution for each phase. On the other hand, the CNN is a completely deterministic system. However, the orbits in chaotic itinerancy state have the irregularity that can be collectively characterized as a probabilistic distribution. This formalization with the invariant density leads us to investigate the correspondence between the stochastic model and the deterministic model with chaos.

The residence time distribution on each effective dimension is shown in Fig. 8.1. They all converge to exponential distributions as the long-term time average. Since the stochastic model for 2-phase search is also defined by an exponential residence time distribution for each phase, it is possible to realize the same kind of dynamics with CNN in terms of the long-term statistics.

As for the temporal structure of the orbit, we should consider the temporal dependency of the deterministic CNN. Though, the chaotic itinerancy states expresses the dependency on the past states during only a short time period. It is because of the long term instability of chaotic orbits. This means that the chaotic itinerancy can be approximated with higher order Markov process in

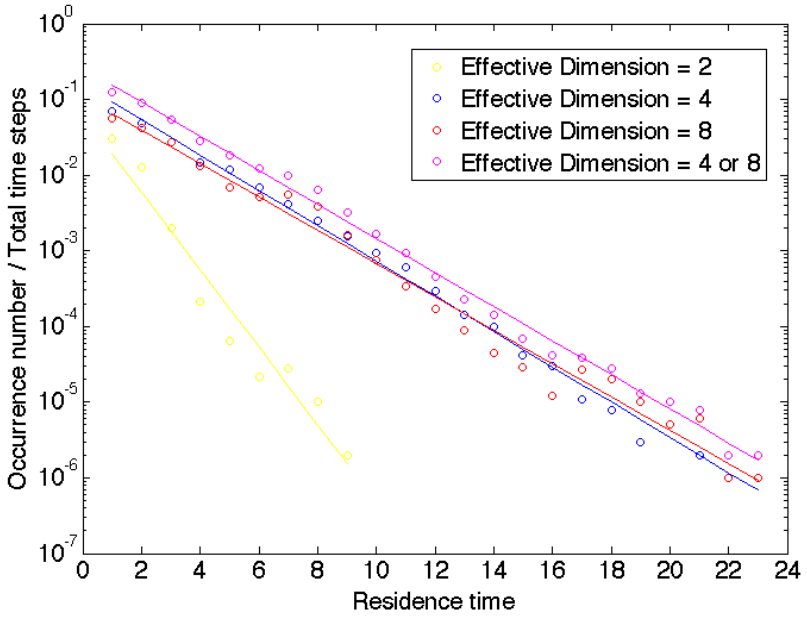


Figure 8.1: **Residence time distributions of each effective dimension in CNN.** The distributions are obtained from 1000,000 steps of output from CNN with $k_r = 0.98$, $\tau = 1.2$.

terms of the predictability.The effective scale of temporal dependency is related to the Lyapunov exponent in simple chaotic systems.

Suppose that the chaotic itinerancy in chosen parameters is approximately equivalent to a k-th order Markov process M defined as follows:

$$P(t+k) = M(P(t), \cdots, P(t+k-1)), \tag{8.8}$$

where $P(t) = (P_1(t), \cdots, P_6(t)) \in [0,1]^6$ is the joint probability distribution, which represents the appearance rate of the six patterns $\mathbf{P^1}$, $\mathbf{P^2}$, $\mathbf{P^3}$, $\mathbf{P^4}$, $\mathbf{P^5}$, $\mathbf{P^6}$ at time t.

We now consider the expression of (8.8) with Perron-Frobenius operator $F : [0,1]^8 \to [0,1]^8$ of CNN $\Phi : \mathbf{R^8} \times \mathbf{R^8} \to \mathbf{R^8} \times \mathbf{R^8}$ as follows:

$$p(t+k) = F^k(p(t)), \tag{8.9}$$

$$p(t) = \lim_{N\to\infty} \frac{1}{N} \sum (\eta(\mathbf{t}), \zeta(\mathbf{t})), \tag{8.10}$$

where $p(t)$ is the sum of the state $(\eta(\mathbf{t}), \zeta(\mathbf{t}))$ of CNN at time t over a infinite number of initial conditions. Since $P(t)$ and $p(t)$ have a correspondence between them, we formally describe this relation as $P(t) = fp(t)$. Note f is not necessary one-to-one projection, because it is invariant under the permutation between η_i and ζ_i with arbitrary $1 \leq i \leq 8$.

The invariant density p_{inv} of F is then defined as follows:

$$p_{inv} = F(p_{inv}), \tag{8.11}$$

where

$$fp_{inv} = M(fp_{inv}, \cdots, fp_{inv}). \tag{8.12}$$

The distribution $P_{inv} = fp_{inv}$ is the stationary distribution of the k-th order Markov process.

Table 8.1: **Transition probability matrix between the six patterns of CNN.** The transition matrix as a first-order Markov process is calculated from 1000,000 time steps of CNN retrievals. The last line is the overall appearance time for each pattern.

	$\mathbf{P^1}$	$\mathbf{P^2}$	$\mathbf{P^3}$	$\mathbf{P^4}$	$\mathbf{P^5}$	$\mathbf{P^6}$
$\mathbf{P^1}$	0.5652	0.0248	0.1014	0.1043	0.1026	0.1044
$\mathbf{P^2}$	0.0249	0.5650	0.1044	0.1013	0.1044	0.1022
$\mathbf{P^3}$	0.1037	0.1023	0.5657	0.0256	0.1023	0.1012
$\mathbf{P^4}$	0.1018	0.1037	0.0254	0.5651	0.1018	0.1016
$\mathbf{P^5}$	0.1026	0.1011	0.1006	0.1022	0.5640	0.0251
$\mathbf{P^6}$	0.1014	0.1028	0.1022	0.1012	0.0246	0.5652
#Appearance	156841	156676	156421	155943	154953	155588

To obtain the concrete P_{inv}, we investigated the transition matrix between the six patterns of the CNN dynamics. The result for the first-order Markov process is described in Tab. 8.1. As a property of dynamical system, the asymmetric short-time dependency exists, though since the overall appearance rate of the defined six patterns in CNN is uniform, the stationary distribution is assumed to exist as

$$P_{inv} = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right). \tag{8.13}$$

This means that the stationary distribution of the chaotic itinerancy state is equivalent to that of the symmetric random walk on a triangular lattice. The same statistics also holds for $ED = 4$ or 8 and $ED = 2$ among the nearest patterns of the orbit. We will later use this result to obtain the optimal solution in one-dimensional projection of 2-phase search with CNN.

We now consider the implementation of the 2-phase search in a roving robot with CNN. As in the previous chapter, we control the three motors outputs with the six memorized patterns. Since the residence time of each effective dimension follows exponential distribution, we can find a natural correspondence with the search and moving phase. We define that the CNN is in the search phase if the effective dimension is 4 or 8, and in the moving phase if 2. We fitted the residence time distribution of the effective dimension 4 or 8 as $P_S(t_r)$, and 2 as $P_M(t_r)$, respectively, as follows:

$$P_S(t_r) = A_S e^{a_S t_r}, \tag{8.14}$$
$$P_M(t_r) = A_M e^{a_M t_r}, \tag{8.15}$$

where t_r is the residence time, and A_S, a_S, A_M, and a_M are the fitting parameters. Comparing with the definition of f_S and f_M, we have

$$f_S = 1 - e^{a_S}, \tag{8.16}$$
$$f_M = 1 - e^{a_M}. \tag{8.17}$$

Since the dynamics of CNN with the chosen parameters $k_r = 0.98$ and $\alpha = 1.2$ have longer residence time in the search phase (effective dimension 4 or 8), we consider the optimization in the regime $\mathcal{S}$. We introduce the control parameters γ_S and γ_M defined as follows, to realize the distributions for the optimal 2-phase search.

$$f_S = 1 - e^{\gamma_S a_S}, \tag{8.18}$$
$$f_M = 1 - e^{\gamma_M a_M}. \tag{8.19}$$

This corresponds to elongate the residence time distributions of the two phases with the rate of $1/\gamma_S$ and $1/\gamma_M$.

We define the control sequence of the robot on two-dimensional surface as follows:

Table 8.2: **Correspondence between the nearest pattern and motor output in search phase.** Motor numbers with the positive/negative signs of the output direction are listed in the second line. For example, if the nearest pattern is $\mathbf{P^1}$, the robot moves to the direction of the motor 1. If the nearest pattern is $\mathbf{P^2}$, the moving direction is the opposite to the motor 1 by the inverse rotation. In the search phase, the six moving directions are defined with three motors to have a correspondence between the chaotic itinerancy state and the diffusion of Markov process on a triangular lattice.

Nearest Pattern	$\mathbf{P^1}$	$\mathbf{P^2}$	$\mathbf{P^3}$	$\mathbf{P^4}$	$\mathbf{P^5}$	$\mathbf{P^6}$
Motor Output	+1	-1	+2	-2	+3	-3

1. At each time step t, judge the effective dimension (ED) and define the search mode. If $ED = 4$ or 8,then the robot is in the search phase and is able to detect targets in its vicinity. If $ED = 2$, then the robot is in the motion phase and is unable to detect any target.

2. Judge the nearest pattern $\mathbf{P^j}$ $(1 \leq j \leq 6)$ with respect to the Euclidian distance $\frac{1}{8}\sqrt{\sum_{i=1}^{8}(x_i(t) - p_i^j)^2}$. If the robot is in the search phase, the motor output is defined as Tab. 8.2. If the robot is in the motion phase, the motor output is defined as Tab. 8.3.

We defined the six moving directions for the search phase, and three directions for the motion phase, as depicted in Fig. 8.2 and 8.3. The motion phase is limited to three directions to avoid round trip. The dynamics of the search phase corresponds to the diffusion of the k-th order Markov process on a triangular lattice. Furthermore, from (8.13), the chaotic itinerancy state is equivalent to a symmetric random walk in terms of the stationary distribution. The diffusion coefficient D of a random walk is given by the variance of the stationary distribution. Defining the moving speed of the robot during the search phase as v_S, the variance of the symmetric random walk on a triangular lattice is given by $\frac{2}{3}v_S^2$ for each of three axis.

We consider now the projection of the search dynamics on triangular lattice to one of the three axis, to obtain the optimal solution in one-dimensional search. For simplicity, we define only the one-dimensional moving direction of the robot in the motion phase with the speed $v_M = \nu v_S$. The motion is bi-directional. (In the actual robot, the different three directions have to be considered.) Then the parameter $\tau = \frac{D}{v_M}$ becomes as follows:

$$\tau = \frac{2}{3\nu^2}. \tag{8.20}$$

Defining $f_S^{max} = 1 - e^{\gamma_S a_S}$ with the control parameter γ_S, and in case $\nu \geq 10$, for example, the following condition of the regime $\mathcal{S}$ holds:

$$\frac{1}{\tau} = \frac{3\nu^2}{2} \geq 150 >> 1 > f_S^{max}. \tag{8.21}$$

The optimal solution of the control parameter γ_M for the one-dimensional 2-phase search with CNN is therefore given by

$$f_S^{max} = 1 - e^{\gamma_S a_S}, \tag{8.22}$$

$$\gamma_M = \frac{1}{a_M}\log\left[1 - \left(\frac{4}{3\tau}\right)^{\frac{1}{3}}(1 - e^{\gamma_S a_S})^{\frac{2}{3}}\right], \tag{8.23}$$

with the constraint $f_S < f_M$ that is translated as $\gamma_M > \frac{a_S\gamma_S}{a_M}$. This is the analytical optimal solution for the simplified one-dimensional 2-phase search with the symmetric random walk whose stationary distribution is equivalent to that of the chaotic itinerancy state.

The obtained theoretical solution in one-dimensional space actually contains some problems to be directly applied to the 2-phase search of the robot in two-dimensional surface. First, the

Table 8.3: **Correspondence between the nearest pattern and motor output in motion phase.** Motor numbers with the positive/negative signs of the output direction are listed in the second line. In the motion phase, we do not define the inverse direction of each motor so that the dynamics become irreversible ballistic motions.

Nearest Pattern	$\mathbf{P^1}$	$\mathbf{P^2}$	$\mathbf{P^3}$	$\mathbf{P^4}$	$\mathbf{P^5}$	$\mathbf{P^6}$
Motor Output	+1	+1	+2	+2	+3	+3

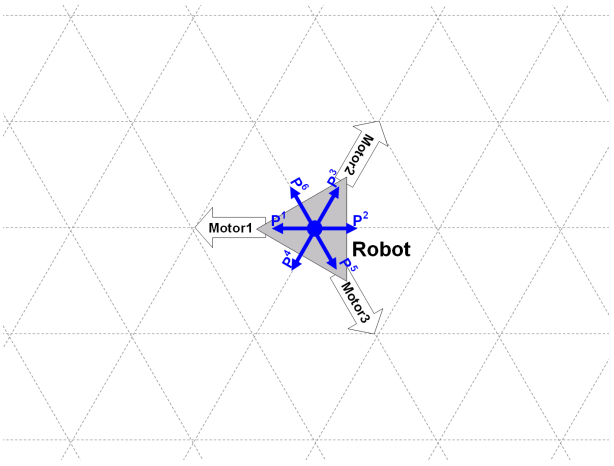


Figure 8.2: **Definition of moving directions with respect to the nearest pattern during search phase.** The side length of the triangular lattice corresponds to the moving speed v_S.

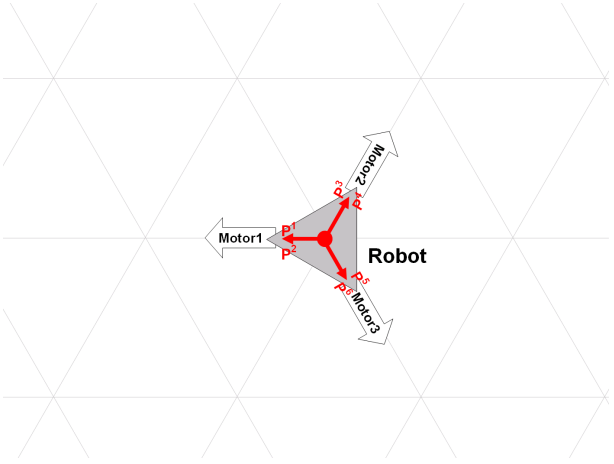


Figure 8.3: **Definition of moving directions with respect to the nearest pattern during motion phase.** The side length of the triangular lattice corresponds to the moving speed v_M.

robot actually moves to three directions in the motion phase to cover the searching surface, and the one-dimensional projection of v_M changes its value and directions. Next, we should consider that in the actual orbit of chaotic itinerancy, there exists high order temporal dependency that is averaged and lost in steady-state distribution. The approximation with the diffusion coefficient contains corresponding fluctuations in a short time period. Therefore, the relation of the long-term average and the temporal dynamics is not necessary straight.

Since these problems are too difficult to solve theoretically, we expect to obtain the optimal solution by a numerical simulation. We simulated the 2-phase search on two-dimensional surface with six directions in the search phase and three directions in the motion phase. We set the parameters $\gamma_S = 0.1$ and $\nu = 12.5$, latter to satisfy the condition of the regime $\mathcal{S}$. The temporal change of the robot speed is shown in Fig. 8.4. We located the targets as information sources with regular distance $L = 200$, and simulated the 2-phase search. This value of L is not sufficient to satisfy the low density limit $L >> \frac{v_M}{f_M}$ in one-dimensional model, but is sufficiently sparse in two-dimensional surface. The vicinity of the robot for the targets detection was set as 3. The result is shown in Figs. 8.5 and 8.6. In case of the two-dimensional search, there exists an optimal peak around $1/\gamma_M = 9$.

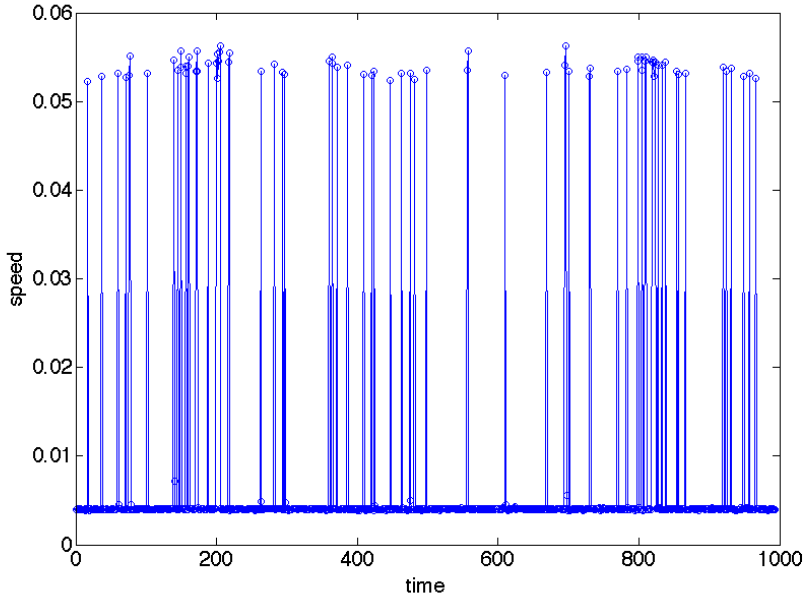


Figure 8.4: **2-phase search speed of the simulated robot.** The speed [m/s] of the up state corresponds to v_M, while that of down state corresponds to v_S. $\nu = 12.5$, $\gamma_S = 1$, $1/\gamma_M = 9$. 1 step of CNN is expressed as 100 time steps of the robot.

8.3 Contact of Robots in Collective Movement with Optimal Intermittent Search

We consider from this section the collective dynamics of the robots in view of application to global collective search and information sharing. We assume the collective search of sporadic information sources in a wide or complicated environment that is beyond the capacity of the single robot search, and define the collective interaction as the local interaction between robots. Local interaction is set to have exchange of search history between robots, so that to deploy more robots to where there exist more information sources. We expect here the bottom-up emergence of collective information sharing from the local interaction.

The benefit of local interaction can be summarized as follows: First, the emerged collective dynamics is the result of the local interaction between robots and the environment. Therefore, we

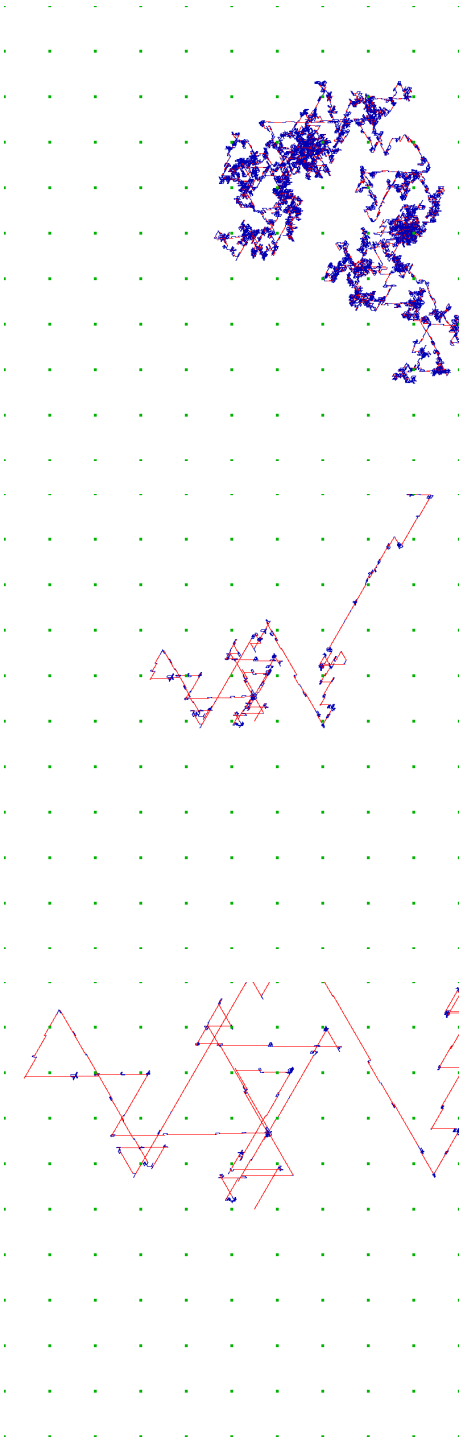


Figure 8.5: **Example of the searching orbit.** Top: $1/\gamma_M = 2$. Middle: $1/\gamma_M = 9$. Bottom: $1/\gamma_M = 22$. The blue orbits are in the search phase, while the red orbits are in the motion phase. The targets are depicted as green square points, and are located on a square lattice with the intertarget distance $L = 200$. $1/\gamma_M = 9$ is the numerical optimal.

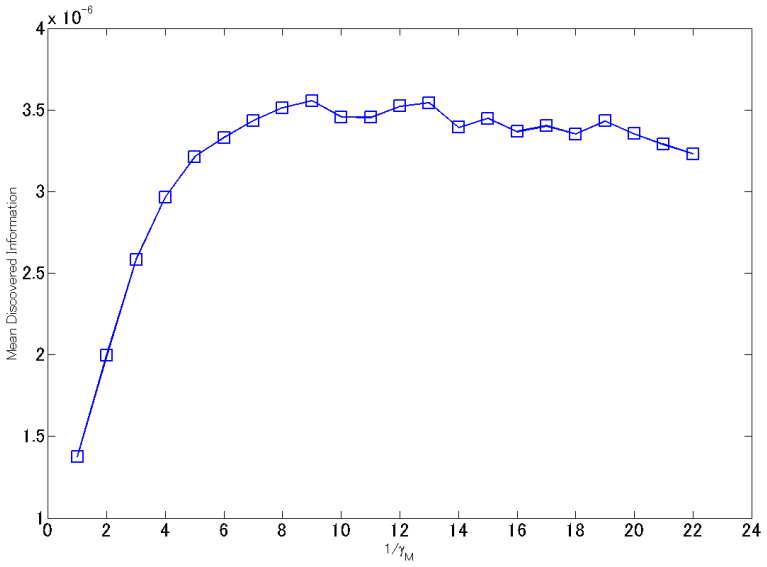


Figure 8.6: **2-phase search result of roving robot in regime $\mathcal{S}$.** Horizontal axis: $1/\gamma_M$. Vertical axis: Mean discovered information. The means of 30 different orbits are depicted. The discovery of the hidden target during the search phase is counted as the information. An optimal peak exists around $1/\gamma_M = 9$. The upper bound of $1/\gamma_M$ is 22.79 in regime $\mathcal{S}$.

do not need *a priori* the global information on the searching space. The group of the robots are expected to fit autonomously to its structure. This is a strong point when we do not know about the optimal searching strategy in a real complicated environment, especially in case the information sources move temporarily. Second, the local interaction can lessen the cost of global collective search. In a wide environment, the centralized control from outside of the robots requires high-power wireless network. Though in practice, conventional Wi-Fi devices, for example, are limited to guarantee the communication of at most 50-100 m distance without barrier. The use of local interaction facilitate the implementation of each robot with the existing technology.

In this section, we simply investigate the contact structure of the emerged collective dynamics from a local interaction rule. Actual simulation of information search is developed in the following section. The local interaction is defined as follows:

1. For each robot, in search phase, at each time step t, scan the neighbor and judge if there exist other robots. This scanning assume the detection of other robots by Wi-Fi network. If there are any, put external input pattern of the nearest direction to CNN. This range is assumed to relate the strongness of the Wi-Fi connection between the robots. We consider only the nearest robot for one-to-one communication channel. We call this additional phase as "communication phase".

2. For each robot, in communication phase, at each time step t, scan the neighbor and judge if there exist other robots that are sufficiently close to establish communication. This range is assumed to relate the strongness of the Wi-Fi connection between the robots. If there are any, establish a contact for further information exchange.

We first simulated the group of robots with this local interaction rule without information sources in the environment. This is to investigate the nature of the emerged collective dynamics without environmental factors. The simulation is performed using the Microsoft Robotics Studio software. To avoid the global divergence of robots, the searching space is confined with a circular wall of 100 m radius. The scan range of the communication phase is 50 m, the establishment of communication is judged by the physical contact between robots. The size of the robots are 80 cm.

The result of the collective dynamics of 10 and 20 robots are shown in Figs. 8.7 and 8.8. Sensitive response to external input of CNN efficiently worked to form contact structure between robots. In a long term superposition, the robots establish almost all-to-all connection, which guarantee the global sharing of information in the long run (Fig. 8.7).

In shorter term, mid-scale hub structures are dominant with large fluctuation for both 10 and 20 robots (Fig. 8.8). The robots are in average keeping mid-degree connection with each others, which would assure the fast and stable spreading/sharing of information in the given confined scale.

The unconfined setting also showed similar statistics, but can lose some robots too far out of others due to the spacial divergence.

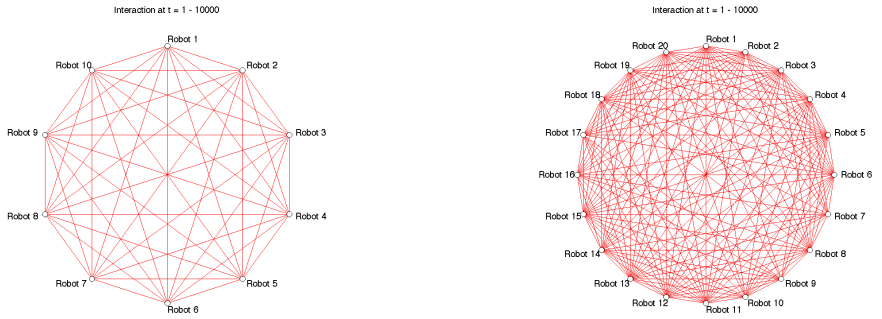


Figure 8.7: **Total communication graph between robots in confined space.** Left: Communication graph between 10 robots during 100000 steps of CNN in confined circle space with radius=100m. Right: Communication graph between 10 robots during 100000 steps of CNN in confined circle space with radius=100m.

8.4 Infotaxis in Information Landscape with Collective Intermittent Search

We apply the collective dynamics of the robots with communication phase to autonomously fit to the information landscape of the environment.

We defined the information gradient dynamics by the superposition of three gaussian distributions that rotate temporarily. Although the 2-phase search optimization is conditioned under sporadic information search, for simplicity we use the continuous information gradient to evaluate its efficiency in averaged term.

By combining the intermittent search and information sharing via collective interaction, we aim to adapt the spacial distribution of the roving robots to the gradient of information in the environment. The group of robots is expected to concentrate and temporally follow the dynamical information landscape.

Figs.8.9 and 8.10 shows schematic representation of the algorithm used in this section. This time, we use the external input to follow the detected information gradient. The algorithm is defined as follows:

1. During search phase, sense information in the environment.

2. After the search phase,communicate with other robots within communication range (50 m) and exchange history of sensed information with global coordinates and time.

3. Calculate gradient vectors using all records within local spatial range (50 m neighbor) with time-dependent exponentially decreasing weight. Decide the external input direction closest to the steepest ascending gradient vector and proceed to motion phase.

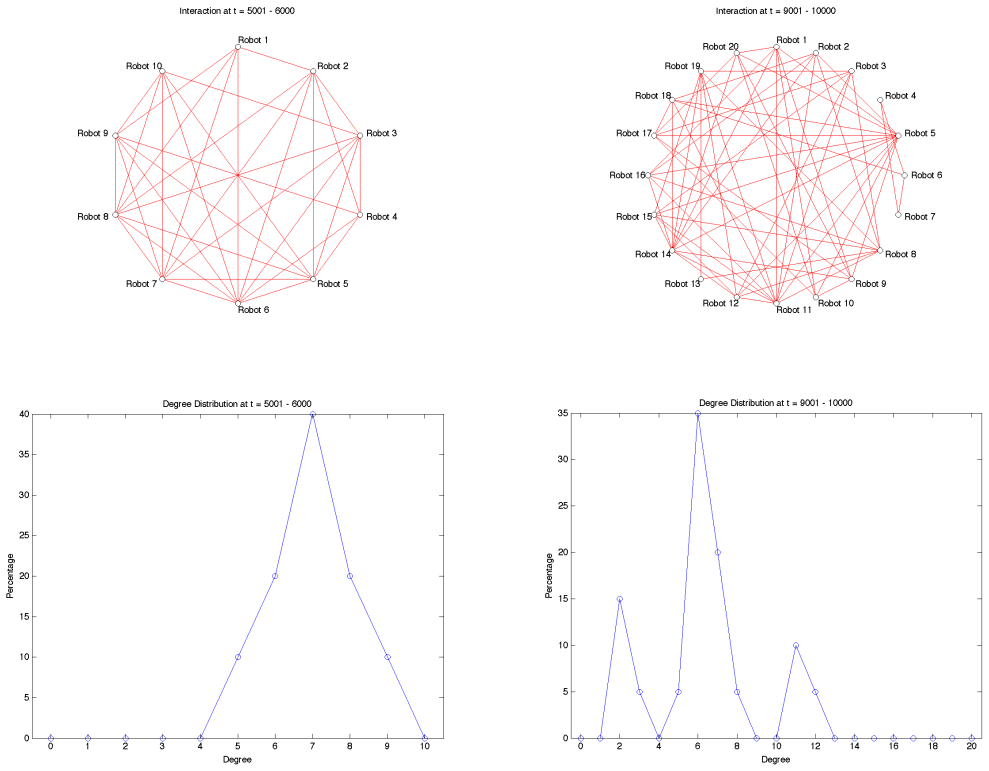


Figure 8.8: **Examples of communication graph and its degree distribution between robots in confined space.** Left Up: Communication graph between 10 robots during 50001-60000th steps of CNN in confined circle space with radius=100m. Left Down: Degree distribution of upper graph. Right Up: Communication graph between 10 robots during 90001-100000th steps of CNN in confined circle space with radius=100m. Right Down: Degree distribution of upper graph.

The global coordinates of the steepest direction (x_M, y_M) with time-dependent weight is defined as follows by calculating $\mathbf{v_M}$. The time dependency is to avoid the persistence of past memory in order to adapt to the dynamical change of information landscape.

$$\mathbf{v_{self}} = (x_{self}, y_{self}, t_{self}, I_{self}), \tag{8.24}$$

$$\mathbf{v_{max}} = (x_{max}, y_{max}, t_{max}, I_{max}) = \max_{\exp(\mu\cdot(t_{self}-t))\cdot I}\{\mathbf{v}\}, \tag{8.25}$$

$$\mathbf{v_M} = (x_M, y_M, t_M, I_M), \tag{8.26}$$

$$\mathbf{v_M} = sgn(I_{max} - I_{self})(\mathbf{v_{max}} - \mathbf{v_{self}}) + \mathbf{v_{self}}, \tag{8.27}$$

where $\{\mathbf{v}\}$ are the 4-dimensional vectors $\{(x, y, t, I)\}$ within communication range, with 2-dimensional global coordinates (x, y), time t, and local information I. $\mathbf{v_{self}}$ is the coordinates of present state of the robot, $\mathbf{v_{max}}$ is those having the maximum I in the past history of communication range with time-dependent penalty weight $\exp(\mu \cdot (t_{self} - t))$. $\mu \geq 0$ is the control parameter of memory decay, which becomes larger to set the optimal as the change of information landscape becomes faster.

This control algorithm only controls the moving direction of motion phase, and conserves perfectly the residence time statistics of the 2-phase search shown in Fig. 8.5 (Middle). It is therefore the optimal for sporadic information search assumed in actual situation.

An example of the results of this algorithm is shown in Fig. 8.11. The robots succeeded to detect the peaks of the information gradient by the collective sharing of search history. A single robot with the proposed algorithm can climb up the information gradient by itself, but the collective sharing of information augments the resolution of the collective grid and realizes quicker search and sharper localization around the information peaks. Figs. 8.12 show an example of the temporal dynamics of sensed information in single robot and collective search. The communication between robots augment in average the quickness of search.

The communication radius also gives the resolution of local minimum avoidance. Local minimums of information landscape less than the communication radius is easily overcome by collective sharing of information (Fig.8.13). Even though the communication radius is limited by hardware (such as Wi-Fi network communication range), the calculation of the gradient vectors in shared information grid is not restricted by the communication manner. Once the sensed information is collectively shared, each robot can independently decide the local spatial range of information grid to decide the next moving direction. Therefore, the algorithm is also adjustable to local minimum problem according to the condition of the environment.

Dynamical information landscape is also traceable by adjusting the temporal decay parameter μ proportional to the changing speed of the information.

8.5 Implementation Plan -Toward the Enhancement of Human Interaction-

Application in real world of the roving robots has many possibilities. The collective adaptation to the information landscape will realize a new way of information presentation in interaction with human population. For example, in public event, information boards can be carried by the robots and automatically distributed to where there are more people. The distribution of information according to the distribution of humans will create the positive loop of concentration between them. Gathering to the populated place to find the information is natural to human behavior.

Fig. 8.14 shows the virtual project during the world exposition Expo 2010 Shanghai China. The robots in the site are deployed automatically according to the human population and inform the people about the on-going events as well as showing the guide map.

Other possibilities include to carry illumination and audio instrument on the robot, and create an adaptive atmosphere according to the temporal location of the people in public space such as outdoor cafeteria or bar. Fig.8.15 shows another virtual project in open-air cafe.

We are planning to seek for the new way of interaction with and among robots through emergent behavior to better enhance human interactions.

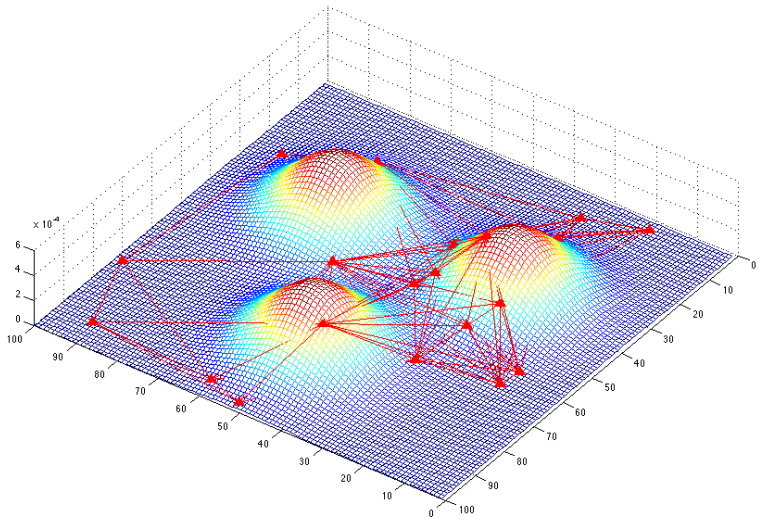


Figure 8.9: **Collective estimation of information landscape by constructing local communication grid between robots.** The robots (red triangles) establish communication with other ones within local communication range (red lines) and exchange information they found during the search phase. Based on these shared information, the robots climb up the information gradient approximated by the collective grid. The global information landscape (mesh grid) is not given explicitly.

8.6 Multi-scale Evaluation of Collective Infotaxis with Respect to the Actual Information Landscape

We investigate a theoretical framework to evaluate the efficiency of the collective infotaxis in actual environment. Actual distribution of information such as human crowd is dynamic and varies both in spatial and temporal scales. The quantification of the cost and benefit is an important issue to evaluate the system.

We consider a short time period for the real-time evaluation, and compare the information and robots distribution with the use of approximation with Gaussian mixture distribution. The actual information landscape can be measured with the use of other empirical technology such as remote sensing. Suppose the actual distribution of information and those sensed by a collective grid of robots are approximated as follows:

$$p^I(\mathbf{x}, j; \theta^I) = \sum_{j=1}^{J} \pi_j^I N(\mu_j^I, \mathbf{\Sigma}_j^I), \tag{8.28}$$

$$p^R(\mathbf{x}, j; \theta^R) = \sum_{j=1}^{J} \pi_j^R N(\mu_j^R, \mathbf{\Sigma}_j^R). \tag{8.29}$$

Where $\theta^I = \{\pi_j^I, \mu_j^I, \mathbf{\Sigma}_j^I | 1 \leq j \leq J\}$ and $\theta^R = \{\pi_j^R, \mu_j^R, \mathbf{\Sigma}_j^R | 1 \leq j \leq J\}$ are the parameters of Gaussian mixture distribution with J components. The $\pi_j N(\mu_j, \mathbf{\Sigma}_j)$ represents the j-th component, which is a 2-dimensional normal distribution with mean vector μ_j and covariance matrix $\mathbf{\Sigma}_j$ multiplied by component weight π_j ($\sum_{j=1}^{J} \pi_j = 1$).

The total efficiency of the collective infotaxis can be measured by temporally integrating the following minimum KL divergence $D[\cdot : \cdot]$ between p^I and p^R as follows:

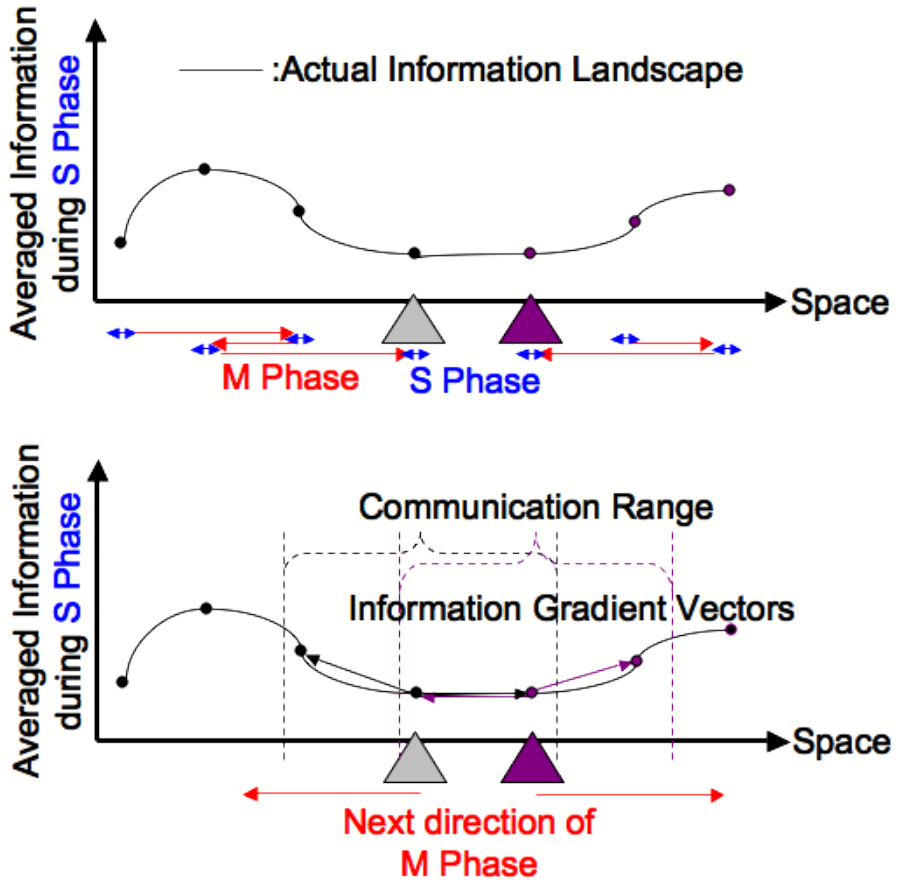


Figure 8.10: **Schematic representation of the algorithm in collective infotaxis.** Top: During the search phase, each robot calculates the local information sensed in the environment. Bottom: At the communication phase after each search phase, each robot detects other robots in the defined communication range, and exchange all history of sensed information. Each robot calculates the information gradient vectors using these shared records, and decide the next moving direction of motion phase to the steepest ascending direction in collective information grid.

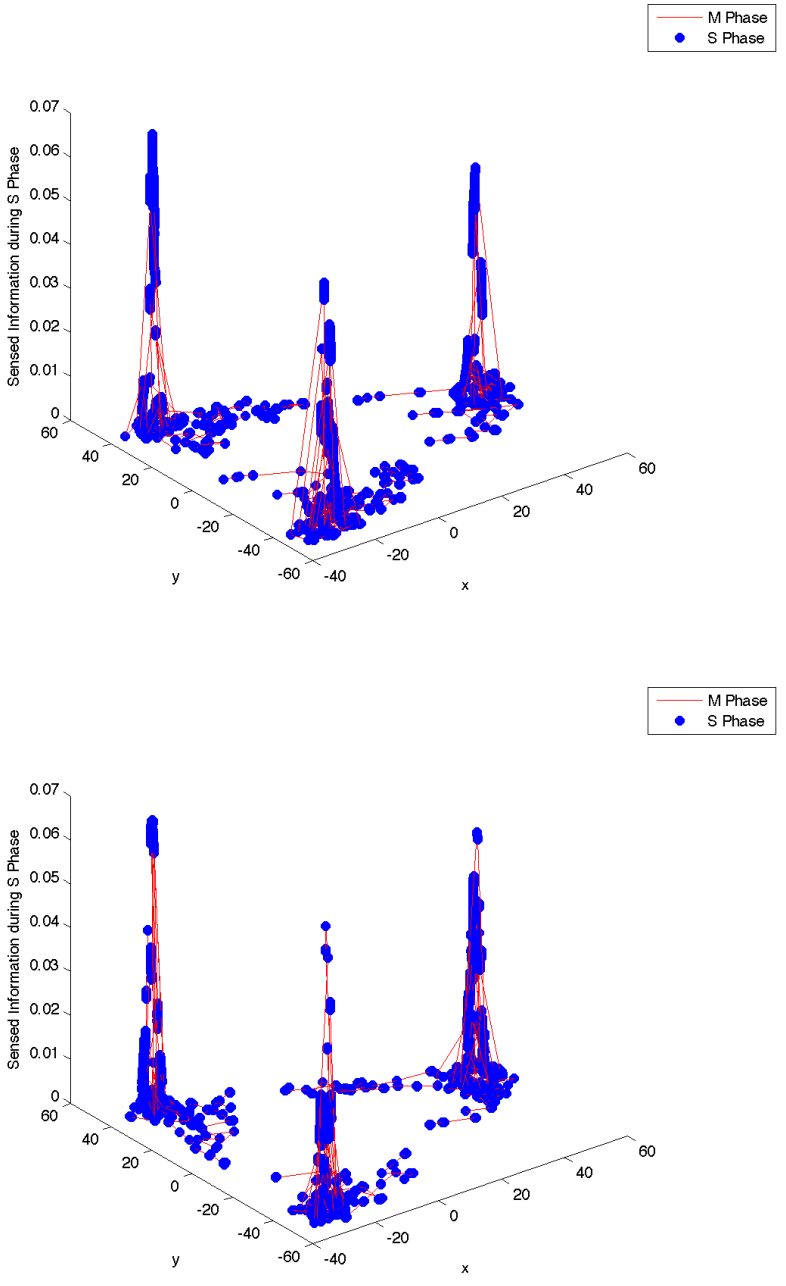


Figure 8.11: **Example of collective infotaxis.** Top: Superposition of single robot search without communication. Bottom: 10 robots collective search. In collective search, the robots communicate with 50 m communication range and exchange history of search phase within globally confined space of 100 m radius. 3 peaks of the gaussian information gradient are aligned equidistantly on the circumference of 50 m radius, which are successfully detected. Total searching time is 60 minutes. Information landscape is stable and $\mu = 0$. The ballistic motion speed v was adjusted to fulfill the low density condition $L >> \frac{v}{f_M}$, preserving the motion/search speed ratio $\nu = 12.5$.

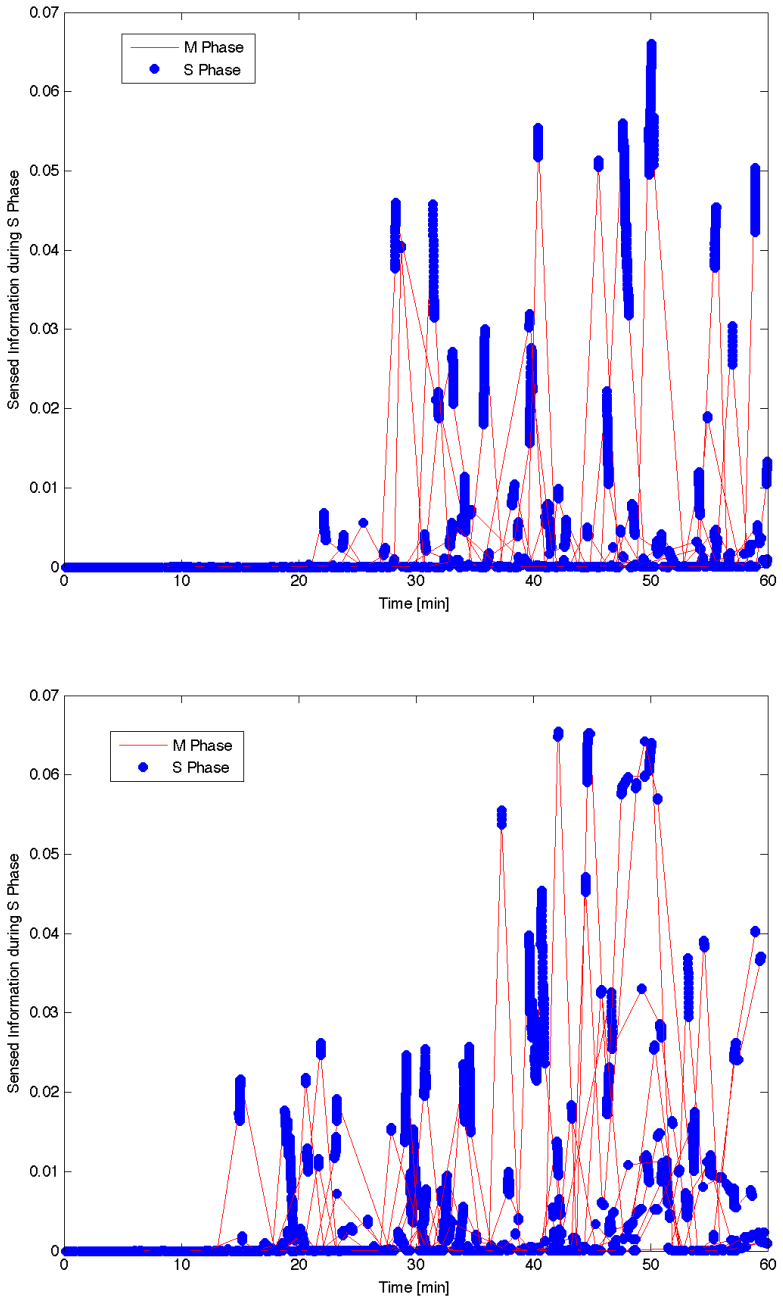


Figure 8.12: **Dynamics of sensed information in single robot and collective search.** Total searching time vs sensed information during search phase is depicted. Top: Superposition of single robot search without communication (corresponds to Fig.8.11 Top). Bottom: 10 robots collective search (corresponds to Fig.8.11 Bottom). Collective search with communication between robots realizes in average quicker search.

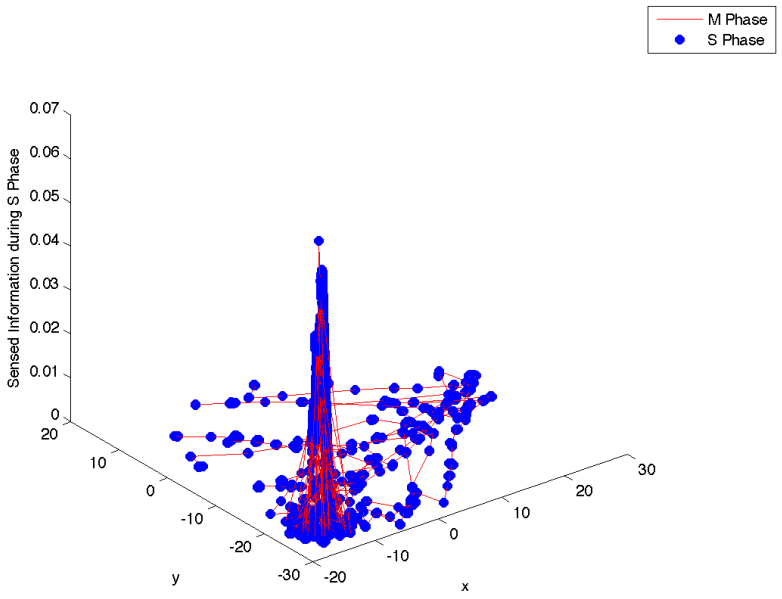


Figure 8.13: **Example of local minimum avoidance in 10 robots collective search.** The robots communicate with 50 m communication range and exchange history of search phase within globally confined space of 50 m radius. 3 peaks of the gaussian information gradient are aligned equidistantly on the circumference of 25 m radius. Only 1 peak is detected due to the winner-take-all principle in dense communication grid. Total searching time is 60 minutes. Information landscape is stable and $\mu = 0$. The ballistic motion speed v was adjusted to fulfill the low density condition $L >> \frac{v}{f_M}$, preserving the motion/search speed ratio $\nu = 12.5$.

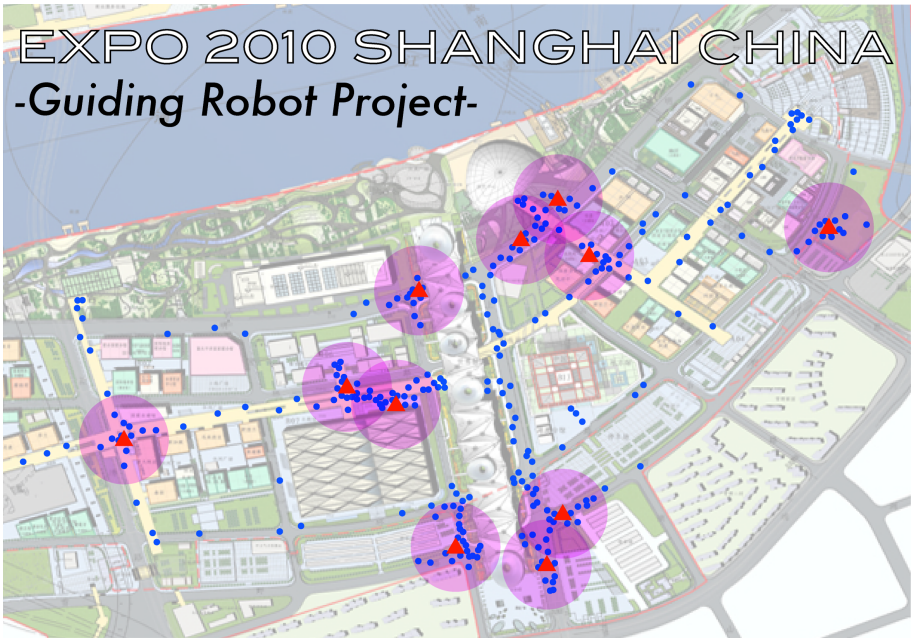


Figure 8.14: **Virtual project 1.** The robots sense the density of human population and automatically adapt to the crowd distribution. The communication between robots facilitate the dynamic deployment without a priori information. The robots display guiding information.

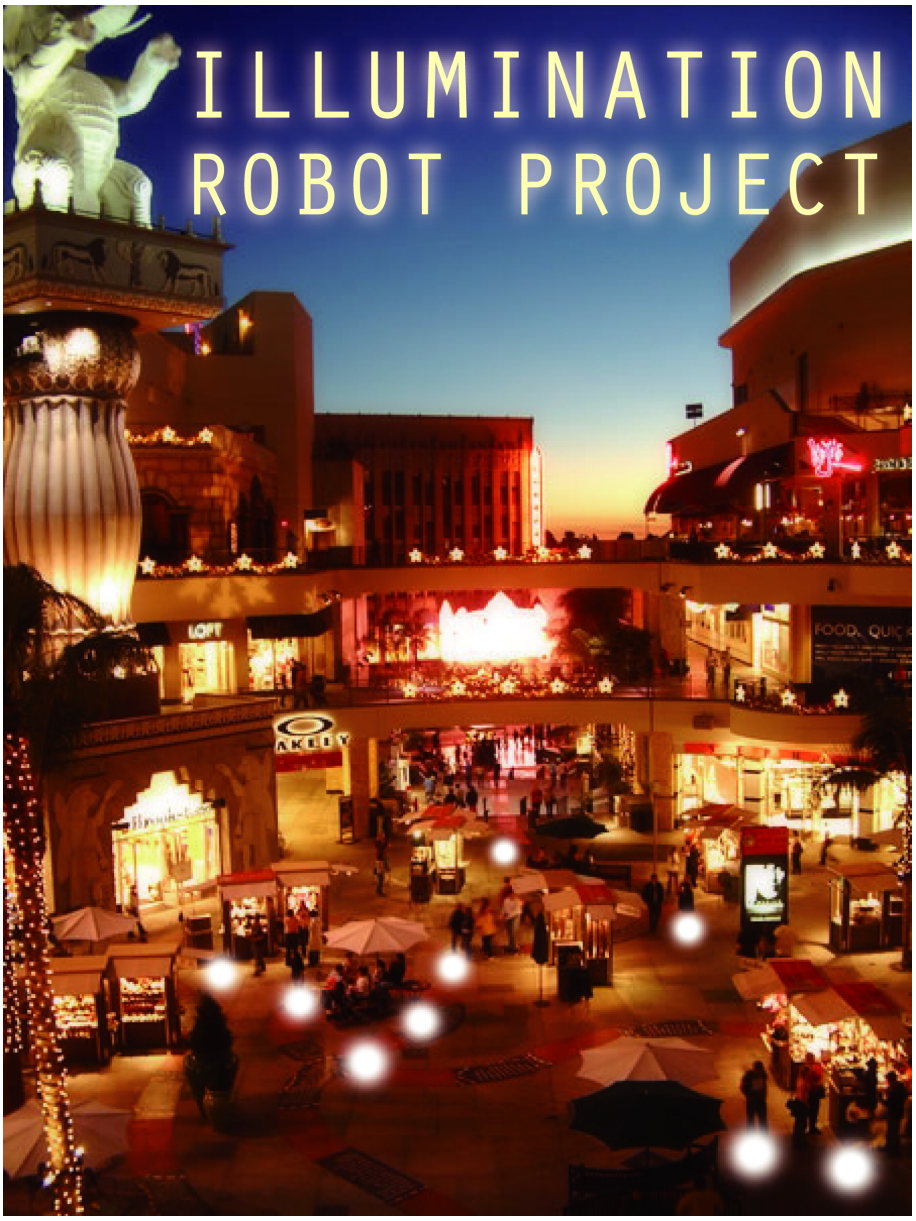


Figure 8.15: **Virtual project 2.** The robots with lighting apparatus adapt to the residence pattern of human in outdoor cafeteria. The robots also function as interface to command food and beverage. Nine illuminations at the ground floor are the robots.

$$
\begin{aligned}
D[p^I : p^R] &= \int d\mathbf{x} \sum_{j=1}^{J} \pi_j^I N(\mu_j^I, \mathbf{\Sigma}_j^I) \log \frac{\pi_j^I N(\mu_j^I, \mathbf{\Sigma}_j^I)}{\pi_j^R N(\mu_j^R, \mathbf{\Sigma}_j^R)} \\
&= \sum_{j=1}^{J} \pi_j^I \left(D[N(\mu_j^I, \mathbf{\Sigma}_j^I) : N(\mu_j^R, \mathbf{\Sigma}_j^R)] + \log \frac{\pi_j^I}{\pi_j^R} \right). \qquad (8.30)
\end{aligned}
$$

Where the component subscript j is chosen to minimize the $D[p^I, p^R]$ according to a given time period. This quantity measures the deviation of the actual information landscape with respect to the information sensed by the collective search. It represents the remaining difficulty to estimate the actual information landscape with the use of the sensed information as an information theoretical measure. The smaller the value of $D[p^I, p^R]$ is than the random search, the more effective the collective search is.

The evaluation of the infotaxis is possible to analyze in multi-scale by both ranging time period and spacial scale by decomposing the $D[p^I, p^R]$ into component-wise terms. From equation (8.30), the following quantity $D_j[p^I, p^R]$ gives the j-th component term of the total discrepancy $D[p^I, p^R]$:

$$
D_j[p^I : p^R] = D[N(\mu_j^I, \mathbf{\Sigma}_j^I) : N(\mu_j^R, \mathbf{\Sigma}_j^R)] + \log \frac{\pi_j^I}{\pi_j^R}. \qquad (8.31)
$$

This measure gives the component-wise comparison between the sensed and actual information landscape, which would represent the spacial heterogeneity of search efficiency induced by an environmental condition. The strength and defect of the collective search are then possible to detect in actual environment. Spacial neighborhood of an important event can also be isolated by choosing appropriate set of $D_j[p^I, p^R]$ representing the necessary components of p^I covering the region.

The component-wise comparison is further possible to decompose into different orders of statistics of the Gaussian component. With the use of the mixture coordinate in information geometry, the following decomposition of $D_j[p^I, p^R]$ between the mean value and the variance component is derived:

$$
\begin{aligned}
D_j[p^I : p^R] &= D[N(\mu_j^I, \mathbf{\Sigma}_j^I) : N(\mu_j^R, \mathbf{\Sigma}_j^I)] \\
&+ D[N(\mu_j^R, \mathbf{\Sigma}_j^I) : N(\mu_j^R, \mathbf{\Sigma}_j^R)] \\
&+ \log \frac{\pi_j^I}{\pi_j^R}. \qquad (8.32)
\end{aligned}
$$

Where the first term represents the deviation with respect to the difference of the mean value, and the second term is that of the variance. The theoretical foundation of this decomposition will be explained in detail in Part 6. This decomposition serves to distinguish the efficiency of the collective search between the detection of the local peak (mean value μ_j^I) and the sensitivity to the spatial extension (variance $\mathbf{\Sigma}_j^I$). If the mean value component of the deviation is large, the robots are missing to detect the local peak of the actual information, while if the variance component is large, the chaotic diffusion of the search and/or motion phase is too large or small with respect to the actual spacial distribution of the information. The higher localization around the information peak gives higher value of the averaged sensed information, but may not be appropriate to enhance human interaction because it contains less diffusive movement. The variance component of the deviation is therefore not a simple measure of search efficiency but rather an index of spacial fluctuation necessary to introduce for human interactions besides the simple infotaxis. Such functional distinction is totally omitted if we simply use the total efficiency defined in (8.30).

8.7 Conclusion

We constructed 2-phase optimal intermittent search strategy of roving robot for sporadic information using the hierarchical statistics of effective dimension in CNN dynamics. We also investigated

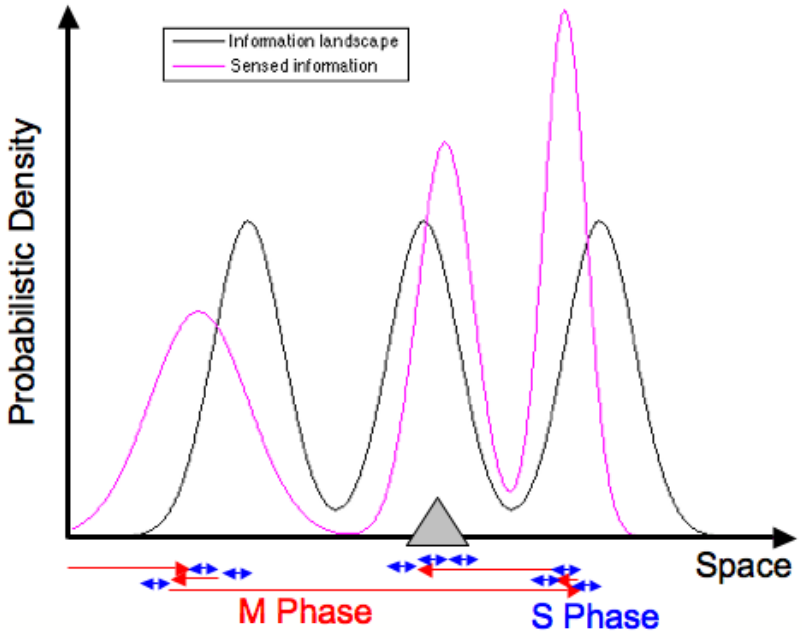


Figure 8.16: **Multi-scale evaluation of collective infotaxis with respect to the actual information landscape.** Both actual information landscape (black line) and sensed information (magenta line) are approximated as Gaussian mixture distribution. The KL divergence between them allows us to quantify the efficiency of the collective search varying both in temporal and spacial resolution, as well as between the mean and variance components.

the collective interaction between the robots and applied to the collective infotaxis in dynamical environment without a priori information. The efficiency of the proposed strategy is shown by a simulation.

Part IV

Complexity Measures and Application to Analysis of Social Network Dynamics

Chapter 9

Order-wise Correlation Dynamics in Text Data

Abstract

We applied information geometry to treat the correlation beyond second order in network co-word analysis. We also defined the meta-level dynamics of these correlations based on all hierarchical combination of the variables, in order to encompass all possible relations. The result shows a rich variety of dynamics depending both on the order of correlation and the order of meta-level dynamics.

Keywords: Hierarchical clustering, Information geometry, co-occurrence network, political weblog, epistemic community

Methodology: Definition of order-wise correlations as contexts and their meta-contexts with information geometry → Trend analysis of the context and meta-context dynamics as dynamical systems → Return map and local stability analysis of the context and meta-context dynamics with dynamical system formulation

9.1 Introduction

Network analysis is one of the prominent tool for addressing a wide range of complex systems objects. Social networks have been extensively studied since the mid twentieth century [73], lexical networks have been introduced in the 80's [74], biological systems are now commonly analyzed through network representation, etc [75, 76]. In this kind of framework, nodes represent elements, and links, eventually weighted, represent the strength of the interactions between those elements. For example, in social networks links are often assumed to represent the number of physical contacts or collaborations ; in lexical networks links could represent the number of co-occurrences of terms in a corpus ; in biological networks, links could represent genes co-expression ; etc. It is noteworthy that in many cases, the value of links are processed over a large number of events where interactions do not take place in a one-to-one way, but rather involve potentially several elements at the same time (co-authors, co-words, etc.). This information conveyed by the patterns of interaction is lost when representing a link has an average over pairwise interactions (9.1). Thus, the standard way to think about links in networks only encapsulates relationships between pairs of nodes and fails to take into account an information that might be critical for further analysis. In particular, several clustering technics aim precisely at recovering communities of nodes that are supposed to interact in the same context (community of co-authors, community of words defining a scientific fields, community of genes coding for a phenotype, etc.).

In this paper, we propose a theoretical framework from information geometry that enables to take into account the fine grained structure of links. By distinguishing different orders of correlations between the variables associated with nodes, we consider links with the multiplicity of contexts in which they appear, and propose a method to identify the most relevant contexts. In the following, a context will be defined as a set of nodes such that their interactions are strongly

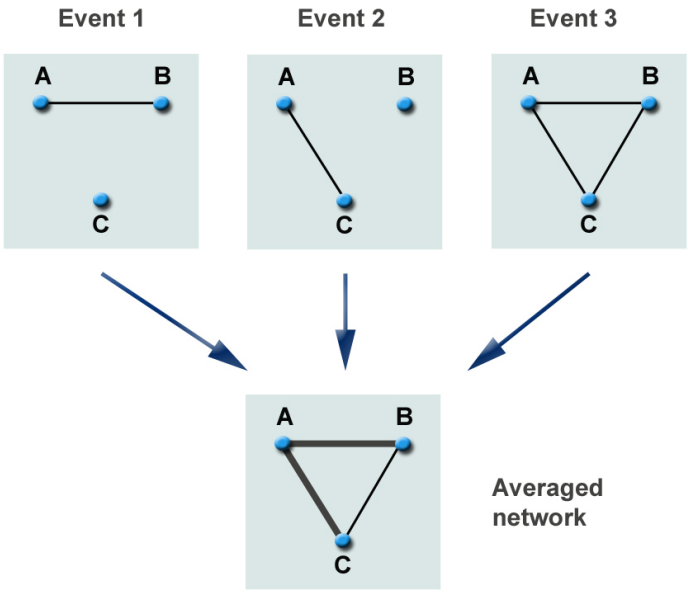


Figure 9.1: **Example of conventional network presentation.** In many cases, the value of links of a network are processed over events where interactions do not take place pairwise, but often involve several elements at the same time (co-authors, co-words, etc.). Thus, the information conveyed by the patterns of interaction is lost when representing a link has an average over pairwise interactions

correlated. Reconstructing all relevant contexts can be viewed as performing an overlapping clustering that takes into account all patterns of interaction at the event level, rather than patterns of interaction on the network structure averaged over all events. In a second section, we propose to extend this approach to higher order correlation computing meta-contexts, which is analogous to performing multi-level clustering.

9.2 Correlations Beyond the Second Order

Although the network representation is intuitively powerful to grasp the structure of complex data , it is not appropriate to distinguish different orders of correlations between the variables associated with nodes. In real data, there exist correlations beyond the second order, which are very important to analyze collective stochastic dynamics.

Let us take for example a network with 3 nodes, each node taking binary values $\{0,1\}$ at a given time t. The dynamics of these variables as a stochastic process can be completely defined by choosing the following 7 parameters of joint distributions at time t:

$$p(1,0,0),\ p(0,1,0),\ p(0,0,1),\ p(1,1,0),\ p(1,0,1),\ p(0,1,1),\ p(1,1,1). \tag{9.1}$$

Note that $p(0,0,0)$ is automatically defined from the normalization $\sum_{i,j,k\in\{0,1\}} p(i,j,k) = 1$.

Next, we consider the following hierarchical marginal distribution η.

$$\begin{aligned}
\eta_1 &= \sum_{i=1,\ j,k\in\{0,1\}} p(i,j,k), \qquad \eta_2 = \textstyle\sum_{j=1,\ i,k\in\{0,1\}} p(i,j,k),\\
\eta_3 &= \sum_{k=1,\ i,j\in\{0,1\}} p(i,j,k), \qquad \eta_{12} = \textstyle\sum_{i,j=1,\ k\in\{0,1\}} p(i,j,k),\\
\eta_{13} &= \sum_{i,k=1,\ j\in\{0,1\}} p(i,j,k), \qquad \eta_{23} = \textstyle\sum_{j,k=1,\ i\in\{0,1\}} p(i,j,k),\\
&\qquad\qquad \eta_{123} = \textstyle\sum_{i,j,k=1} p(i,j,k).
\end{aligned} \tag{9.2}$$

η_1, η_2, and η_3 are the marginal distributions of each variable. η_{12}, η_{13}, and η_{23} represent the degrees of second-order correlations between nodes 1, 2, nodes 1,3, and nodes 2,3, respectively. In the same way, η_{123} defines the degree of third-order correlation. These η can also be obtained by the Legendre transformation of natural parameters of exponential family. This fact later implies a crucial role of these coordinates.

Considering the links between nodes as a representation of interactions between nodes, it comes out straight forward that the third-order correlation included in η_{123} cannot be decided by defining the other η parameters. All the 7 parameters must be defined independently so that to specify a single model. This causes a problem to graph representation when using single kind of link between nodes, because the distinction among different orders of correlations is not considered.

The same argument also holds for the dynamics on graph with n nodes taking binary variables. The k-tuple correlation ($3 \le k \le n$) can not be represented by assembling the correlations less than k-tuple ones. The higher-order correlations are independent from the lower-order ones, so that it should be measured separately for each order. (For example, even if there exists no significant second order correlation, it is possible to observe the correlation among three variables.) This class of model can be generally formulated by log-linear model, with the use of natural parameter θ.

$$\begin{aligned}
\log p(x_1, x_2, \cdots, x_n) = \quad & \textstyle\sum_{i=1}^n \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j + \cdots \\
& + \textstyle\sum_{i<j<\cdots<k} \theta_{ij\cdots k} x_i x_j \cdots x_k + \cdots + \theta_{12\cdots n} x_1 x_2 \cdots x_n - \psi, \qquad (9.3)\\
\psi = \quad & -\log p(0, 0, \cdots, 0). \qquad (9.4)
\end{aligned}$$

For example, θ are defined as follows when $n = 3$.

$$\begin{aligned}
\theta_1 &= \log \frac{p(1,0,0)}{p(0,0,0)}, \qquad & \theta_2 &= \log \tfrac{p(0,1,0)}{p(0,0,0)},\\
\theta_3 &= \log \frac{p(0,0,1)}{p(0,0,0)}, \qquad & \theta_{12} &= \log \tfrac{p(1,1,0)p(0,0,0)}{p(1,0,0)p(0,1,0)},\\
\theta_{13} &= \log \frac{p(1,0,1)p(0,0,0)}{p(1,0,0)p(0,0,1)}, \qquad & \theta_{23} &= \log \tfrac{p(0,1,1)p(0,0,0)}{p(0,1,0)p(0,0,1)},\\
& & \theta_{123} &= \log \tfrac{p(1,1,1)p(1,0,0)p(0,1,0)p(0,0,1)}{p(1,1,0)p(1,0,1)p(0,1,1)p(0,0,0)}.
\end{aligned} \tag{9.5}$$

9.3 Order-wise Decomposition of Correlations

9.3.1 Definition of Context as Order-wise Correlation

Although the correlation between variables can be defined up to the order of total nodes number, little is known on the correlation higher than the second order. Information geometry gives us a clear way to decompose the correlation orthogonally into different order of statistics [77]. Let p_n^n be a data distribution of n binary variables allowing up to the n-th order of correlation, and p_n^1 be the independent joint distribution with the same marginal distribution $(\eta_1, \cdots, \eta_n)$ as the data. The superscript of p represents the possible highest order of interaction, while the subscript of p is the dimension of distribution. Then the Kullback-Leibler divergence $D[p : p^1]$ between the 2

probabilities can be decomposed in the following way:

$$D[p_n^n : p_n^1] = \sum_{k=2}^{n} D_k(p_n^n), \tag{9.6}$$

where $D_k(p_n^n)$ represents the overall degree of interaction purely among the k variables.

The above decomposition is obtained with the use of the dual flat coordinates η and θ. The duality and the flatness of these coordinates can be defined with respect to the Fisher information matrix as metric, and dual $\alpha = \pm 1$ connections. The element $D_n(p_n^n)$, for example, is calculated by the composition of the following mixture coordinate ζ and ζ^0:

$$\begin{aligned}
&(\zeta_1, \cdots, \zeta_n; \zeta_{12}, \cdots, \zeta_{n-1n}; \cdots\cdots\cdots; \zeta_{12\cdots n-1}, \cdots, \zeta_{2\cdots n}; \zeta_{1\cdots n}) \\
&\qquad = (\eta_1, \cdots, \eta_n; \eta_{12}, \cdots, \eta_{n-1n}; \cdots\cdots\cdots; \eta_{12\cdots n-1}, \cdots, \eta_{2\cdots n}; \theta_{1\cdots n}), \qquad (9.7) \\
&(\zeta_1^0, \cdots, \zeta_n^0; \zeta_{12}^0, \cdots, \zeta_{n-1n}^0; \cdots\cdots\cdots; \zeta_{12\cdots n-1}^0, \cdots, \zeta_{2\cdots n}^0; \zeta_{1\cdots n}^0) \\
&\qquad = (\eta_1, \cdots, \eta_n; \eta_{12}, \cdots, \eta_{n-1n}; \cdots\cdots\cdots; \eta_{12\cdots n-1}, \cdots, \eta_{2\cdots n}; 0), \qquad (9.8)
\end{aligned}$$

and

$$D_n(p_n^n) = D[\zeta : \zeta^0]. \tag{9.9}$$

With this example, we reduce the general form (9.6) to the following relation.

$$D[p_n^n : p_n^1] = D_n(p_n^n) + D[\zeta^0 : p_n^1]. \tag{9.10}$$

We now try to isolate the arbitrary k-tuple interaction among k variables. By reducing the model to arbitrary k nodes ($2 \leq k \leq n$), we obtain the same relation using the reduced mixture coordinates ζ' and $\zeta^{0'}$:

$$\begin{aligned}
&(\zeta_1', \cdots, \zeta_k'; \zeta_{12}', \cdots, \zeta_{k-1k}'; \cdots\cdots\cdots; \zeta_{12\cdots k-1}', \cdots, \zeta_{2\cdots k}'; \zeta_{1\cdots k}') \\
&\qquad = (\eta_1, \cdots, \eta_k; \eta_{12}, \cdots, \eta_{k-1k}; \cdots\cdots\cdots; \eta_{12\cdots k-1}, \cdots, \eta_{2\cdots k}; \theta_{1\cdots k}), \qquad (9.11) \\
&(\zeta_1^{0'}, \cdots, \zeta_k^{0'}; \zeta_{12}^{0'}, \cdots, \zeta_{k-1k}^{0'}; \cdots\cdots\cdots; \zeta_{12\cdots k-1}^{0'}, \cdots, \zeta_{2\cdots k}^{0'}; \zeta_{1\cdots k}^{0'}) \\
&\qquad = (\eta_1, \cdots, \eta_k; \eta_{12}, \cdots, \eta_{k-1k}; \cdots\cdots\cdots; \eta_{12\cdots k-1}, \cdots, \eta_{2\cdots k}; 0), \qquad (9.12) \\
&\qquad D_k(p_k^k) = D[\zeta' : \zeta^{0'}], \qquad (9.13) \\
&\qquad D[p_k^k : p_k^1] = D_k(p_k^k) + D[\zeta^{0'} : p_k^1]. \qquad (9.14)
\end{aligned}$$

Where $D_k(p_k^k)$ is the interaction purely among k variables, while $D[\zeta^{0'} : p_k^1]$ represents the rest of the interactions lower than the k-th order. Note that ζ' is not a submanifold of ζ. In ζ', we ignore completely the data outside of k nodes subgraph.

This decomposition is also compatible to chi-squared test to judge the significance of order-wise correlation. Considering only the highest order of interaction in k nodes subgraph, the following holds asymptotically with enough large sample number N.

$$\lambda = 2ND_k(p_k^k) \sim \chi^2(1). \tag{9.15}$$

Where $\chi^2(1)$ is the chi-squared distribution with one degree of freedom.

We also consider the distinction between positive and negative correlation by changing the positive/negative sign of the statistics λ:

$$\lambda = \mathrm{sgn}(\eta_{12\cdots k} - \eta'_{12\cdots k}) \cdot 2ND_k(p_k^k). \tag{9.16}$$

Where $\mathrm{sgn}(\cdot)$ is the signum function, and $\eta'_{12\cdots k}$ is the solution of the $\theta_{12\cdots k} = 0$ condition with respect to the $\eta_{12\cdots k}$ variable. This means that the 5 % significance level of positive/negative correlation is judged by the λ being above/below the +/- 5 % threshold (3.841), respectively.

By changing the k, we are able to evaluate all possible sets of pure k-tuple interaction among specific set of k variables. We call this λ the significance of k-th order context among k variables.

Pragmatically, the sample used to calculate λ is not necessary limited to each time step of data. In weblog data, for example, it is common to keep the history of about the past 1 week to 1 month in the same page. Therefore, the distribution at time t should also reflect these periods in the calculation by setting proportional length of the time window. This prescription also ameliorate the sensitivity of the test.

For the following analysis based on text data, we define the context as the significance of keywords co-occurence. The context of k words corresponds to the k-th order correlation of the occurrence of k words, which is evaluated by (9.16). Therefore, there exist $n_c = \sum_{k=1}^{n} {}_nC_k$ possible contexts in the graph of n nodes. We also define the subcontext of a context when the k-th order context of binary variables $(x_1, \cdots, x_k)$ and the k_{sub}-th order context of binary variables $(x_1, \cdots, x_{k_{sub}})$ exist, where $2 \leq k_{sub} < k$ and $\{x_1, \cdots, x_{k_{sub}}\} \subset \{x_1, \cdots, x_k\}$ are satisfied. The k_{sub}-th order context is then a subcontext of the k-th order context. This gives us a way of clustering taking the different orders of interaction into consideration. We applied this theory to distinguish different order of correlations on the log-linear model with $n = 4$ binary variables.

9.3.2 Definition of Meta-Context

Furthermore, contexts between contexts can also be defined in the same manner. This is a kind of hierarchical clustering taking the result of the previous level. From the calculation of the n_c contexts' significance, we first judge for each order k and each period t which contexts are significant or not. Next, we obtain the occurrence matrix of the contexts, that specifies with a binary variable if a context is significant (true, = 1) or not (false, = 0). This new binary data has the same format as the raw data used to calculate the contexts, except the number of variables being n_c instead of n.

It is now a question if these significant contexts also have correlations among each other. This information cannot be treated with context only, but requires to establish relations across contexts. It is observed in the relation between contexts. It can also be considered as the relation among n_c links of the n nodes graph. We call this co-occurrence of contexts as "meta-context." The relation between the binary variables, the contexts, and the meta-contexts is depicted in FIg. 9.2. In the same way as the number of contexts, $n_c^{meta} = \sum_{k=1}^{n_c} {}_{n_c}C_k$ meta-contexts are possible in n nodes graph. As an expansion, performing the same operation on meta-context occurrence matrix, there exists $n_c^{meta^2} = \sum_{k=1}^{n_c^{meta}} {}_{n_c^{meta}}C_k$ meta-meta-contexts. For simplicity, we denote such expansion as meta^2-context, which is possible to generalize to the arbitrary finite levels as meta^l-context $(1 \leq l < \infty)$.

It is clear that the number of meta^l-context grows more than exponentially with l when $n \geq 4$. Though, it is natural to assume that there exists a certain limit of l, namely l_{max}, where there is no significant meta$^{l_{max}}$-context. Ascending the hierarchy of meta^l-context up to this limit, we are able to detect all relativistic relations unconsciously associated with the text data.

9.4 Data Set

The campaign for the French presidential election that ended in May 2007 was marked by the strengthening of new modes of public expression. As the access to new communication media was democratizing, new patterns of public intervention have been invented. The main outcome of this new mode of intervention was that citizens could express their own views about social, political or economical topics concerning the campaign that in turn modified deeply the political agenda. The political blogosphere, in particular, is one of the main places of appropriation of public debate.

We have collected the content and hypertext links of a selection of 120 political blogs during the first six months of 2007. The database allows each blog to provide a time series of semantic attributes (quantified using a vector of occurrences of 190 key terms) that evolve according to new posts. This provides us a daily semantic characterization of each blog from January the 1^{st} to June the 30^{th} reflecting the current concerns of each blogger. The correspondence between the date of important issues and the daily time axis is presented in Tab. 10.1.

We have analyzed these 120 weblog text data focusing on the 4 candidates (Ségolène Royal, François Bayrou, Nicolas Sarkozy, and Jean-Marie Le Pen) of the last French presidential. For

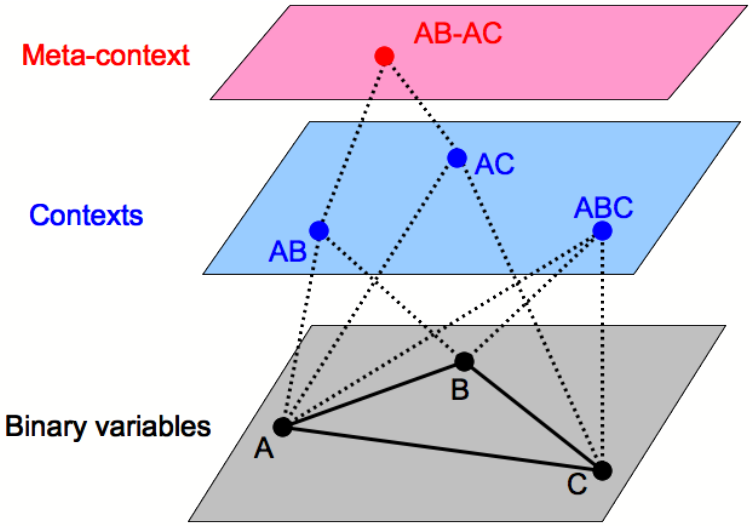


Figure 9.2: **Hierarchical relation between binary variables, contexts, and meta-contexts.** As an example, three nodes network with binary variables A, B, and C is shown. The horizontal planes and dashed lines schematically defines the combinations to define the above level of interactions. Assume that the second order contexts AB, AC, and third order context ABC are significant. Then the meta-contexts are defined on the co-occurrence of the contexts, in which, for example, only AB-AC is significant.

simplicity, we will take the abbreviation of their family name, R,B,S,L to denote each candidate. We will use these abbreviations as subscript for the binary occurrence vector (x_R, x_B, x_S, x_L) aligned from left to right according to their political position being leftist to rightist.We set the first or family name of each candidate as the key terms to scan the weblogs and obtain the data. Each of 120 blogs was crawled each day to see whether these names appeared or not in their posts, and coded accordingly as 1 (occurrence) and 0 (non-occurrence). For example, if a blog post contained the words "Royal" and "Sarkozy" but none of the others, then the occurrence data for this blog in this day will be $(x_R, x_B, x_S, x_L) = (1, 0, 1, 0)$. Using this occurrence vectors, we extracted significant order-wise correlation as context.

9.5 Result and Discussion

9.5.1 Context Dynamics

For the clarity, among the dynamics of all $n_c = 11$ contexts, we only focus in this paper on some interesting sets of contexts. Here, we are limited to give a case study on the above mentioned data, and further study is needed for a generalization.

One of the novelties of this method is that we are able to distinguish the significance of a context from its subcontexts. In other word, we can compare the different orders of correlation on a specific set of variables. This raise an issue about whether higher-order correlation could play an important role to characterize social dynamics. To focus on this issue, we take here 2 opposite examples, the one where higher-order correlation is stronger (Fig.9.4), and the other where lower-order correlation or the subcontexts are more dominant (Fig.9.3). Using the abbreviation to denote the contexts, Fig.9.3 shows the dynamics of the 2nd order contexts RB, RS, BS and the 3rd order context RBS. While Fig.9.4 shows the dynamics of the 2nd order contexts RS, RL, SL and the 3rd order context RSL. Since little negative correlation is observed, only positive correlation is depicted. We took three different time windows of daily, weekly, and biweekly scales, and the distribution for each day was calculated taking the data of 1, 7, and 14 previous days, respectively.

These contexts remains mostly significant during the campaign except the context RBS. The context RSL is shown to be always superior to the significance of its subcontexts. This means there

Table 9.1: **List of important issues during the election.** For simplicity of the later analysis, we introduced a daily time axis to give a decimal notation. The source in French: http://www.presidentielle-2007.net/calendrier.php

issue	Date	Decimal date
Deadline for inscription to the list of candidates	31 Dec 2006	0
Decree of summons to candidates	22 Feb 2007	53
Dispatch of bulletins of sponsorship to candidates	22 Feb 2007	53
Deadline for registration of 500 sponsorships	16 Mar 18:00	75
Publication of candidate list	19 Mar 2007	78
Start of official campaign	9 Avr 2007	99
Publication of sponsorships	10 Avr 2007	100
End of official campaign	20 Avr 2007	110
First round of presidential	22 Avr 2007	112
Second round of presidential	6 Mai 2007	127
Start of new presidential term of office	17 Mai 2007	137
General elections	10 & 17 Jun 2007	161 & 168

was a general tendency to put the three names R,S, and L altogether instead of only two of them in blog posts. Since R and S were the two persons who actually went to the final referendum, the superiority of the context RSL suggests that L was a popular catalyst of the debate about R and S, which was actually the case in many situations.

The context RBS, on the other hand, only shows intermittent significance, and mostly behind its subcontexts. This shows the tendency of people who would rather compare each 2 persons among R,B, and S instead of discussing on these 3 candidates together. Since B is considered to be a moderate candidate, the inferiority of the context RBS may represent the lack of the arguments with comprehensive perspective ranging from left to right wing.

The empirical results of the context dynamics are shown in Figs.9.5 9.6 9.7 and 9.8. In all positive correlations, generally the peak comes at the second round (127th day) where the new president S is finally chosen, except the 4th order context RBSL. The relatively low 4th order context may be a characteristic of intense discussion. The steepness of the peak around 127th day is relatively large in the 2nd order contexts, less in the 3rd order, and decreases in the 4th order. This implies a tendency that as people get more involved in the discussion, the significance of the contexts increases but in some sense inversely proportional to the orders of correlation. In other word, the discussion becomes intense to compare every set of the candidates in detail, but less effort is payed on more comprehensive perspectives.

There exists also daily fluctuation, which diminish as the temporal scale increases. Since the intervals of the important issues are in most cases less than two weeks, the weekly scale would be appropriate for the characterization, both to avoid daily fluctuation and keep the temporal resolution. Indeed, most weblogs contain the articles of past several days, which supports the plausibility of calculation in weekly time scale. Negative correlations tend to appear in daily fluctuations, and are stronger in 4th and 3rd orders. Significant negative correlations in weekly scale appear after the second round (127th day). Negative 2nd order correlations never become significant with respect to the 5 % threshold.

9.5.2 Meta-Context Dynamics

We analyze here the contexts between the contexts RB, RS, BS, and RBS, or the meta-contexts of R,B, and S. Figs.9.9 and 9.10 show separately the positive and negative correlation of these meta-contexts. The distribution at each day was calculated taking the data of 30 previous days. Therefore, the significance is judged on a monthly scale basis.

From this viewpoint, the whole election campaign can be characterized in 3 distinctive parts. The first 78 days before the publication of the candidate list are dominated by many negative correlations around 5 % significant level. This means the significant lack of meta-context dynamics much more than an independent trial. There exists however one constant positive correlation of

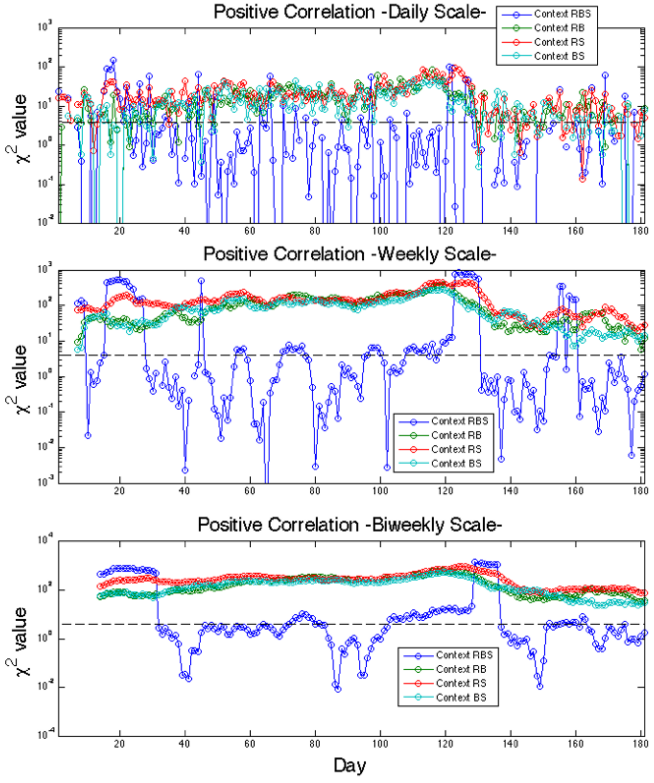


Figure 9.3: **Context dynamics among R, B, and S.**The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only positive correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

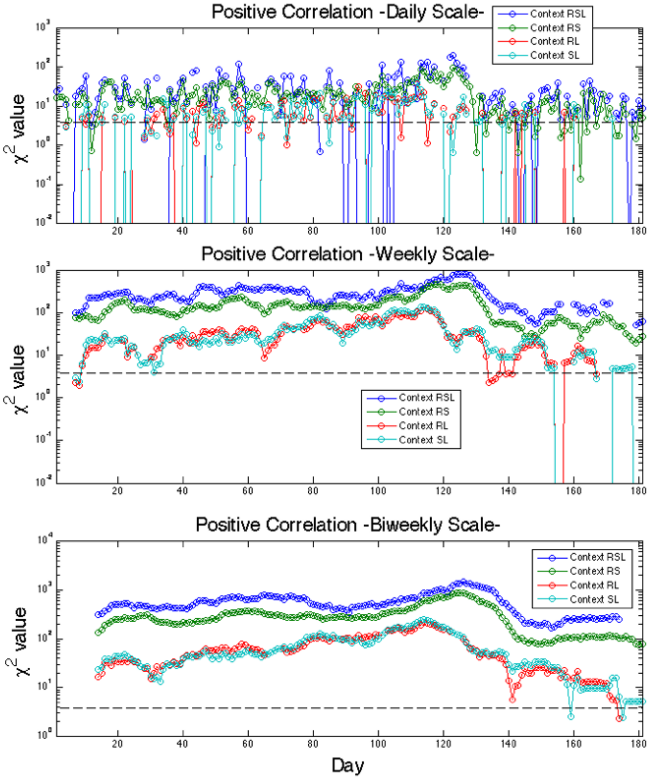


Figure 9.4: **Context dynamics among R, S, and L.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only positive correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

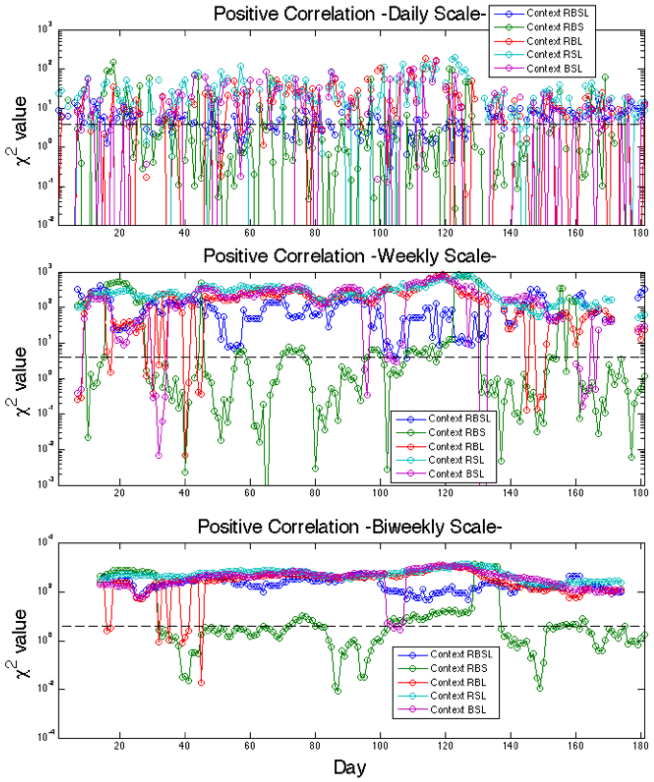


Figure 9.5: **High order positive context dynamics.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only positive correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

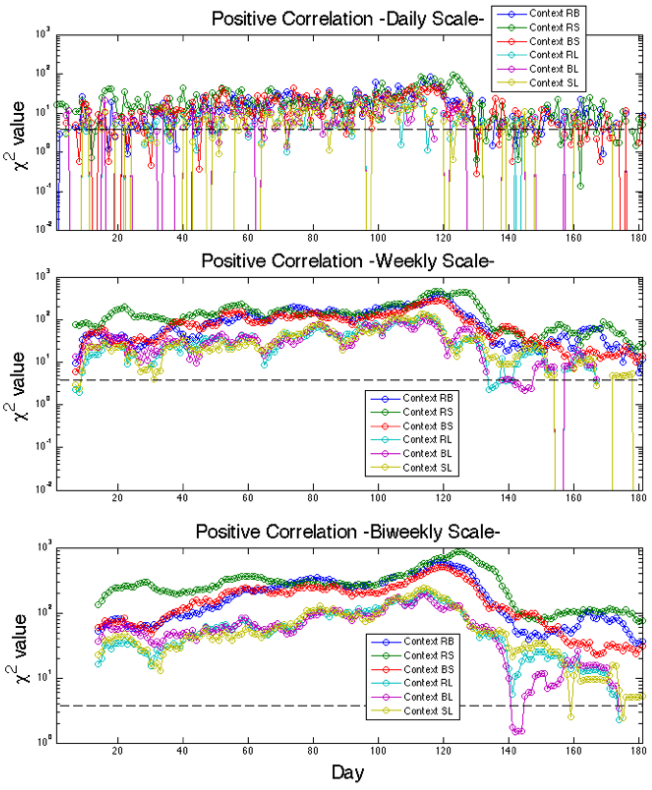


Figure 9.6: **Pair-wise positive context dynamics.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only positive correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

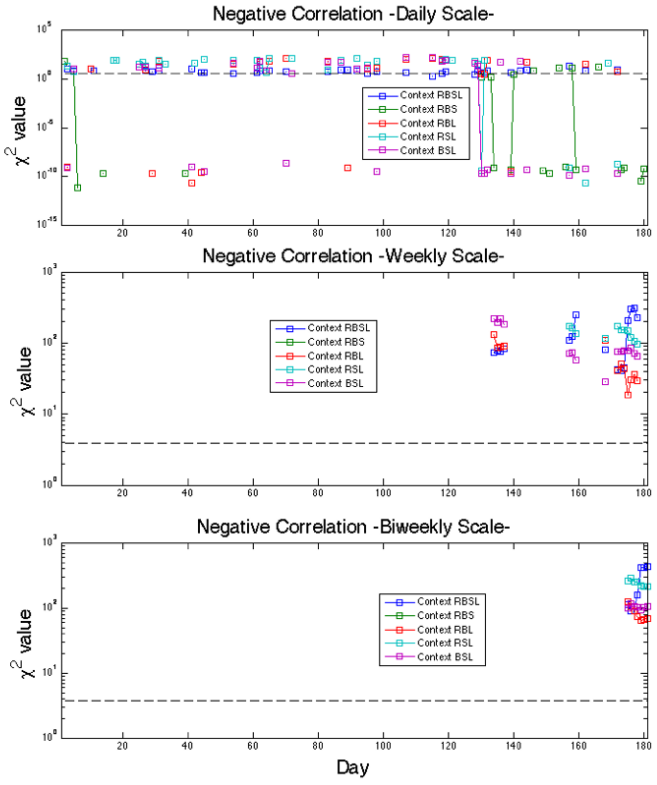


Figure 9.7: **High order negative context dynamics.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only negative correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

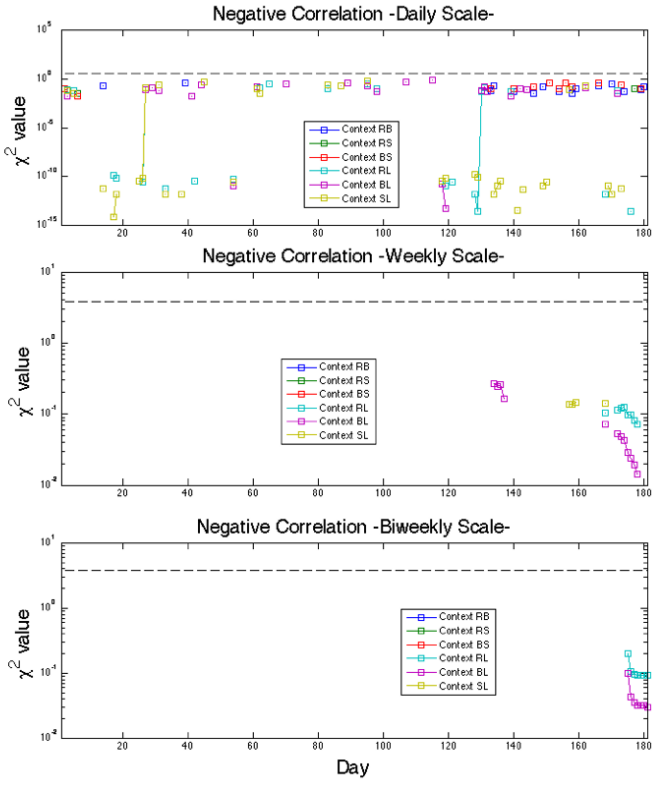


Figure 9.8: **Pair-wise negative context dynamics.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only negative correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

RB-BS. This was then the only popular topic at meta-context level in this period.

Possible interpretation of the meta-context RB-BS in the first third may be an anticipation of the second round (127th day). At the first round (112th day), people are supposed to choose two candidates who will fight the second round. Therefore, there should be the discussion to compare between different pairs of candidates that better represent the voter's will, or that better compete with the worst candidate for them. The second order meta-contexts between second order contexts may reflect such arguments.

The mid term from 78th to 127th day including the official campaign shows the lack of significant correlation neither positive nor negative, except the 4th-order meta-context RB-RS-BS-RBS nearly at the 5 % threshold. Although the contexts between R,B, and S are relatively high in this period, their meta-dynamics are less significant. Though, the significance of the meta-context dynamics is in some sense proportional to their order, which is contrary to the result of contexts dynamics.

The last term is dominated by the 2nd-order meta-contexts around positive significance level, and the intermittent significance of some other higher-order meta-contexts. As the new president was chosen to be S at the 127th day, 2nd-order meta-contexts highly emerged.

The results of the 2nd-order meta-contexts of R, S, and L are also shown in Figs. 9.11 and 9.12. There exist also qualitative division of dynamics in three parts. The first period is about before 63rd day, which we can not find specific event from the official important issues. The second period is between 63rd to 127th day, whose end again coincides with the second round. The third period is the same as that of R,B, and S.

The first period of the meta-contexts between R,S, and L are dominated with many 2nd-order positive correlations. Some of them drop down during the second period, and regain the significance at the third period. Higher-order meta-contexts tend to be higher in the second period, as observed also in R,B, and S. On the other hand, most of the negative correlations only appear at the 3rd period.The fourth-order meta-context RS-SL-SL-RSL remains constantly high below the threshold during the whole sequence, regardless of the three periods.

The results of meta-contexts show completely different dynamics from that of contexts. The remarkable thing is that there exist 3 terms that are sharply separated along the time axis, and show the sudden drastic changes that corresponds to the important issues of the election..Such property is not apparent in context dynamics. This implies the importance and independence of dynamics in meta-context level, and leads us to form a hypothesis if the meta-context dynamics can serve for the global characterization in temporal scale. If the meta-context dynamics has generally such tendency to change discretely with typical transition patterns, and to be maintained for a certain period, it may allow us in real time to forecast in which phase of opinion formation the community is situated.

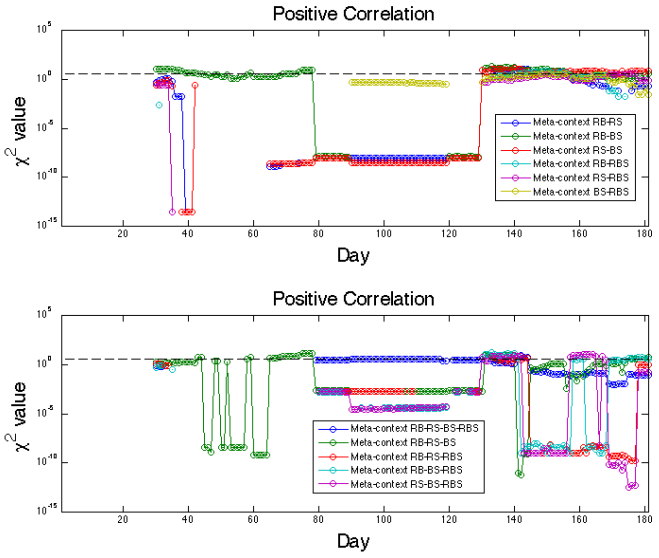


Figure 9.9: **Positive meta-context dynamics among the contexts RB, RS, BS, and RBS.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only positive correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

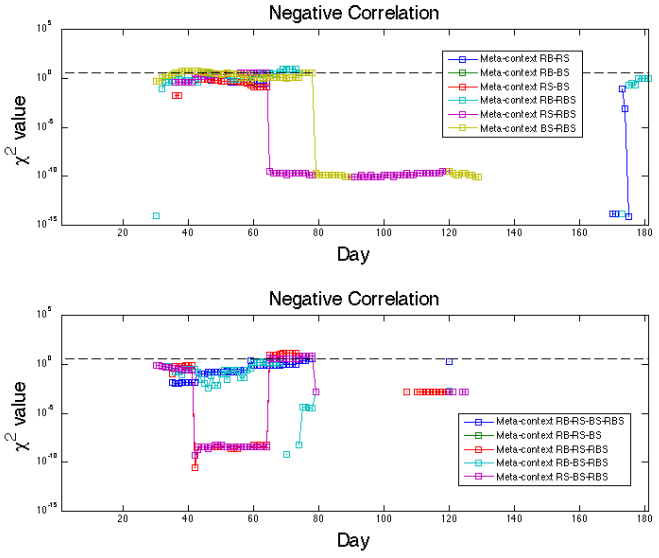


Figure 9.10: **Negative meta-context dynamics among the contexts RB, RS, BS, and RBS.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only negative correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

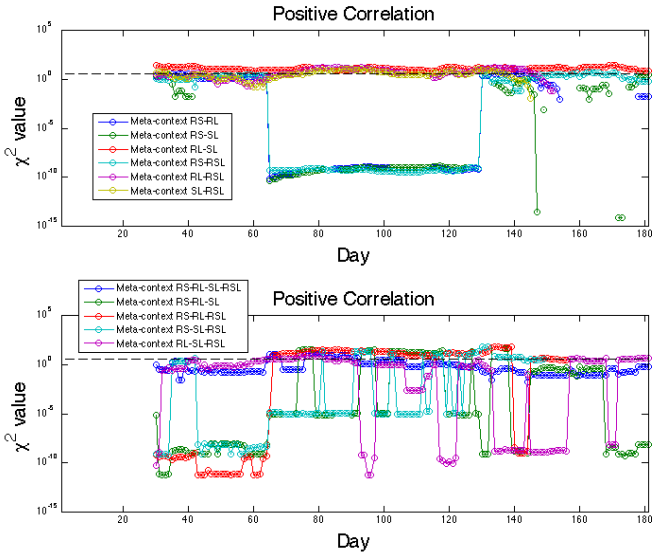


Figure 9.11: **Positive meta-context dynamics among the contexts RS, RL, SL, and RSL.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only positive correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

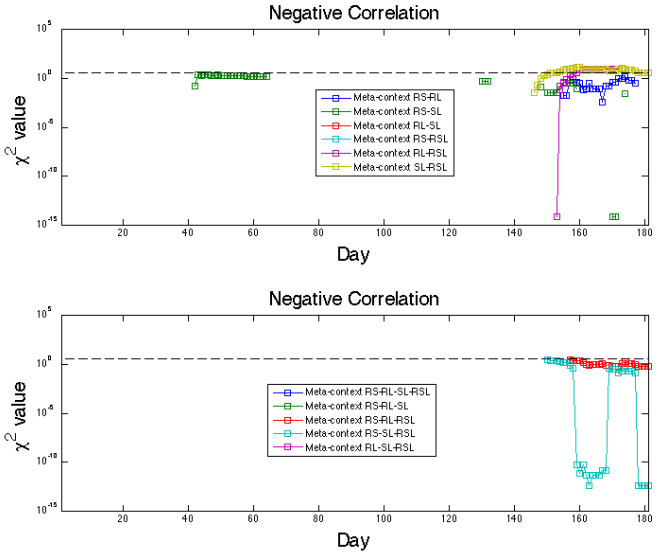


Figure 9.12: **Negative meta-context dynamics among the contexts RS, RL, SL, and RSL.** The absolute value of λ defined in (9.16) is plotted in log-scale vertical axis. Only negative correlation is shown. The horizontal axis is the day counted from the beginning of the campaign. The dashed line is the 5 % significant level 3.841 of χ^2 test.

9.5.3 Return Map Analysis of Context Dynamics

The existence of approximately 1-dimensional trajectory insists that there exists a function $f : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ that determines the value of $\lambda(t+T)$ from $\lambda(t)$ loosely depending on t:

$$\lambda(t+T) = f(\lambda(t), t). \tag{9.17}$$

The dependence on t should satisfy certain constraint such as smoothness to exclude completely random dynamics. This function can be modeled as an almost everywhere continuous and differentiable function with respect to $\lambda(t)$ and t. This function is conceptually a lower-dimensional projection of the whole social dynamics as a high-dimensional nonlinear dynamical system. The possible class of this function and its nature is an interesting open question to characterize the collective dynamics on social networks.

The plausibility of this hypothesis can be measured by comparing with random dynamics. If there exists no dynamical relation, the trajectory on the return map is expressed as a random walk and does not show global structure (Fig.9.13).

Assuming the presence of such function locally, taking analogy of linear stability analysis, we can consider the development of perturbation δ in the neighborhood of data orbit.

We first numerically estimate the function $f(\cdot,\cdot)$ by smoothing 3 adjacent points $\lambda(t-1), \lambda(t), \lambda(t+1)$ for all t. Next, we obtain the moment Lyapunov exponent of $f(\lambda(t), t)$ at time t by calculating the value of $\log\left|\frac{\partial f(\lambda(t),t)}{\partial \lambda(t)}\right|$. Since $f(\lambda(t)+\delta, t) = \lambda(t+T) + \frac{\partial f(\lambda(t),t)}{\partial t}\delta + O(\delta^2)$ holds, the moment Lyapunov exponent represent the linear stability(-) and unstability(+) of the estimated dynamics in 1 step of T interval. Following the development of δ for n step, we obtain

$$\lambda(t+nT) = \lambda(t+T) + \Pi_{k=0}^{n-1} \frac{\partial f(\lambda(t+kT), t+kT)}{\partial \lambda(t+kT)}\delta + O(\delta^2). \tag{9.18}$$

Therefore, the Lyapunov exponent LE representing global stability/instability of the dynamics in n steps can be defined taking the average of both the time evolution and the overlapping T distributions as follows:

$$LE = \frac{1}{nT}\sum_t \log\left|\frac{\partial f(\lambda(t), t)}{\partial \lambda(t)}\right|. \tag{9.19}$$

Hence, in this definition, the dependence of the function f to the time t is also taken into account, while the classical method of Lyapunov exponent estimation from time series usually takes as a theoretical assumption the temporal invariance of f inside of an bounded attractor [78].

The Lyapunov exponent can also be interpreted as the loss of information associated with the evolution of dynamics. Therefore, this gives a measure of complexity in terms of predictability of social dynamics.

Note that the above discussion holds only when there exists clear trace of 1 dimensional curve in the return map figure. Most of the context dynamics in biweekly scale can satisfy such condition (Fig.(9.14)). Irregular scattering on the return map reduces the plausibility of the estimation of $f(\cdot,\cdot)$. We investigated the dynamics of the moment Lyapunov exponent and calculated the value of the Lyapunov exponent for each context dynamics in biweekly scale (Fig.(9.15)). Although the value of λ varies, the moment Lyapunov exponents mostly keep fluctuation between constant boundaries. The dynamics keep similar dynamics in instantaneous complexity, but sometimes go through steep changes observed as sharp positive peaks. These can be interpreted as the points where fluctuation of individual opinion might lead to the macroscopic change of social trend. The Lyapunov exponents are positive in all contexts, implying the difficulty of long-term prediction. It can be considered as the impossibility of prediction and information compression in social dynamics. This fact also supports one of the definitions of complex systems that there is no better simulator than the real dynamics itself.

The relation between the moment Lyapunov exponent and the significance λ was also investigated in Fig.(9.16)). Although we need to further accumulate similar data to be convinced, the greater moment Lyapunov exponents, especially the drastic positive peaks tend to be observed

rather close to the 5 % threshold 3.848 of λ, compared to the whole range of the dynamics. The high values of λ, on the other hand, tend to be concentrated near the line where the moment Lyapunov exponent is 0. This may imply that once the significance of a context is highly settled, there are cases where the dynamics autonomously diminish unstability and form a relatively stable, short term popular topic. Since the estimated long term Lyapunov exponents are positive, the stability of the contexts should be evaluated in term of the weak unstability. The high significance of a context can be considered as a stagnation of the dynamics, which is still weakly unstable, but not as turbulent as the rest.

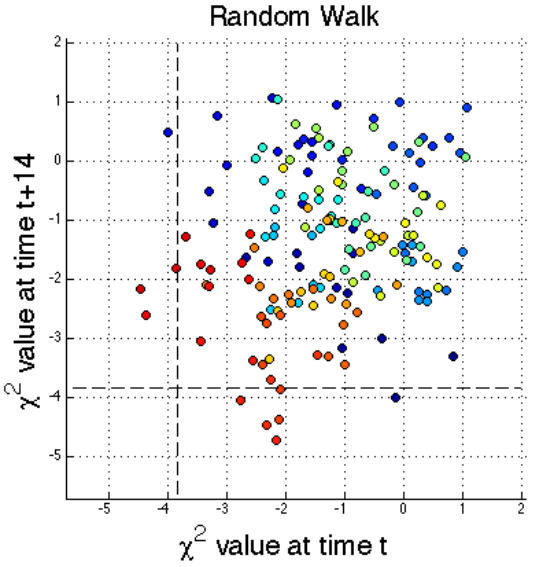


Figure 9.13: **Return map of one-dimensional random walk.** Each step change of the dynamics follows the uniform distribution between $[-1, 1]$. The color indicate the value of the time t.

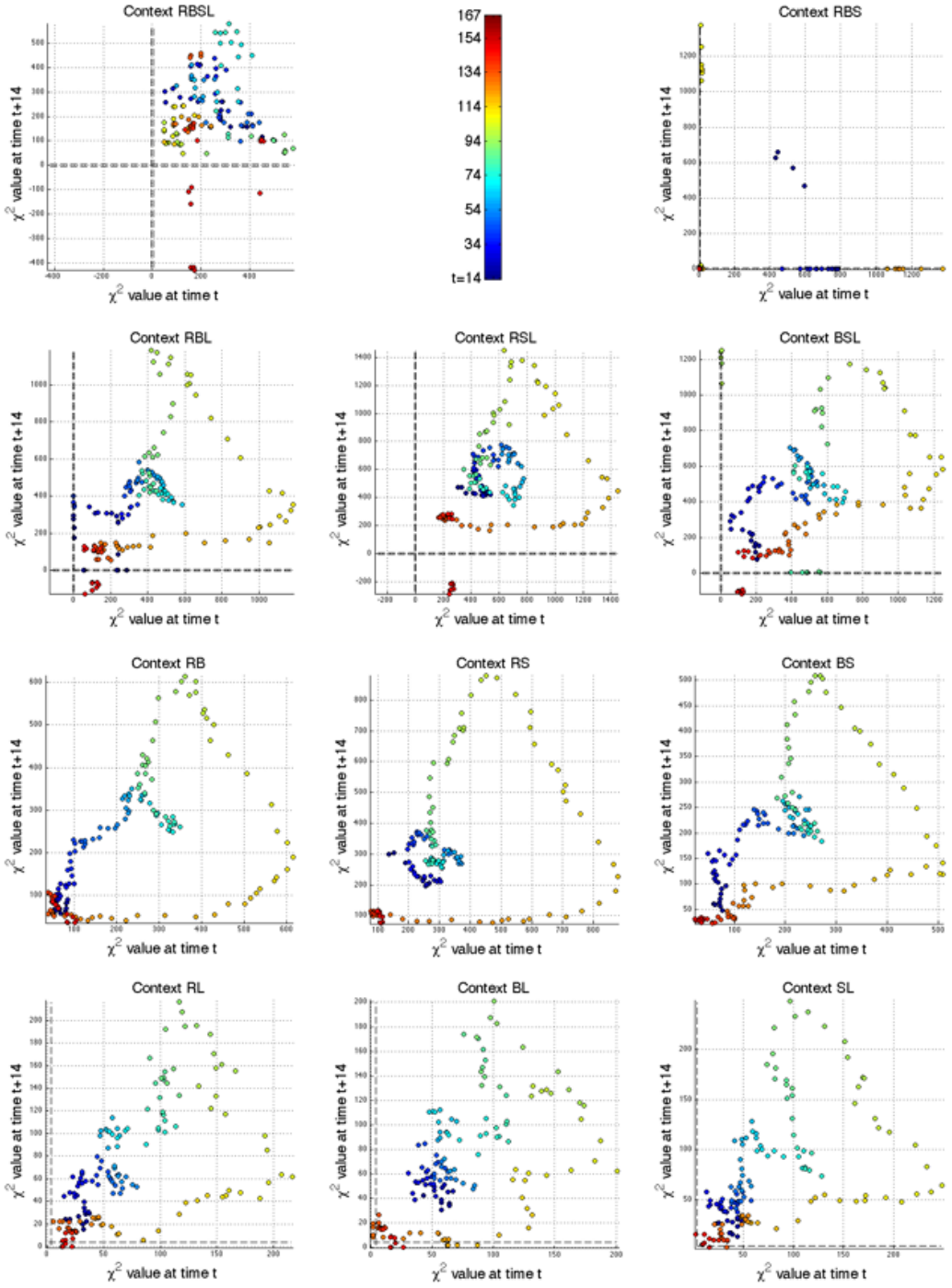


Figure 9.14: **Return maps of the context dynamics among R, B, S and L.** The value of λ defined in (9.16) at time t vs. $t + 14$ is plotted. The color indicate the value of the time t.

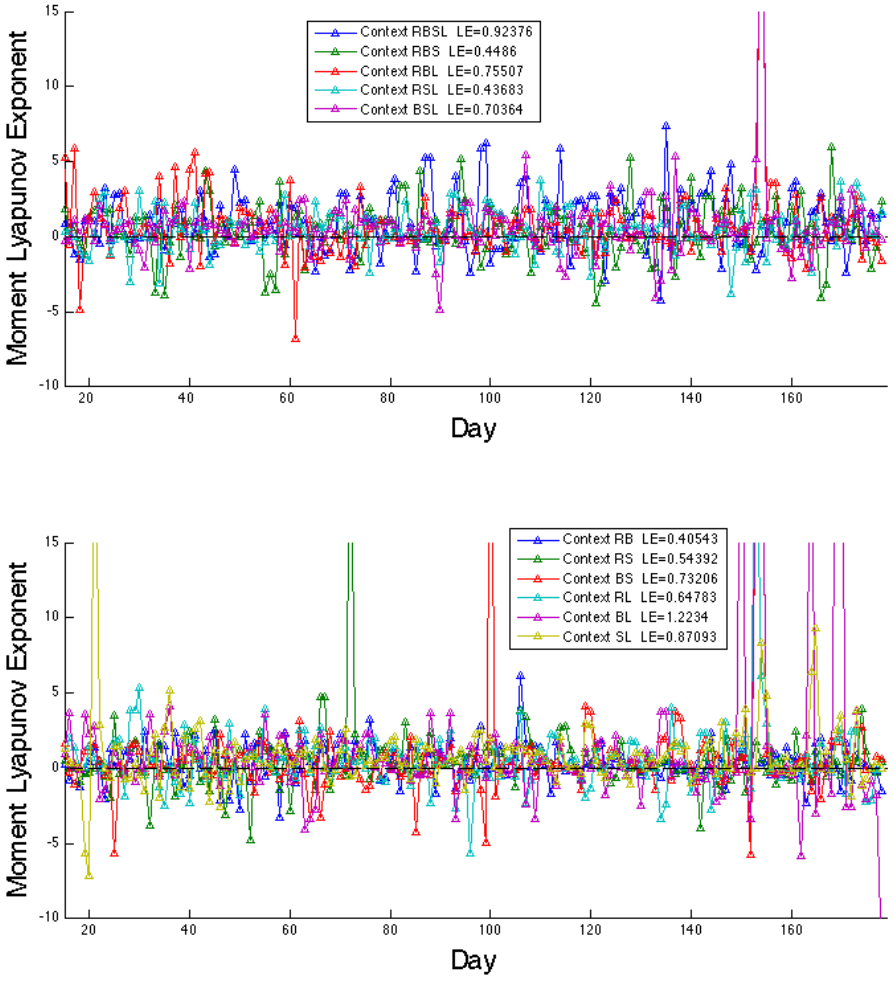


Figure 9.15: **Evolution of the moment Lyapunov exponent in biweekly scale.** The value of the Lyapunov exponents are shown in the legend.

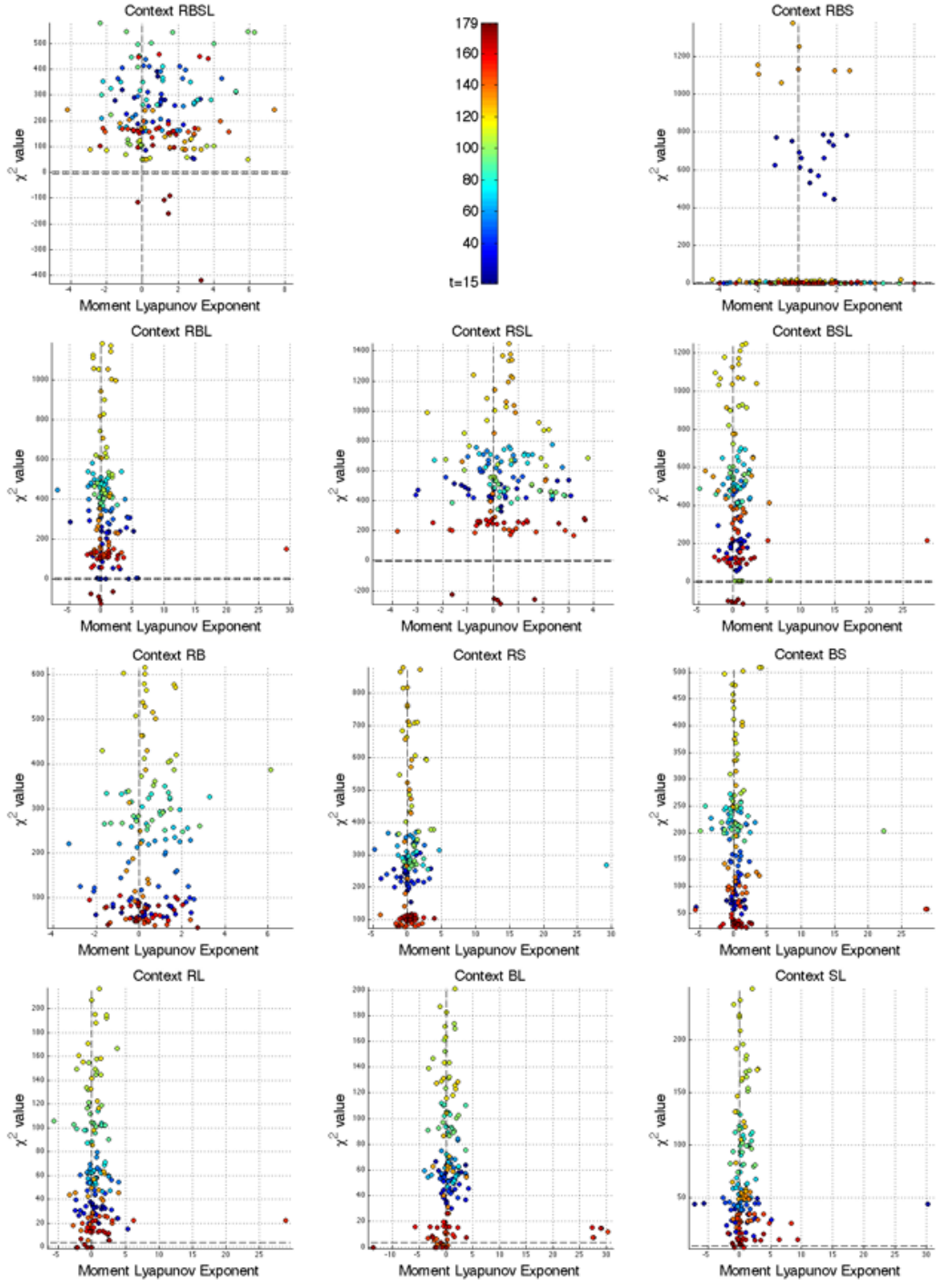


Figure 9.16: **Relation between instantaneous linear stability and significance in biweekly scale.** Horizontal axis: Moment Lyapunov exponent, Vertical axis: The value of λ defined in (9.16) . The color indicate the time t. The horizontal dashed lines are the 5 % significant level ± 3.841 of χ^2 test. The vertical dashed line is the changing point 0 between stability and unstability of the moment Lyapunov exponent.

9.6 Conclusion

We have investigated all possible orders of interaction on the combinational hierarchy of 4-dimensional binary variables with the use of political weblog data, and discussed their significance. The method provided a novel way of hierarchical clustering according to the orders of interactions and its meta-level combinations, reflecting data structure into the definition of co-occurrence matrix. The defined setting provided a wide range of viewpoints, which is also applicable in general to the characterization of epistemic community from text data.

Chapter 10

Network Decomposition: An Information Theoretical Approach

Abstract

We consider the graph representation of the stochastic model with n binary variables, and develop an information theoretical framework to measure the degree of interactions existing between subsystems as well as the ones represented by each edge of the graph representation. A case study on political weblog data is demonstrated.

Keywords: Information geometry, Complex network, Hierarchical clustering

Methodology: Formulation of system decomposition and edge cutting with information geometry → Trend analysis of political weblog data as dynamical systems

10.1 Introduction

Complex systems sciences emphasize on the importance of non-linear interactions which can not be easily approximated linearly. In other word, the degree of non-linear interactions are the source of complexity. The classical reductionism approach generally decomposes a system into its components with linear interactions, and tries to evaluate whether the whole property of the system can still be reproduced. If this decomposition of a system destroys too much information to reproduce the system's whole property, the plausibility of such reductionism is lost. Inversely, if we can evaluate how much information is ignored by the decomposition, we can assume how much complexity of the whole system is lost. This gives us a way to measure the complexity of a system with respect to the system decomposition.

In stochastic systems described as a set of joint distributions, the interaction can basically be expressed as the correlations between the variables. The simplest reductionism approach is to separate the whole system into some subsets of variables, and assume the independence between them. If such decomposition does not affect the system's property, the isolated subsystem is independent from the rest. On the other hand, if the decomposition loses too much information, then the subsystem is inside of a larger subsystem with strong internal interactions and can not be easily separated.

The stochastic models have often been represented with the use of graph representation, and treated with the name of complex network [75, 76, 73]. Generally, the nodes represent the variables and the weights on the edges are the interactions between them. However, if we consider the information contained in the different orders of interactions among variables, the graph with a single kind of edges is not sufficient to express the whole information of the system [79]. An edge of a graph with n nodes contains the information of the interactions up to the n-th order interactions among n variables. If we try to decompose the system independently by cutting the interaction, we have to consider what it means to cut the edge of the graph from the information theoretical point of view.

In this paper, we consider the stochastic system with binary variables and theoretically develop a way to measure the information between subsystems, which is consistent to the information represented by the edges of the graph representation. A case study on the political weblog data is also presented.

10.2 System Decomposition

Let us consider the stochastic system with n binary variables $\mathbf{x} = (x_1, \cdots, x_n)$ where $x_i \in \{0, 1\}$ $(1 \leq i \leq n)$. We denote the joint distribution of $\mathbf{x}$ by $p(\mathbf{x})$. We define the decomposition $p^{dec}(\mathbf{x})$ of $p(\mathbf{x})$ into two subsystems $\mathbf{y^1} = (x_1^1, \cdots, x_{n_1}^1)$ and $\mathbf{y^2} = (x_1^2, \cdots, x_{n_2}^2)$ $(n_1 + n_2 = n$, $\mathbf{y^1} \cup \mathbf{y^2} = \mathbf{x}$, $\mathbf{y^1} \cap \mathbf{y^2} = \phi)$ as follows:

$$p^{dec}(\mathbf{x}) = p(\mathbf{y^1})p(\mathbf{y^2}), \tag{10.1}$$

where $p(\mathbf{y^1})$ and $p(\mathbf{y^2})$ are the joint distributions of $\mathbf{y^1}$ and $\mathbf{y^2}$, respectively. For simplicity, hereafter we denote the system decomposition using the smallest subscript of variables in each subsystem. For example, in case $n = 4$, $\mathbf{y^1} = (x_1, x_3)$ and $\mathbf{y^2} = (x_2, x_4)$, we describe the decomposed system $p^{dec}(\mathbf{x})$ as $< 1212 >$. The system decomposition means to cut all interactions between the two subsystems, which is expressed as setting the independent relation between them.

We will further consider the equation (10.1) in terms of the graph representation. We define the undirected graph $\Gamma := (V, E)$ of the system $p(\mathbf{x})$, whose vertices $V = \{x_1, \cdots, x_n\}$ and edges $E = V \times V$ represent the variables and the interaction, respectively. To express the system, we set the value of each vertex as the value of the corresponding variable, and the weight of each edge as the degree of interaction between the connected variables.

There is however a problem considering the representation with a single kind of edge. The interactions among variables are not only between two variables, but can be independently defined among plural variables up to the n-th order. Therefore, the exact definition of the weight of the edges remains unclear. To clarify these problematics, we consider the hierarchical marginal distributions η as another coordinates of the system $p(\mathbf{x})$ as follows:

$$\eta = (\eta^{\mathbf{1}}; \eta^{\mathbf{2}}; \cdots; \eta^{\mathbf{n}}), \tag{10.2}$$
$$\tag{10.3}$$

where

$$\eta^{\mathbf{1}} = (\eta_1, \cdots, \eta_i, \cdots, \eta_n), (1 < i < n), \tag{10.4}$$
$$\eta^{\mathbf{2}} = (\eta_{12}, \cdots, \eta_{ij}, \cdots, \eta_{n-1n}), (1 < i < j < n), \tag{10.5}$$
$$\vdots$$
$$\eta^{\mathbf{n}} = \eta_{12\cdots n}, \tag{10.6}$$
$$\tag{10.7}$$

and

$$\eta_1 = \sum_{i_2,\cdots,i_n \in \{0,1\}} p(1, i_2, \cdots, i_n), \tag{10.8}$$

$$\vdots$$

$$\eta_n = \sum_{i_1,\cdots,i_{n-1} \in \{0,1\}} p(i_1, \cdots, i_{n-1}, 1), \tag{10.9}$$

$$\eta_{12} = \sum_{i_3,\cdots,i_n \in \{0,1\}} p(1, 1, i_3, \cdots, i_n), \tag{10.10}$$

$$\vdots$$

$$\eta_{n-1n} = \sum_{i_1,\cdots,i_{n-2} \in \{0,1\}} p(i_1, \cdots, i_{n-2}, 1, 1), \tag{10.11}$$

$$\vdots$$

$$\eta_{12\cdots n} = p(1, 1, \cdots, 1). \tag{10.12}$$

Since the definition of η is a linear transformation of $p(\mathbf{x})$, both coordinates have the degrees of freedom $\sum_{k=1}^{n} {}_nC_k$.

The subcoordinates $\eta^{\mathbf{1}}$ is simply the set of marginal distributions of each variable. The subcoordinates $\eta^{\mathbf{k}}$ $(1 < k \leq n)$ include the interactions among k variables, which can not be expressed with the coordinates less than the k-th order. This means that the different interaction exist independently in each order among the corresponding sets of the variables. The interaction represented by the weight of a graph edge $\{x_i, x_j\}$ is therefore the superposition of the different interactions defined on every subset of $\mathbf{x}$ including x_i and x_j.

To measure the degree of interaction in each order, the information geometry established the following setting [77]. We first define another coordinates $\theta = (\theta^{\mathbf{1}}; \theta^{\mathbf{2}}; \cdots; \theta^{\mathbf{n}})$ which are the Legendre transformation of η mediated by the function $\psi(\theta)$ and $\phi(\eta)$ as follows:

$$\theta^{\mathbf{1}} = (\theta_1, \cdots, \theta_n), \tag{10.13}$$

$$\theta^{\mathbf{2}} = (\theta_{12}, \cdots, \theta_{n-1n}), \tag{10.14}$$

$$\vdots$$

$$\theta^{\mathbf{n}} = \theta_{12\cdots n}, \tag{10.15}$$

$$\tag{10.16}$$

where

$$\psi(\theta) = \log \frac{1}{p(0, \cdots, 0)}, \tag{10.17}$$

$$\phi(\eta) = \sum_i \theta_i \eta_i + \sum_{i<j} \theta_{ij} \eta_{ij} + \cdots + \theta_{12\cdots n} \eta_{12\cdots n} - \psi(\theta), \tag{10.18}$$

$$\theta_i = \frac{\partial \phi(\eta)}{\partial \eta_i}, (1 \leq i \leq n), \tag{10.19}$$

$$\theta_{ij} = \frac{\partial \phi(\eta)}{\partial \eta_{ij}}, (1 \leq i < j \leq n), \tag{10.20}$$

$$\vdots$$

$$\theta_{12\cdots n} = \frac{\partial \phi(\eta)}{\partial \eta_{12\cdots n}}. \tag{10.21}$$

Using the coordinates θ, the system is described in the form of the exponential family as follows:

$$p(\mathbf{x}) = \sum_i \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j + \cdots + \theta_{12\cdots n} x_1 x_2 \cdots x_n - \psi(\theta). \tag{10.22}$$

The information geometry revealed that the exponential family of probability distribution forms a manifold with a dual-flat structure. More precisely, the coordinates θ form a flat manifold with respect to the Fisher information matrix as the Riemannian metric, and α-connection with $\alpha = 1$. Dually to θ, the coordinates η are flat with respect to the same metric but α-connection with $\alpha = -1$. It is known that θ and η are orthogonal to each other with respect to the Fisher information matrix. This structure give us a way to decompose the degree of interactions among variables into separated elements of arbitrary orders. We define the so-called k-cut mixture coordinates $\zeta^{\mathbf{k}}$ as follows [80].

$$\zeta^{\mathbf{k}} = (\eta^{\mathbf{k}-}; \theta^{\mathbf{k}+}), \tag{10.23}$$
$$\eta^{\mathbf{k}-} = (\eta^{\mathbf{1}}, \cdots, \eta^{\mathbf{k}}), \tag{10.24}$$
$$\theta^{\mathbf{k}+} = (\theta^{\mathbf{k+1}}, \cdots, \theta^{\mathbf{n}}). \tag{10.25}$$

We also define the k-cut mixture coordinates $\zeta^{\mathbf{k}}_{\mathbf{0}} = (\eta^{\mathbf{k}-}; 0, \cdots, 0)$ with no interaction above the k-th order. We denote the system specified with $\zeta^{\mathbf{k}}$ and $\zeta^{\mathbf{k}}_{\mathbf{0}}$ as $p(\mathbf{x}, \zeta^{\mathbf{k}})$ and $p(\mathbf{x}, \zeta^{\mathbf{k}}_{\mathbf{0}})$, respectively.

Then the degree of the interaction more than the k-th order in the system can be measured by the Kullback-Leibler (KL-) divergence $D[p(\mathbf{x}, \zeta) : p(\mathbf{x}, \zeta^{\mathbf{k}}_{\mathbf{0}})]$. This is also asymptotically compatible with the classical χ^2 test of the degree of freedom $\sum_{i=k+1}^{n} {}_nC_i$ with enough sample number N:

$$2N \cdot D[p(\mathbf{x}, \zeta) : p(\mathbf{x}, \zeta^{\mathbf{k}}_{\mathbf{0}})] \sim \chi^2(\sum_{i=k+1}^{n} {}_nC_i). \tag{10.26}$$

Here, the decomposition is performed according to the orders of interactions, which does not spatially distinguish the vertices. If we define the weight of an edge $\{x_i, x_j\}$ with the KL-divergence, the above k-cut coordinates $\zeta^{\mathbf{k}}$ are not appropriate to measure the information represented in each edge. We need to set another mixture coordinates so that to separate only the existing information between x_i and x_j regardless of its order.

Let us return to the definition of the system decomposition and consider on the dual-flat coordinates θ and η. The independence between the two decomposed systems $\mathbf{y^1} = (x^1_1, \cdots, x^1_{n_1})$ and $\mathbf{y^2} = (x^2_1, \cdots, x^2_{n_2})$ can be expressed on the new coordinates $\eta^{\mathbf{dec}}$ as follows:

$$\eta^{dec}_i = \eta_i, (1 \le i \le n), \tag{10.27}$$
$$\eta^{dec}_{ij} = \begin{cases} \eta_{ij}, (1 \le i < j \le n), & \text{if } \{x_i, x_j\} \subseteq \mathbf{y^1} \text{ or } \subseteq \mathbf{y^2} \\ \eta_i \eta_j, (1 \le i < j \le n), & \text{else} \end{cases}, \tag{10.28}$$
$$\eta^{dec}_{ijk} = \begin{cases} \eta_{ijk}, (1 \le i < j < k \le n), & \text{if } \{x_i, x_j, x_k\} \subseteq \mathbf{y^1} \text{ or } \subseteq \mathbf{y^2} \\ \eta_{ij}\eta_k, (1 \le i < j < k \le n), & \text{else if } \{x_i, x_j\} \subseteq \mathbf{y^1} \text{ or } \subseteq \mathbf{y^2} \\ \eta_i \eta_{jk}, (1 \le i < j < k \le n), & \text{else if } \{x_j, x_k\} \subseteq \mathbf{y^1} \text{ or } \subseteq \mathbf{y^2} \\ \eta_j \eta_{ik}, (1 \le i < j < k \le n), & \text{else (if } \{x_i, x_k\} \subseteq \mathbf{y^1} \text{ or } \subseteq \mathbf{y^2}) \end{cases}, \tag{10.29}$$
$$\vdots$$
$$\eta^{dec}_{12\cdots n} = \eta_{s[ik_1\cdots k_{n_1-1}]} \eta_{s[jl_1\cdots l_{n_2-1}]}, (x_i \in \mathbf{y^1}, x_j \in \mathbf{y^2}), \tag{10.30}$$

where $s[\cdots]$ is the ascending sort of the internal sequence.

The definition of $\eta^{\mathbf{dec}}$ means to decompose the hierarchical marginal distributions η into the products of the subsystems' marginal distributions, in case the subscripts traverse the two subsystems. Therefore, only the interactions between two subsystems are set to be independent, while the internal interactions of each subsystem remain unchanged. This is analytically equivalent to

compose another mixture coordinates ξ, namely the $< \cdots >$-cut coordinates, with proper description of the system decomposition with $< \cdots >$. The ξ consist of the η coordinates with subscripts which does not traverse the subsystems, and the θ coordinates whose subscripts traverse the subsystems. For simplicity, we only describe here the case $n = 4$ and the decomposition $< 1133 >$ (the set of the first, second, and the third, fourth nodes each form a subsystem). The system $p(\mathbf{x})$ is expressed with the $< 1133 >$-cut coordinates ξ as

$$\xi_1 = \eta_1, \tag{10.31}$$
$$\vdots$$
$$\xi_4 = \eta_4, \tag{10.32}$$
$$\xi_{12} = \eta_{12}, \tag{10.33}$$
$$\xi_{13} = \theta_{13}, \tag{10.34}$$
$$\xi_{14} = \theta_{14}, \tag{10.35}$$
$$\xi_{23} = \theta_{23}, \tag{10.36}$$
$$\xi_{24} = \theta_{24}, \tag{10.37}$$
$$\xi_{34} = \eta_{34}, \tag{10.38}$$
$$\xi_{123} = \theta_{123}, \tag{10.39}$$
$$\vdots$$
$$\xi_{234} = \theta_{234}, \tag{10.40}$$
$$\xi_{1234} = \theta_{1234}. \tag{10.41}$$

The decomposed system with no interactions between two subsystems have the following coordinates $\xi^{\mathbf{dec}}$, which is, in any decomposition, equivalent to set all θ in ξ as 0 :

$$\xi_1^{dec} = \eta_1, \tag{10.42}$$
$$\vdots$$
$$\xi_4^{dec} = \eta_4, \tag{10.43}$$
$$\xi_{12}^{dec} = \eta_{12}, \tag{10.44}$$
$$\xi_{13}^{dec} = 0, \tag{10.45}$$
$$\xi_{14}^{dec} = 0, \tag{10.46}$$
$$\xi_{23}^{dec} = 0, \tag{10.47}$$
$$\xi_{24}^{dec} = 0, \tag{10.48}$$
$$\xi_{34}^{dec} = \eta_{34}, \tag{10.49}$$
$$\xi_{123}^{dec} = 0, \tag{10.50}$$
$$\vdots$$
$$\xi_{234}^{dec} = 0, \tag{10.51}$$
$$\xi_{1234}^{dec} = 0. \tag{10.52}$$

This is analytically equivalent to the definition of the decomposition (10.27)-(10.30) in case of $< 1133 >$. Therefore, the KL-divergence $D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{\mathbf{dec}})]$ measures the information lost by the system decomposition. The following asymptotic agreement to χ^2 test also holds.

$$2N \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{\mathbf{dec}})] \sim \chi^2(\sharp_\theta(\xi)), \tag{10.53}$$

where $\sharp_\theta(\xi)$ is the number of θ coordinates appearing in the ξ coordinates.

10.3 Edge Cutting

We further expand the concept of system decomposition to eventually quantify the total amount of information expressed by an edge of the graph. Let us consider to cut an edge $\{x_i, x_j\}$ $(1 \leq i < j \leq n)$ of the graph with n vertices. Hereafter we call this operation as the edge cutting $i-j$. In the same way as the system decomposition, the edge cutting corresponds to modify the η coordinates to produce $\eta^{\mathbf{ec}}$ coordinates as follows:

$$\eta^{ec}_{ij} = \eta_i \eta_j, \tag{10.54}$$
$$\eta^{ec}_{s[ijk_1]} = \eta_{s[ik_1]} \eta_{s[jk_1]}, \tag{10.55}$$
$$\eta^{ec}_{s[ijk_1k_2]} = \eta_{s[ik_1k_2]} \eta_{s[jk_1k_2]}, \tag{10.56}$$
$$\vdots$$
$$\eta^{ec}_{s[ijk_1\cdots k_{n-2}]} = \eta_{s[ik_1\cdots k_{n-2}]} \eta_{s[jk_1\cdots k_{n-2}]}, \tag{10.57}$$
$$(\{i, j, k_1, \cdots, k_{n-2}\} = \{1, \cdots, n\}),$$

and the rest of $\eta^{\mathbf{ec}}$ remains the same as those of η.

The formation of η^{ec} from η consists of replacing the k-th order elements ($k \geq 3$) of η including both i and j in its subscripts, with the product of the $k-1$-th order η in maximum subgraphs ($k-1$ vertices) each including i or j. This means that all orders of interactions including the variables x_i and x_j are set to be independent only between them. Other interactions which do not include simultaneously x_i and x_j remain unchanged.

Certain combinations of edge cuttings coincide with system decompositions. For example, in case $n = 4$, the edge cuttings $1-2$, $1-3$, and $1-4$ are equivalent to the system decomposition $< 1222 >$.

We define the $i-j$-cut mixture coordinates ξ for orthogonal decomposition of the interaction represented by the edge $\{x_i, x_j\}$. Although actual calculation can be performed only with η coordinates, this generalization is necessary to have a geometrical definition of the orthogonality. For simplicity, we only describe the ξ in the case of $n = 4$:

$$\xi_1 = \eta_1, \tag{10.58}$$
$$\vdots$$
$$\xi_4 = \eta_4, \tag{10.59}$$
$$\xi_{12} = \theta_{12}, \tag{10.60}$$
$$\xi_{13} = \eta_{13}, \tag{10.61}$$
$$\xi_{14} = \eta_{14}, \tag{10.62}$$
$$\xi_{23} = \eta_{23}, \tag{10.63}$$
$$\xi_{24} = \eta_{24}, \tag{10.64}$$
$$\xi_{34} = \eta_{34}, \tag{10.65}$$
$$\xi_{123} = \theta_{123}, \tag{10.66}$$
$$\xi_{124} = \theta_{124}, \tag{10.67}$$
$$\xi_{134} = \eta_{134}, \tag{10.68}$$
$$\xi_{234} = \eta_{234}, \tag{10.69}$$
$$\xi_{1234} = \theta_{1234}, \tag{10.70}$$

where orthogonality between the elements of η and θ holds with respect to the Fisher information matrix.

Calculating the dual coordinates $\theta^{\mathbf{ec}}$ of $\eta^{\mathbf{ec}}$, we can define the coordinates $\xi^{\mathbf{ec}}$ of the system after the edge cutting $1-2$ as follows:

$$\xi^{ec}_1 = \eta_1, \tag{10.71}$$
$$\vdots$$
$$\xi^{ec}_4 = \eta_4, \tag{10.72}$$
$$\xi^{ec}_{12} = \theta^{ec}_{12}, \tag{10.73}$$
$$\xi^{ec}_{13} = \eta_{13}, \tag{10.74}$$
$$\xi^{ec}_{14} = \eta_{14}, \tag{10.75}$$
$$\xi^{ec}_{23} = \eta_{23}, \tag{10.76}$$
$$\xi^{ec}_{24} = \eta_{24}, \tag{10.77}$$
$$\xi^{ec}_{34} = \eta_{34}, \tag{10.78}$$
$$\xi^{ec}_{123} = \theta^{ec}_{123}, \tag{10.79}$$
$$\xi^{ec}_{124} = \theta^{ec}_{124}, \tag{10.80}$$
$$\xi^{ec}_{134} = \eta_{134}, \tag{10.81}$$
$$\xi^{ec}_{234} = \eta_{234}, \tag{10.82}$$
$$\xi^{ec}_{1234} = \theta^{ec}_{1234}. \tag{10.83}$$

Note that the edge cutting can not be defined simply by setting the corresponding elements of $\theta^{\mathbf{ec}}$ as 0.

Then the KL-divergence $D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{\mathbf{ec}})]$ represent the total amount of information represented by the edge $1 - 2$.

The following asymptotic agreement to χ^2 test also holds:

$$2N \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{\mathbf{ec}})] \sim \chi^2(1 + \sum_{k=1}^{n-2} {}_{n-2}C_k). \tag{10.84}$$

We call this χ^2 value or the KL divergence itself as edge information of edge $1 - 2$.

10.4 Case Study on Political Weblog Data

We applied the system decomposition and the edge cutting for the network analysis in actual data. As a case study, we have calculated all possible system decompositions and edge cuttings for the following data set with $n = 4$ binary variables.

10.4.1 Data Set

The campaign for the French presidential election which ended in May 2007 was marked by the strengthening of new modes of public expression. As the access to new communication media was democratizing, new patterns of public intervention have been invented. The main outcome of this new mode of intervention was that citizens could express their own views about social, political or economical topics concerning the campaign which in turn modified deeply the political agenda. The political blogosphere, in particular, is one of the main places of appropriation of public debate.

We have collected the content and hypertext links of a selection of 120 political blogs during the first six months of 2007 [1]. The database allows each blog to provide a time series of semantic attributes (quantified using a vector of occurrences of 190 key terms) that evolve according to new posts. This provides us a daily semantic characterization of each blog from January the 1^{st} to June the 30^{th} reflecting the current concerns of each blogger. The correspondence between the date of important issues and the daily time axis is presented in Tab. 10.1.

We have analyzed these 120 weblog text data focusing on the 4 candidates (Ségolène Royal, François Bayrou, Nicolas Sarkozy, and Jean-Marie Le Pen) of the last French presidential. For

[1] The data were collected by Jean-Philippe Cointet.

Table 10.1: **List of important issues during the election.** For simplicity of the later analysis, we introduced a daily time axis to give a decimal notation. The source in French: http://www.presidentielle-2007.net/calendrier.php

issue	Date	Decimal date
Deadline for inscription to the list of candidates	31 Dec 2006	0
Decree of summons to candidates	22 Feb 2007	53
Dispatch of bulletins of sponsorship to candidates	22 Feb 2007	53
Deadline for registration of 500 sponsorships	16 Mar 18:00	75
Publication of candidate list	19 Mar 2007	78
Start of official campaign	9 Avr 2007	99
Publication of sponsorships	10 Avr 2007	100
End of official campaign	20 Avr 2007	110
First round of presidential	22 Avr 2007	112
Second round of presidential	6 Mai 2007	127
Start of new presidential term of office	17 Mai 2007	137
General elections	10 & 17 Jun 2007	161 & 168

simplicity, we will take the abbreviation of their family name, R,B,S,L to denote each candidate. We will use these abbreviations as subscript for the binary occurrence vector (x_R, x_B, x_S, x_L) aligned from left to right according to their political position being leftist to rightist.We set the first or family name of each candidate as the key terms to scan the weblogs and obtain the data. Each of 120 blogs was crawled each day to see whether these names appeared or not in their posts, and coded accordingly as 1 (occurrence) and 0 (non-occurrence). For example, if a blog post contained the words "Royal" and "Sarkozy" but none of the others, then the occurrence data for this blog in this day will be $(x_R, x_B, x_S, x_L) = (1, 0, 1, 0)$. Using this occurrence vectors, we defined the system $p(x_1, x_2, x_3, x_4) = p(x_R, x_B, x_S, x_L)$.

10.4.2 Result and Discussion

We have calculated the ξ^2 value of the equation (10.84) for the edge cuttings, and that of the equation (10.53) for the system decompositions. We used three time scales, daily, weekly, and biweekly ones to obtain the distribution $p(\mathbf{x})$. Concretely, the distribution of each day was calculated using the data of following 7 and 14 days for the weekly and biweekly scale, respectively. The result are shown in Figs. 10.1-10.5.

Comparing the three different time scales, the weekly scale seems to be appropriate for the characterization of the dynamics. The daily scale contains much fluctuation which blurs the global dynamics, and the biweekly scale loses time resolution too much, since most of the intervals between important events are less than two weeks. Indeed, most of the weblogs generally contain the history of the past several days in the same page, so that it is natural to consider the weekly time scale to obtain the distribution.

Some general tendencies are observed in the temporal dynamics of all results. The χ^2 values tend to augment continuously as the campaign goes on, and reached its peak at between the start of the official campaign (99th day) and the second round of the presidential (127th day). This reflects well the interest of the people on the presidential, as the degree of interactions are augmenting toward the peak at the decisive period. All edge cuttings and system decompositions loses highly significant information with respect to the χ^2 test. This is a strongly interacting system, and it is difficult to apply the reductionism to the joint distribution. The exception is the daily fluctuation and the dynamics after the second round. Significant breakdown of interest comes only after the second round (127th day), where S was finally chosen as the new president.

The result of the edge cuttings is shown in Fig.10.1. Until the publication of the candidate list (78th day), there exists approximately three levels of dynamics in weekly and biweekly scale, and different edge cuttings are migrating among them. This rather discontinuous hopping of edge informations may be interpreted as the dynamics among different rankings of the topics. The mechanism of discontinuity is of further question.

Edges between R-S is the strongest at the beginning, therefore the result of the first round where R and S were chosen for the second round reflects well the advance reviews of this term. While during the official campaign (99th to 110th day), all edges contain similarly high information, which signifies the discussion are not localized only between R and S.

The result of the system decompositions are shown in Figs.10.2-10.4. In Fig.10.2, the dynamics of $< 1222 >$ and $< 1131 >$ are almost synchronized and express the highest degree of interaction. This means that R and S are each highly interacting with the rest of the system. In other word, one can not isolate R or S without destroying the system's strongest interactions. Later we will see that the strongest one is between R and S. The highest value of $< 1222 >$ and $< 1131 >$ are observed between the first round (112th day) and the second round (127th day), where R and S had one-on-one competition including TV debate.

Fig.10.3 and 10.4 show the result of system decompositions into two half and three parts, respectively. The decomposition $< 1212 >$ in Fig.10.3 and $< 1214 >$ in Fig. 10.4 have the lowest degree of interaction. This means that it is rather easier to isolate R and S as a subsystem from the rest. Combining with the result of Fig.10.2, R and S are highly interacting with each other, but rather less with the rest. Although, in any period until the decision of the mew president, all system decompositions are highly significant, which does not allow to perform any decomposition without losing significant information.

Such hierarchical comparison of system decompositions can provide a novel way of clustering. In this case, for example, the R and S have the strongest interaction with each other, then the second strongest with B, and the third with L. Choosing an threshold, one can hierarchically cluster the four candidates according to the degree of interactions.

The degree of interactions with respect to the total decomposition of the system into $< 1234 >$ is shown in Fig. 10.5. The dynamics represents the evolution of complexity with respect to the total decomposition of the system. This is the total amount of information arising from the interactions among variables. The peak of the complexity is observed around the first round (112th day), which can be estimated as the peak of public interest. Although the result of the second round (127th day) was not easy to predict, it did not evoke further augmentation of total information.

Finally, we revisit the result of order-wise interaction in this data and compare the result of edge cutting with that of simple second-order interaction. Edge cutting contains more information than the simple second-order interaction between two nodes, since it takes also the interaction higher than second order in the expanded subgraphs into account. With this respect, we can observe how the 4th- and 3rd-order interactions are incorporated in edge cutting. Figs. 10.6 are the reprint of order-wise interaction dynamics in biweekly scale. Compared to the Fig. 10.1 Bottom, irregularity of the edge cutting dynamics appear to be affected by the information higher than the second order. Edge cutting dynamics show another way to express these higher-order interactions, preserving the graph expression with a single kind of link.

10.5 Further Consideration: Pythagorean Relations in System Decomposition and Edge Cutting

We further look back at the system decomposition and edge cutting in terms of the Pythagorean relation between KL divergences, which is based on the orthogonality between θ and η coordinates.

In system decomposition, the distribution of decomposed system is analytically obtained from the product of subsystems' η coordinates, which is equivalent to set all θ^{dec} parameters as 0 in mixture coordinate ξ^{dec}. From the consistency of θ^{dec} parameters in ξ^{dec} being 0 in all system decompositions, we have the Pythagorean relation according to the inclusion relation of system decomposition. For example, the following holds:

$$
\begin{aligned}
D[< 1111 >:< 1234 >] &= D[< 1111 >:< 1222 >] && (10.85)\\
&+ D[< 1222 >:< 1233 >] && (10.86)\\
&+ D[< 1233 >:< 1234 >]. && (10.87)
\end{aligned}
$$

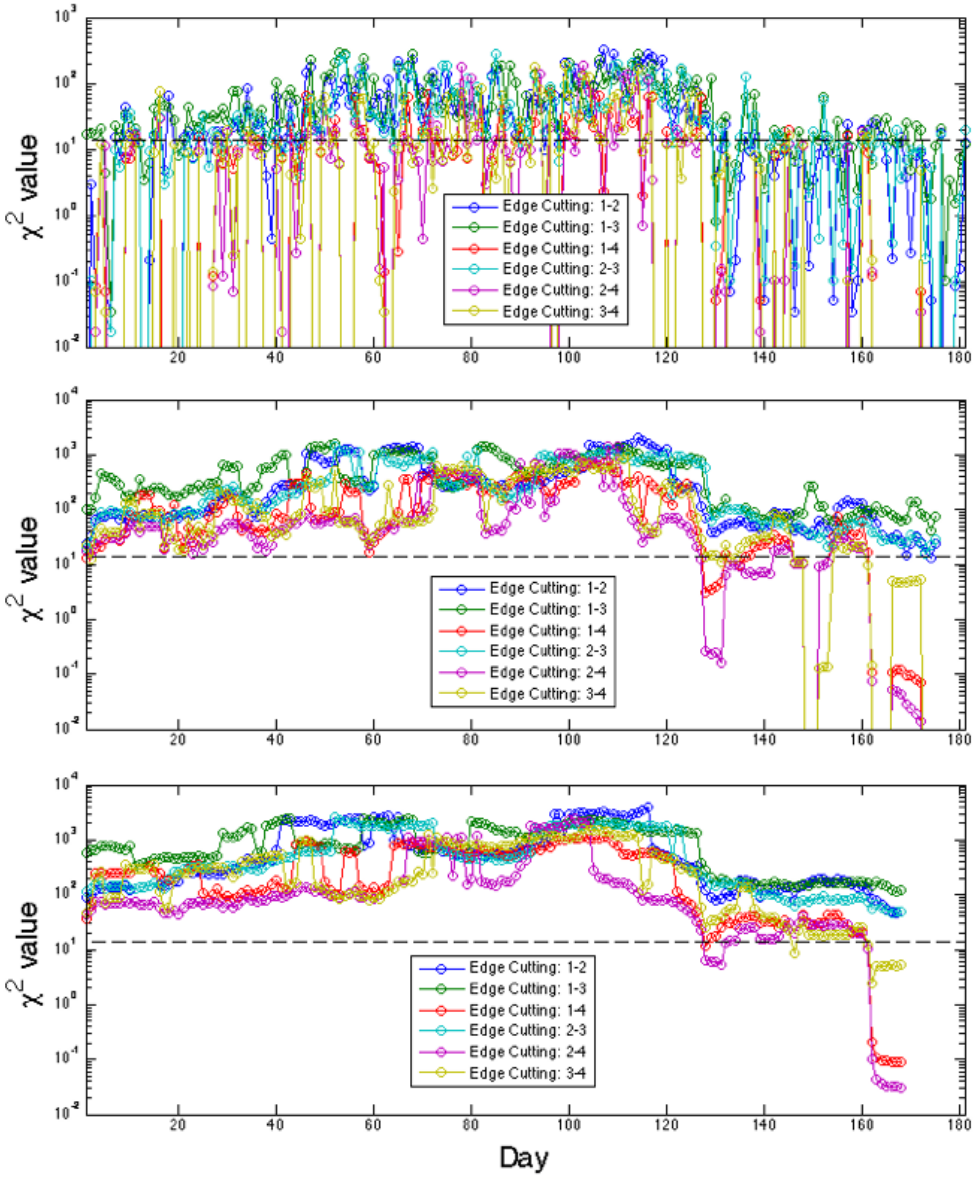


Figure 10.1: **Degree of interaction lost by edge cuttings.** Top: $p(\mathbf{x})$ was taken as daily scale distribution. Middle: Weekly scale. Bottom: Biweekly scale. Vertical axis: The χ^2 value of the equation (10.84). Horizontal axis: Decimal date in Tab.10.1. The dashed line is the 5 % threshold of χ^2 test. The node numbers 1,2,3,4 correspond to R,B,S,L, respectively.

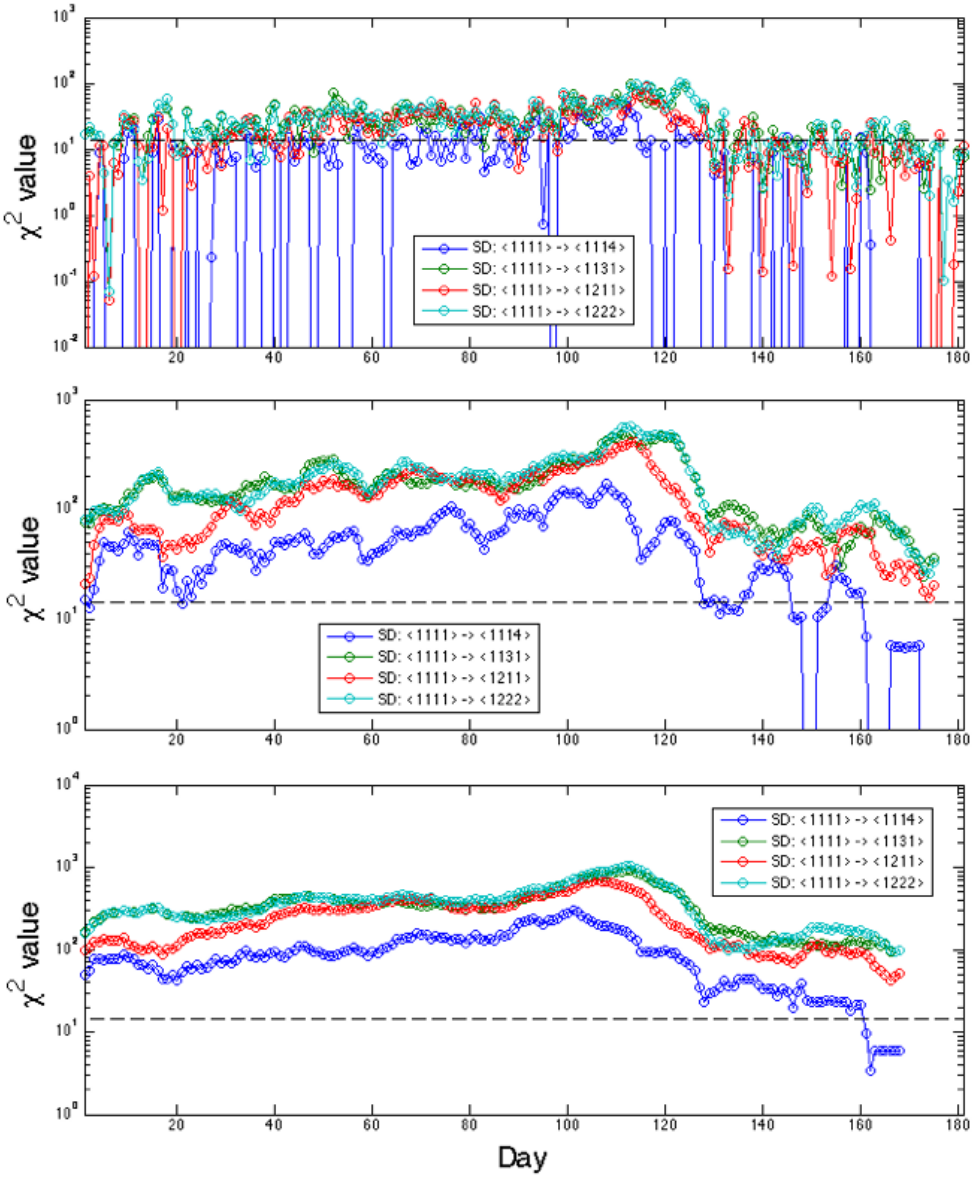


Figure 10.2: **Degree of interaction lost by system decompositions into 1- and 3-nodes subsystems.** Top: $p(\mathbf{x})$ was taken as daily scale distribution. Middle: Weekly scale. Bottom: Biweekly scale. Vertical axis: The χ^2 value of the equation (10.53). Horizontal axis: Decimal date in Tab.10.1. The dashed line is the 5 % threshold of χ^2 test.

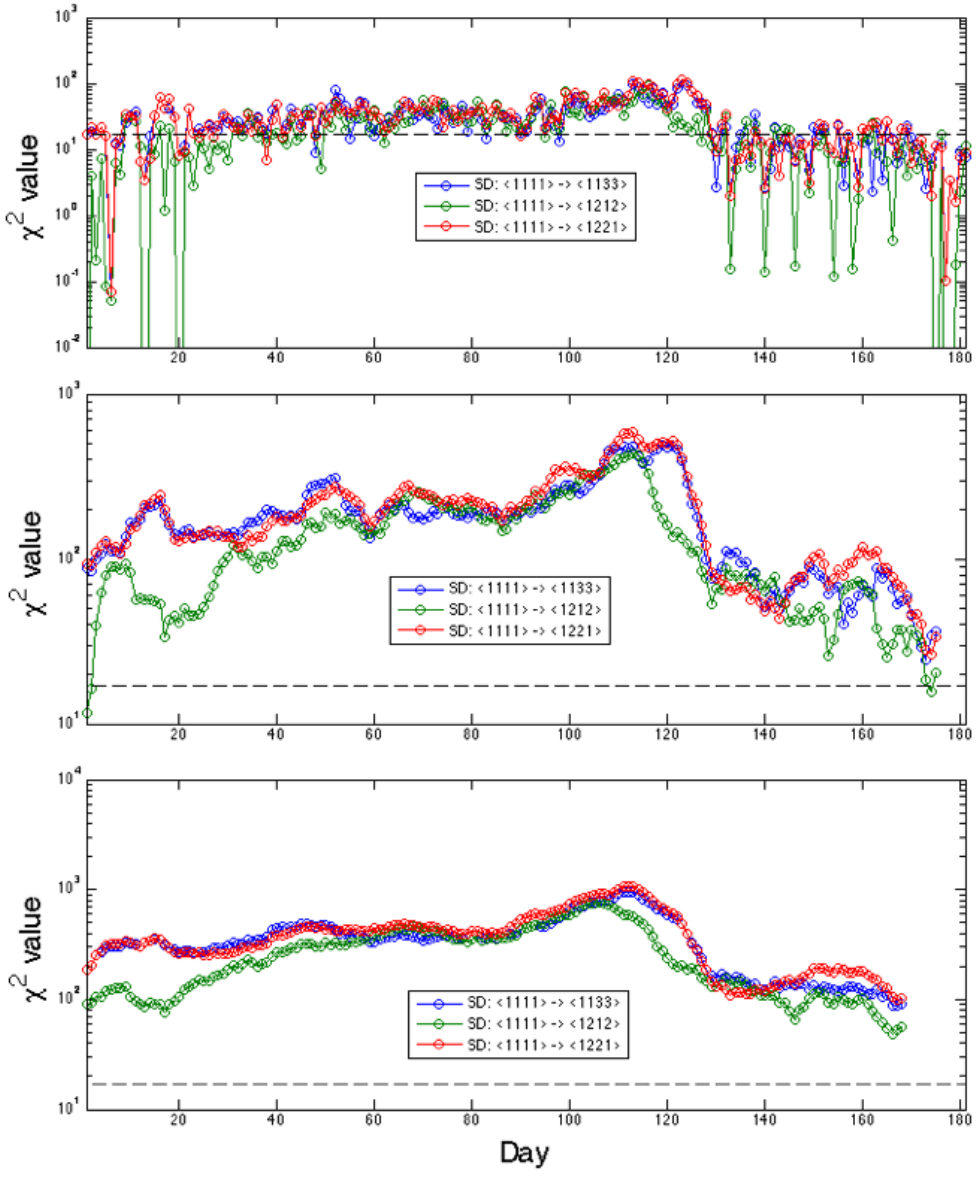


Figure 10.3: **Degree of interaction lost by system decompositions into two 2-nodes subsystems.** Top: $p(\mathbf{x})$ was taken as daily scale distribution. Middle: Weekly scale. Bottom: Biweekly scale. Vertical axis: The χ^2 value of the equation (10.53). Horizontal axis: Decimal date in Tab.10.1. The dashed line is the 5 % threshold of χ^2 test.

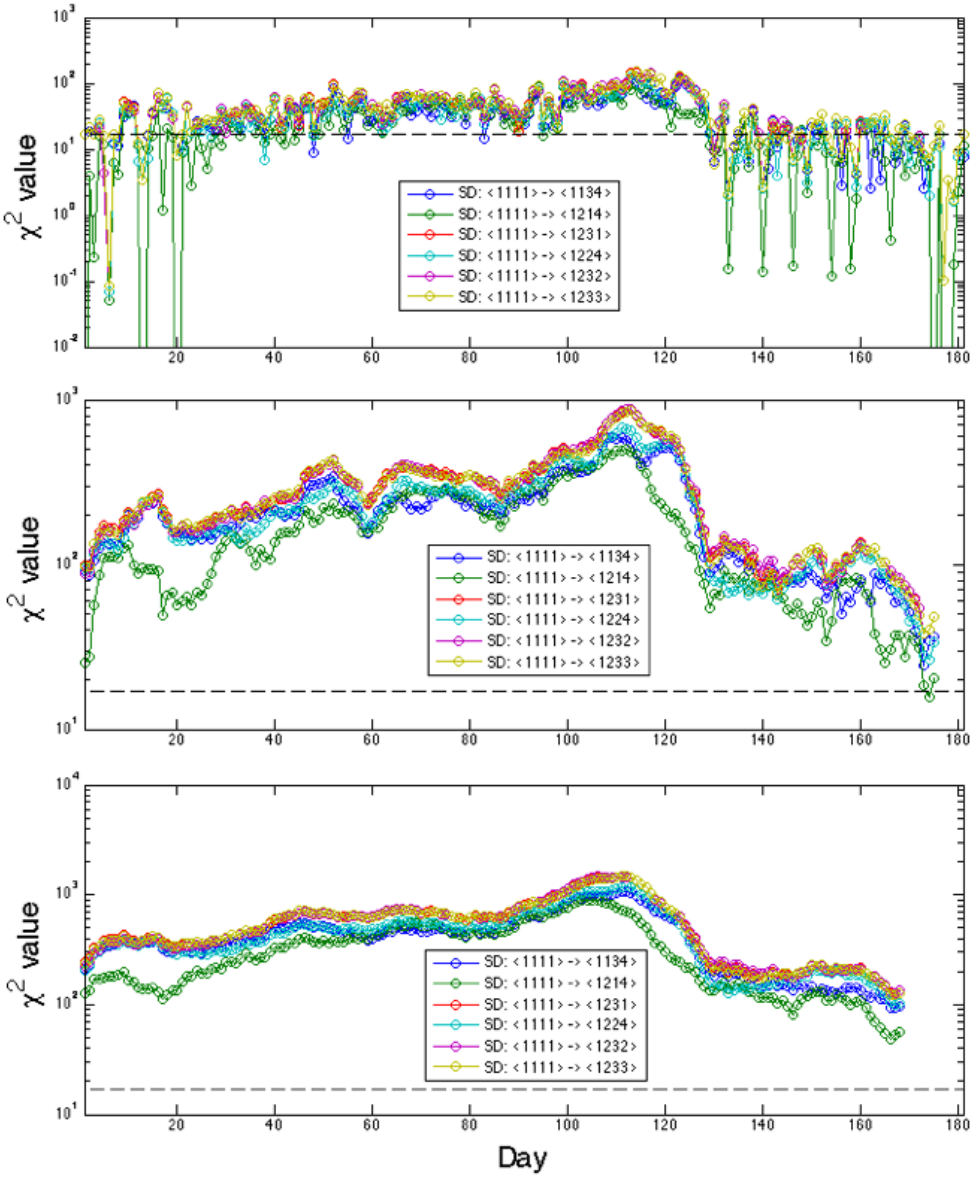


Figure 10.4: **Degree of interaction lost by system decompositions into three subsystems.** Top: $p(\mathbf{x})$ was taken as daily scale distribution. Middle: Weekly scale. Bottom: Biweekly scale. Vertical axis: The χ^2 value of the equation (10.53). Horizontal axis: Decimal date in Tab.10.1. The dashed line is the 5 % threshold of χ^2 test.

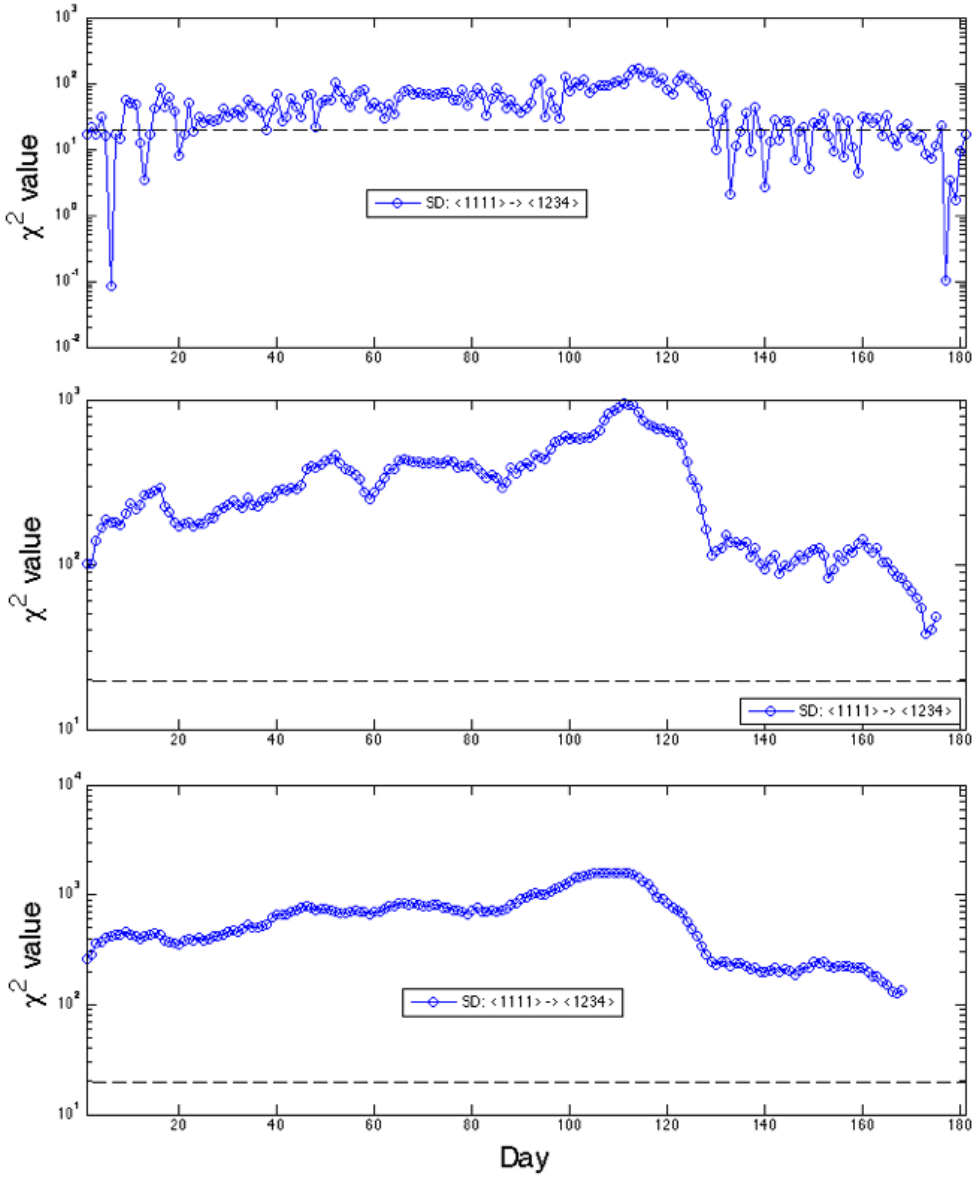


Figure 10.5: **Degree of interaction lost by system decomposition** $< 1234 >$**.** Top: $p(\mathbf{x})$ was taken as daily scale distribution. Middle: Weekly scale. Bottom: Biweekly scale. Vertical axis: The χ^2 value of the equation (10.53). Horizontal axis: Decimal date in Tab.10.1. The dashed line is the 5 % threshold of χ^2 test.

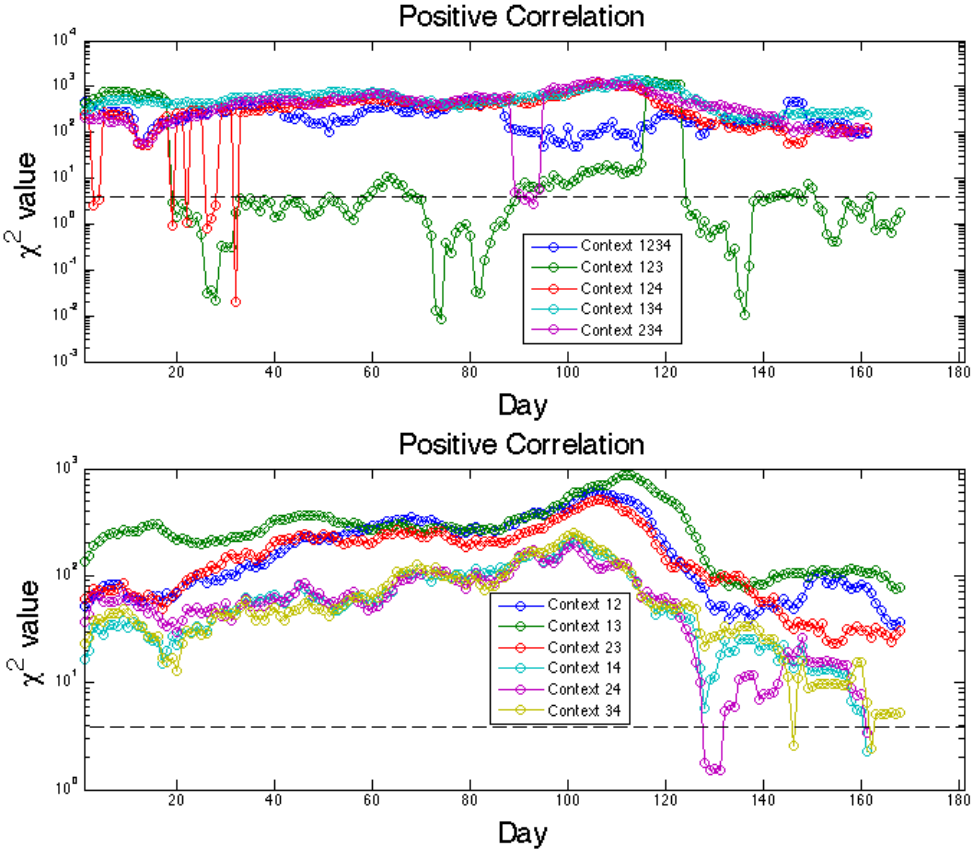


Figure 10.6: **Reprint of order-wise interaction dynamics in biweekly scale in the previous chapter.** Top: 4th- and 3rd-order contexts. Down: 2nd-order contexts.

The proof is in the same way as k-cut coordinates isolating k-tuple interaction between variables [80].

On the other hand, the edge cutting previously defined using the product of remaining maximum cliques' η coordinates does not coincides with the $\theta^{ec} = 0$ condition in mixture coordinates ξ^{ec}. We have defined the η^{ec} values of edge cutting based only on the orthogonal relation between η and θ coordinates, by generalizing the rule of system decomposition in η^{ec} coordinates, and did not consider the Pythagorean relation between different edge cuttings.

It is then possible to define another way of edge cutting using $\theta^{ec} = 0$ condition in ξ^{ec}. Indeed, in k-cut mixture coordinates, $\theta^{k+} = 0$ condition is derived from the independent condition of the variables in all orders, and k-tuple interaction is measured by reestablishing the η parameters for the interaction up to $k-1$-tuple order. In the same way, we can set $\theta^{dec} = 0$ condition for ξ^{dec} of a system decomposition, and reestablish edges with respect to the η parameters, except the one in focus for edge cutting.

As a simple example, consider the system decomposition $< 1222 >$ and edge cutting $1-2$ in 4-node graph. We have the mixture coordinate ξ^{dec} for the system decomposition as follows:

$$\xi^{dec}_{12} = \theta^{dec}_{12} = 0, \tag{10.88}$$
$$\xi^{dec}_{13} = \theta^{dec}_{13} = 0, \tag{10.89}$$
$$\xi^{dec}_{14} = \theta^{dec}_{14} = 0, \tag{10.90}$$
$$\xi^{dec}_{123} = \theta^{dec}_{123} = 0, \tag{10.91}$$
$$\xi^{dec}_{134} = \theta^{dec}_{134} = 0, \tag{10.92}$$
$$\xi^{dec}_{1234} = \theta^{dec}_{1234} = 0, \tag{10.93}$$

where all the rest of ξ^{dec} coordinates is equivalent to that of η coordinates.

We then consider the new way of edge cutting $1-2$ by recovering the interaction in edges $1-3$ and $1-4$ from system decomposition $< 1222 >$, orthogonally to the interaction of edge $1-2$. The new mixture coordinate ξ^{EC} changes to the following:

$$\xi^{EC}_{12} = \theta^{EC}_{12} = 0, \tag{10.94}$$
$$\xi^{EC}_{13} = \eta_{13}, \tag{10.95}$$
$$\xi^{EC}_{14} = \eta_{14}, \tag{10.96}$$
$$\xi^{EC}_{123} = \theta^{EC}_{123} = 0, \tag{10.97}$$
$$\xi^{EC}_{134} = \eta_{134}, \tag{10.98}$$
$$\xi^{EC}_{1234} = \theta^{EC}_{1234} = 0, \tag{10.99}$$

and the rest is equivalent to that of η coordinates.

This new ξ^{EC} is also compatible with k-cut coordinates formalization for its simple $\theta^{EC} = 0$ conditions. To obtain ξ^{EC} for arbitrary edge cutting $i-j$, one should take θ^{EC} containing i and j in its subscript, set them to 0, and combine with η coordinates for the rest of the subscript. For plural edge cuttings $i-j, \cdots, k-l$ $(1 \leq i, j, k, l \leq n)$, it suffices to take θ^{EC} containing i and j, ... , k and l in its subscript respectively, then set them to 0.

We finally obtain the Pythagorean relation between edge cuttings. Denoting the general edge cutting(s) coordinates as $\xi^{i-j,\cdots,k-l}$, the following holds for the example of system decomposition $< 1222 >$:

$$\begin{aligned} D[< 1111 >:< 1222 >] &= D[< 1111 >: p(\xi^{1-2})] & (10.100)\\ &+ D[p(\xi^{1-2}) : p(\xi^{1-2,1-3})] & (10.101)\\ &+ D[p(\xi^{1-2,1-3}) : p(\xi^{1-2,1-3,1-4})]. & (10.102) \end{aligned}$$

Despite the consistency with the dual structure between θ and η, we do not generally have analytical solution to determine η^{EC} values from $\theta^{EC} = 0$ conditions. We should call for some

numerical algorithm to solve $\theta^{EC} = 0$ conditions with respect to η^{EC} values, which are in general high-degree simultaneous polynomials. Furthermore, numerical convergence of the solution have to be very strict, since tiny deviation from the conditions can become nonnegligible by passing fractional function and logarithmic function of θ coordinates.

On the other hand, the previously defined edge cutting with ξ^{ec} using the product between subgraphs' η coordinates is analytically simple and does not need to consider the other edges' recovery from system decomposition or independence hypothesis. We then chose the previous way of edge cutting for both calculability and clarity of the concept.

10.6 Conclusion

We have theoretically developed a framework to measure the degree of interactions existing between subsystems as well as the ones represented by each edge of the graph representation. We applied the theory to characterize the dynamics of the political weblog data, and revealed the hierarchical structure of subsystems according to the degree of interaction. Further empirical analysis on actual data is needed to evaluate the utility of the proposed method.

Chapter 11

Multi-Node Reference Reconstruction of Edge Information Including Substantive Order-Wise Interaction

Abstract

We reconsider the edge information defined in the previous chapter to reconstruct large scale complex networks. Expansions of edge information measure are further derived so that to distinguish between positive/negative order-wise interactions with Pythagorean relation. The unified measure of edge information is also defined. The possibility of dimension reduction according to the substantive information and the utility of proposed edge information measures are investigated with political weblog data.

Keywords: Information geometry, Order-wise interaction, Edge information, Parameter reduction, Political weblog, Social network

Methodology: Decomposition of order-wise edge information and formulation of multi-node reference reconstruction of correlation graph with information geometry → Application to political weblog data as dynamical systems

11.1 Introduction

We consider the application of previously defined edge information to reconstruct in general large scale network from given co-occurrence data between node variables including order-wise interactions.

The first problematics is how to express different orders of interaction, including positive/negative variation, with a single kind of edge weight defined as edge information. The second one is how to reduce the combinatory explosion of parameter number without losing essential information when the system size increases.

For these purposes, we first develop theoretically the decomposition of edge information into different orders, and propose an unified measure of edge information including positive/negative signs of each order.

Second, treating actual data, we evaluate the amount of order-wise interactions to justify the order limit of the model to avoid combinatory explosion of parameters. The result of reconstructed graph with unified edge information as a single kind of edge weight is shown to be accessible for further clustering.

The significance of social dynamics with respect to the order-wise interactions and unified edge informations are also discussed.

11.2 Order-wise Interaction in Edge Information

We revisit the edge information in case of the system size $n = 3$. Following the previous definition, the $1-2$-cut mixture coordinates ξ and that of after edge cutting $1-2$ ξ^{ec} are given as follows:

$$\xi_1 = \eta_1, \tag{11.1}$$
$$\xi_2 = \eta_2, \tag{11.2}$$
$$\xi_3 = \eta_3, \tag{11.3}$$
$$\xi_{12} = \theta_{12}, \tag{11.4}$$
$$\xi_{13} = \eta_{13}, \tag{11.5}$$
$$\xi_{23} = \eta_{23}, \tag{11.6}$$
$$\xi_{123} = \theta_{123}, \tag{11.7}$$

and

$$\xi^{ec}_1 = \eta_1, \tag{11.8}$$
$$\xi^{ec}_2 = \eta_2, \tag{11.9}$$
$$\xi^{ec}_3 = \eta_3, \tag{11.10}$$
$$\xi^{ec}_{12} = \theta^{ec}_{12}, \tag{11.11}$$
$$\xi^{ec}_{13} = \eta_{13}, \tag{11.12}$$
$$\xi^{ec}_{23} = \eta_{23}, \tag{11.13}$$
$$\xi^{ec}_{123} = \theta^{ec}_{123}, \tag{11.14}$$

where θ^{ec} are calculated as the Legendre transformation of the following η^{ec}:

$$\eta^{ec}_1 = \eta_1, \tag{11.15}$$
$$\eta^{ec}_2 = \eta_2, \tag{11.16}$$
$$\eta^{ec}_3 = \eta_3, \tag{11.17}$$
$$\eta^{ec}_{12} = \eta_1\eta_2, \tag{11.18}$$
$$\eta^{ec}_{13} = \eta_{13}, \tag{11.19}$$
$$\eta^{ec}_{23} = \eta_{23}, \tag{11.20}$$
$$\eta^{ec}_{123} = \eta_{13}\eta_{23}, \tag{11.21}$$

We consider the KL divergence $D[p(\mathbf{x},\xi) : p(\mathbf{x},\xi^{ec})]$ which define the amount of edge information. The KL divergence can be decomposed into different orders in the following way, according to the Pythagorean relation:

$$\begin{aligned} & D[p(\mathbf{x},\xi) : p(\mathbf{x},\xi^{ec})] \\ = \; & D[p(\mathbf{x},\xi) : p(\mathbf{x},\xi^{3})] \\ + \; & D[p(\mathbf{x},\xi^{3}) : p(\mathbf{x},\xi^{23})], \end{aligned} \tag{11.22}$$

where the new mixture coordinates ξ^{23} and ξ^3 are defined as follows:

$$\xi^{23}_1 = \eta^{23}_1 = \eta_1, \tag{11.23}$$
$$\xi^{23}_2 = \eta^{23}_2 = \eta_2, \tag{11.24}$$
$$\xi^{23}_3 = \eta^{23}_3 = \eta_3, \tag{11.25}$$
$$\xi^{23}_{12} = \eta^{23}_{12} = \theta^{ec}_{12}, \tag{11.26}$$
$$\xi^{23}_{13} = \eta^{23}_{13} = \eta_{13}, \tag{11.27}$$
$$\xi^{23}_{23} = \eta^{23}_{23} = \eta_{23}, \tag{11.28}$$
$$\xi^{23}_{123} = \theta^{23}_{123} = \theta^{ec}_{123}, \tag{11.29}$$

which are equivalent to denote $\xi^{23} = \xi^{ec}$, and

$$\xi^3_1 = \eta^3_1 = \eta_1, \tag{11.30}$$
$$\xi^3_2 = \eta^3_2 = \eta_2, \tag{11.31}$$
$$\xi^3_3 = \eta^3_3 = \eta_3, \tag{11.32}$$
$$\xi^3_{12} = \eta^3_{12} = \eta_{12}, \tag{11.33}$$
$$\xi^3_{13} = \eta^3_{13} = \eta_{13}, \tag{11.34}$$
$$\xi^3_{23} = \eta^3_{23} = \eta_{23}, \tag{11.35}$$
$$\xi^3_{123} = \theta^3_{123} = \theta^{ec}_{123}. \tag{11.36}$$

This means that ξ^3 only cut the information between nodes 1 and 2 at the 3rd order, while ξ^{23} cut the information of both 2nd and 3rd orders. Generalizing this notation as $\xi^{k\cdots n}$ $(2 \leq k \leq n)$ being the mixture coordinates to cut edge information from the k-th to n-th orders, we have the following decomposition of KL divergence for general system size $n \geq 3$:

$$\begin{aligned} & D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{ec})] \\ = \; & D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{n})] \\ + \; & \sum_{k=2}^{n-1} D[p(\mathbf{x}, \xi^{k+1\cdots n}) : p(\mathbf{x}, \xi^{k\cdots n})]. \end{aligned} \tag{11.37}$$

In practice, $\xi^{2\cdots n}$ can be obtained from the products of remaining maximum cliques' η, following the analytical definition of edge cutting. On the other hand, we need some numerical solution to obtain $\xi^{k\cdots n}$ in case $(3 \leq k \leq n)$.

We next consider the positive/negative signs of interaction for each order of edge information. For simplicity, we take the $n = 3$ case. The 3rd order interaction in edge information $1-2$ is denoted as $D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^3)]$ in equation (11.22). We define the positive/negative sign of this interaction according to the following value s_3:

$$s_3 = sgn(\eta_{123} - \eta'_{123}), \tag{11.38}$$

where η'_{123} is the solution of $\xi^3_{123} = \theta^{ec}_{123}$ condition with respect to η^3_{123} variable. This definition of s_3 is based on the fact that if $\eta_{123} > \eta'_{123}$ holds, 3rd order cooccurrence of the three variables is more than the independence principle, therefore exists positive 3rd order interaction. While in case $\eta_{123} < \eta'_{123}$, the 3rd order interaction is negative. Using s_3, we express the 3rd order edge information of edge $1-2$ with positive/negative sign as

$$s_3 \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^3)]. \tag{11.39}$$

As for the positive/negative sign of the 2nd order edge information $D[p(\mathbf{x}, \xi^3) : p(\mathbf{x}, \xi^{23})]$ in equation (11.22), we consider in the same way the following value s_2:

$$s_2 = sgn(\eta_{12} - \eta'_{12}), \tag{11.40}$$

where η'_{12} is the solution of $\xi^{23}_{12} = \theta^{ec}_{12}$ condition with respect to η^{23}_{12} variable. The 2nd order edge information of edge $1-2$ with positive/negative sign is then given by

$$s_2 \cdot D[p(\mathbf{x}, \xi^3) : p(\mathbf{x}, \xi^{23})]. \tag{11.41}$$

We finally consider the positive/negative sign of the total edge information $D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{ec})]$. From the Pythagorean relation in equation (11.22), we can equally compare the amount of 2nd

and 3rd order interactions of edge information with additive law. The question is whether the edge information contain rather positive or negative interaction in total of all orders. To judge that, we define the following value s_{23} to judge the positive/negative sign of the total edge information:

$$s_{23} = sgn(s_3 D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^3)] + s_2 D[p(\mathbf{x}, \xi^3) : p(\mathbf{x}, \xi^{23})]). \tag{11.42}$$

This means that we compare the amount of positive and negative interactions in all orders, and take the sign of superior one. We then define the total edge information with positive/negative sign, namely the unified edge information, as

$$s_{23} D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{ec})], \tag{11.43}$$

or

$$s_{23} D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{23})]. \tag{11.44}$$

This is asymptotically compatible with χ^2-test of degree of freedom 2, but with distinction of positive/negative correlation:

$$\lambda = s_{23} \cdot 2N \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{ec})] \sim \chi^2(2). \tag{11.45}$$

Where N is the sample number. We call this λ as χ^2 value of edge information, or simply (the significance of) edge information.

In general form with system size n, the unified edge information is expressed as the followings:

$$s_{2\cdots n} D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{ec})], \tag{11.46}$$

or

$$s_{2\cdots n} D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{2\cdots n})], \tag{11.47}$$

where

$$\begin{aligned} s_{2\cdots n} &= sgn(s_n D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^n)] \\ &+ \sum_{k=2}^{n-1} s_k D[p(\mathbf{x}, \xi^{k+1\cdots n}) : p(\mathbf{x}, \xi^{k\cdots n})]), \end{aligned} \tag{11.48}$$

and the composition of ξ^{ec} and $\xi^{k\cdots n}$ $(2 \leq k \leq n)$ coordinates are chosen according to the edge selection $i - j$ $(1 \leq i, j \leq n)$.

11.3 Multi-Node Reference Reconstruction of Graph

11.3.1 Reduction of Model Parameters According to the Substantive Highest Order of Interaction

In this section, we consider the reconstruction of graph from cooccurrence data using the proposed unified edge information. The interest of graph reconstruction in complex network subjects usually postulate the large number of vertices n. Though, the dual-flat coordinates η θ and their mixture coordinates ξ increase their degrees of freedom in combinatory with $\sum_{k=1}^{n} {}_nC_k$. Even if we can afford to save the memory for these coordinates, the calculation of edge cutting requires the precise numerical solution of high-degree simultaneous polynomials with respect to $1 + \sum_{k=1}^{n-2} {}_{n-2}C_k$ η variables. This quantity apparently exceed the computing capacity as n increases. Some practical strategy to reduce the model size without losing essential information of data is needed.

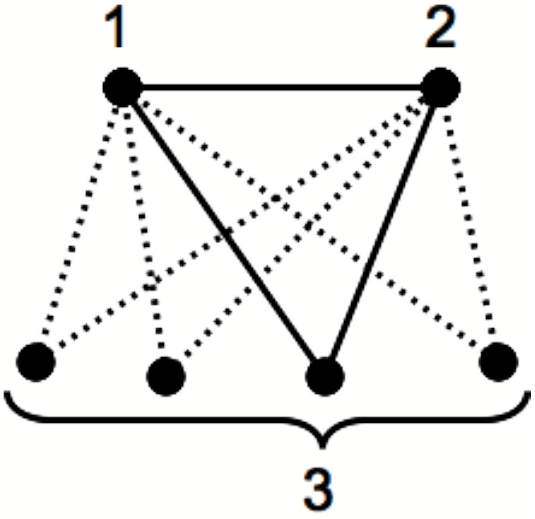


Figure 11.1: **Schematic example of model reduction by multi-node reference.** In case of system size $m = 6$ and substantive highest order $n = 3$ is depicted for edge cutting $1 - 2$. We choose the 3rd node of the model by calculating the unified edge information for each triplet including node 1 and 2, and choose the one which maximize the absolute value of the unified edge information λ in edge $1 - 2$.

Suppose the data are given as an m-dimensional joint distribution. Theoretically, there exist up to the m-th order of interaction between the variables. Though, if the significant interaction exists only up to the n-th order ($n < m$), it suffices to consider the ${}_mC_n$ combinations of n-node subgraphs for practical analysis. In case we can find such n according to the order-wise interaction analysis, we call this n as the substantive highest order of interaction.

Next, suppose we found such $n < m$. Then the unified edge information can be different for the same edge cutting $i - j$ ($1 \leq i, j \leq m$) according to the choice of n-node subgraph. In order to extract as much information as possible from the data, we should choose the subgraph for each edge cutting, such that the one which maximize the absolute value of the unified edge information. This reduction allows us to extract maximum edge information existing up to the highest substantive order. Simple example of this algorithm is depicted in Fig. 11.1.

11.3.2 Practical Problem and Solution

Due to the geometrical constraint of statistical manifold, there exist still some practical problems when calculating the unified edge information and order-wise interactions in actual data. We describe here the most popular one:

In case the cooccurrence data matrix is sparse (sparse network), there exists many zero-parameters which show singularity ($log(0)$ or $log(\infty)$) in KL divergence. This corresponds to the singular points in θ coordinates where the distribution can not find finite parameter values. To avoid these singularity without modifying much information, we implement minimum probability of occurrence for zero parameters of $P(\cdot, \cdots, \cdot)$ (*e.g.* if $P(\cdot, \cdots, \cdot) = 0$ then approximate with $P(\cdot, \cdots, \cdot) = 1.0e - 15$).

11.4 Data Set

We obtained weblog text data written in french from most-cited 10 political weblogs among popular 120 during the 181 days covering the french presidential in 2007 [81]. The occurrence of 190 political keywords were judged on a daily basis with binary value (occurrence:1 non-occurrence:0) for each of the 10 weblogs, and binary cooccurrence matrix between the 10 weblogs was obtained for each

of 181 days and each of 190 keywords.

By applying order-wise interaction and edge information analysis, one can judge in which statistical order the weblogs are correlated according to the occurrence of keywords, which reflect their semantic contents.

The url list of the most-cited 10 weblogs is shown in Tab. 11.1. 191 political keywords are listed in the followings:

11.4.1 Keyword List

35 heures ; action publique ; aide au développement ; Allemagne ; altermondialisme ; antisémitisme ; baisse des prélèvements ; baisses d'impôts ; banlieue ; Banque européenne ; Bayrou ; blogosphère ; blogueurs ; bouclier fiscal ; Bové ; bravitude ; budget de la recherche ; budget de l'Etat ; capitalisme financier ; carte scolaire ; chiffres du chômage ; Chirac ; chômage ; classes moyennes ; CO2 ; collectivités locales ; écolo ; écologie ; communautarisme ; comptes publics ; Conseil d'analyse économique ; contrat de travail ; contribuables ; criminalité ; croissance ; débat public ; dette publique ; déficit budgétaire ; DIABOLISER ; dialogue social ; discrimination positive ; démocratie participative ; démocratie sociale ; Don Quichotte ; dépense publique ; drapeau français ; droit au logement ; éducation ; développement durable ; effet de serre ; emploi ; encadrement militaire ; endettement ; enseignement supérieur ; entreprises ; EPR ; Eric Besson ; Etats membres ; Europe ; finances publiques ; financier ; fiscal ; fiscalité ; FN ; fonction publique ; fonctionnaires ; François Bayrou ; François Hollande ; gauche antilibérale ; gaz à effet de serre ; hausse des prix ; hausse des salaires ; hausse du smic ; heures supplémentaires ; Hollande ; identité française ; identité nationale ; impôt sur les successions ; insécurité ; internautes ; Internet ; intérêt général ; islam ; islamisme ; Jacques Chirac ; Jean-Marie Le Pen ; Jospin ; jurys citoyens ; justice sociale ; Kärcher ; Kyoto ; législatives ; Lionel Jospin ; logement ; logement opposable ; logements sociaux ; LOLF ; maîtrise des dépenses ; MoDem ; monde agricole ; mondialisation ; Nicolas Sarkozy ; Olivier Besancenot ; pacte écologique ; pacte présidentiel ; pauvreté ; petites retraites ; peuple ; PIB ; plein emploi ; pouvoir d'achat ; prélèvements obligatoires ; productivité ; protection sociale ; prévention de la délinquance ; PS ; ps udf ; réchauffement climatique ; recettes fiscales ; recherche ; réforme des retraites ; régimes de retraite ; régimes spéciaux ; réforme ; référendum sur la Constitution ; régularisation ; Royal ; rural ; ruralité ; salaire minimum ; salariés ; sans-abri ; Sarkozy ; Sécurité sociale ; service minimum ; service public ; Ségolène Royal ; socialiste ; solidarité ; sondages ; taux de chômage ; temps de travail ; territoire ; terrorisme ; Tony Blair ; traité constitutionnel ; travail ; TVA sociale ; UDF ; UMP ; Union européenne ; valeur travail ; Valérie Pécresse ; Verts ; vieillissement ; ville ; violences urbaines ; vote utile ; Xavier Bertrand ; zones rurales ; porte-avions ; homosexuel ; égalité des chances ; CSG ; droit de grève ; CNRS ; Gollnisch ; George Bush ; dialogue social ; contrat unique ; assurance maladie ; johnny ; technologies ; anti-Sarkozy ; troisième homme ; délocalisations ; précarité ; Frédéric Nihous ; Marseillaise ; Darfour ; chiffrage ; débats participatifs ; Villepinte ; Clearstream ; mai-68 ; outre-mer ; éléphants ; gare du Nord ; Iran ; Irak

11.5 Result and Discussion

11.5.1 Order-wise Interaction

We first quantified order-wise interactions in weblog data to detect the substantive highest-order interaction. We tested the model of up to $n = 4$ binary variables, which rarely gave 4th-order interaction. We therefore estimate that the substantive interaction exists in at most 3rd order, which justify to limit the model for multi-node reference reconstruction with $n = 3$.

The result of the 2nd and 3rd order interaction is depicted in Figs. 11.2 for the 1st(112th day) and 2nd(127th day) round of the presidential. The whole dynamics of the order-wise χ^2 values are shown in Figs. 11.3.

At the first sight, there exist dominant positive 2nd-order interaction during the whole campaign period, which shows the biggest peak during the 2nd round (113th-127th day). This dynamics is natural because there appear common words in all weblogs due to the discussion on related subjects.

Table 11.1: **Url list of 10 most-cited weblogs among popular 120 during the french presidential 2007.**

Url
http://embruns.net
http://hugues.blogs.com
http://maitre.eolas.free.fr
http://vanb.typepad.com/versac
http://dinersroom.free.fr
http://www.radical-chic.com
http://www.koztoujours.fr
http://www.authueil.org
http://birenbaum.blog.20minutes.fr
http://bruxelles.blogs.liberation.fr

As the chosen keywords consist of politically important terms, ones can not avoid simple 2nd-order correlation even in case they are writing opposite opinions, as long as they use the same keywords.

The 3rd-order interaction, on the other hand, shows relative superiority of negative interaction, which we assume to reflect the semantic diversification between weblogs. Since the chosen weblogs are the winners under the competitive criterion of citation frequency, there should exist distinctive color of opinion in each weblog to assure the semantic variation in weblog sphere. The absence of the 4th-order interaction also supports semantic independency between these weblogs, which gives the upper bound of substantive interaction. Related to the substantive order of interaction, an example of weblog citation network analysis during the american presidential in 2007 shows that topological distance effect has its peak in 4-clique [82]. Such topological information is also useful to limit the maximum subgraph scale for parameter reduction. The existence of such order limit may relate to the limit of our collective interaction scale in social network with respect to the human individual thought.

Judging the χ^2 values with the 5% threshold 3.841 of χ^2 test with degree of freedom 1, the appearance rates of the significant 2nd- and 3rd-order interactions are shown in Figs. 11.4.

The peak of 2nd- and 3rd-order χ^2 values around the 2nd round (127th day) are also expressed in their appearance rates of significance (We will later see that this is not the case for unified edge information). The negative 2nd-order interaction appears to have no significance in this data. The appearance rate of significant 3rd-order negative interaction, on the other hand, tend to relatively decrease during 80th-160th day, which include the 1st and 2nd round of presidential. These facts suggest that the semantic contents in each weblog tend to include broader range of defined keywords as the decisive rounds approach, which semantically means to discuss with more comprehensive perspective. Especially the increase of the positive 3rd-order significance and the decrease of the negative 3rd-order one is a strong indicator that the semantic variation between weblogs became more expressed by the context they use with the keywords, and not by the mere word-level presence/absence.

11.5.2 Unified Edge Information

Based on the substantive highest order of interaction, we next calculated the order-wise and unified edge information using multi-node reference algorithm with subsystem size $n = 3$. The first interests are what would be the degree and sign of interaction between weblogs, if only the information carried by an edge between a pair of vertices are considered, and if different orders are unified and expressed with a single measure. Example of reconstructed interaction graph with unified edge information are depicted in Figs. 11.5, which temporally correspond to the order-wise interaction graph in Figs. 11.2. The whole dynamics of order-wise and unified edge information are shown in Figs. 11.6.

The edge information appeared to be mostly positive in both order-wise and unified ones. This surely relate to the dominant positive 2nd-order interaction, but one can not judge the total positive/negative sign of interaction until we sum up the order-wise elements into unified edge

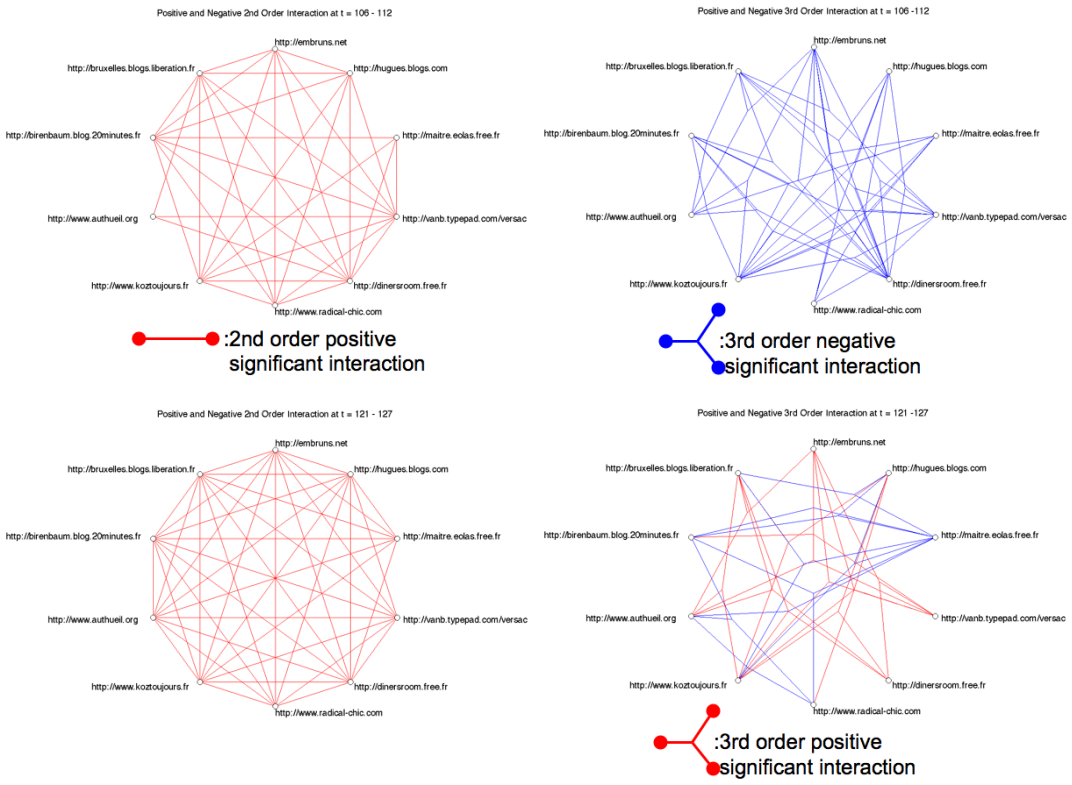


Figure 11.2: **Significant order-wise interaction between most-cited 10 weblogs during the preceding week of the 1st(112th day) and 2nd round (127th day) in french presidential 2007.** Top Left: Graph representation of positive(red) and negative(no exist) significant 2nd-order interaction during the week of the 1st round. Top Right: Graph representation of positive(red) and negative(no exist) significant 3rd-order interaction during the week of the 1st round. Bottom Left: Graph representation of positive(red) and negative(blue) significant 2nd-order interaction during the week of the 2nd round. Bottom Right: Graph representation of positive(red) and negative(blue) significant 3rd-order interaction during the week of the 2nd round.

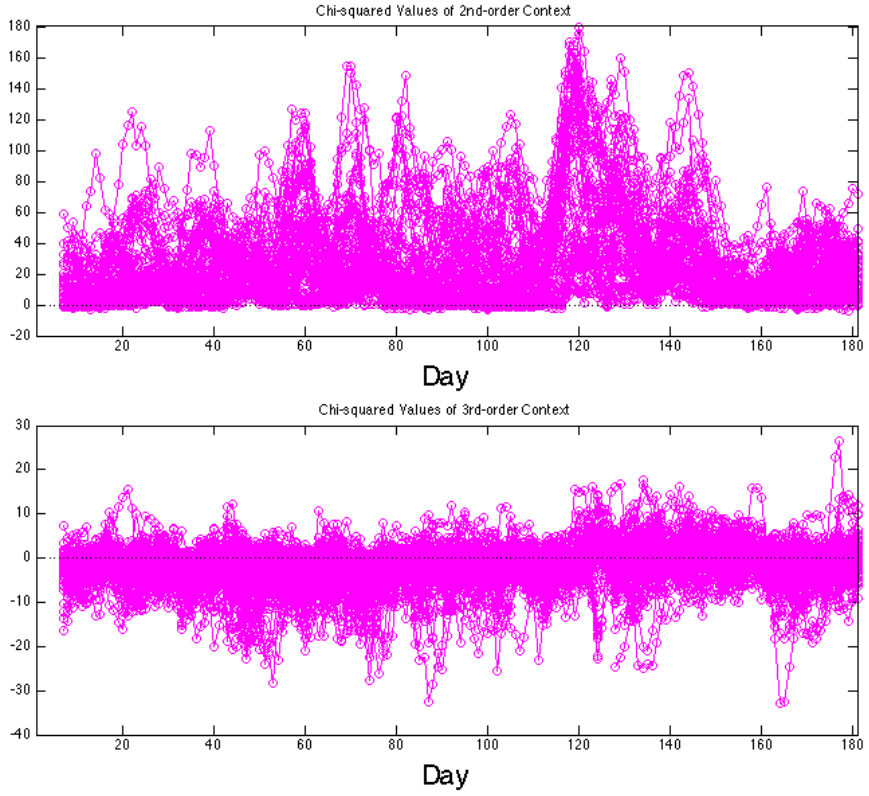


Figure 11.3: **Dynamics of order-wise interaction between most-cited 10 weblogs during french presidential 2007.** Top: Dynamics of χ^2 values with positive/negative signs of 2nd-order interaction. χ^2 values of all edges are superimposed. Bottom: Dynamics of χ^2 values with positive/negative signs of 3rd-order interaction. χ^2 values of all triplet edges are superimposed.

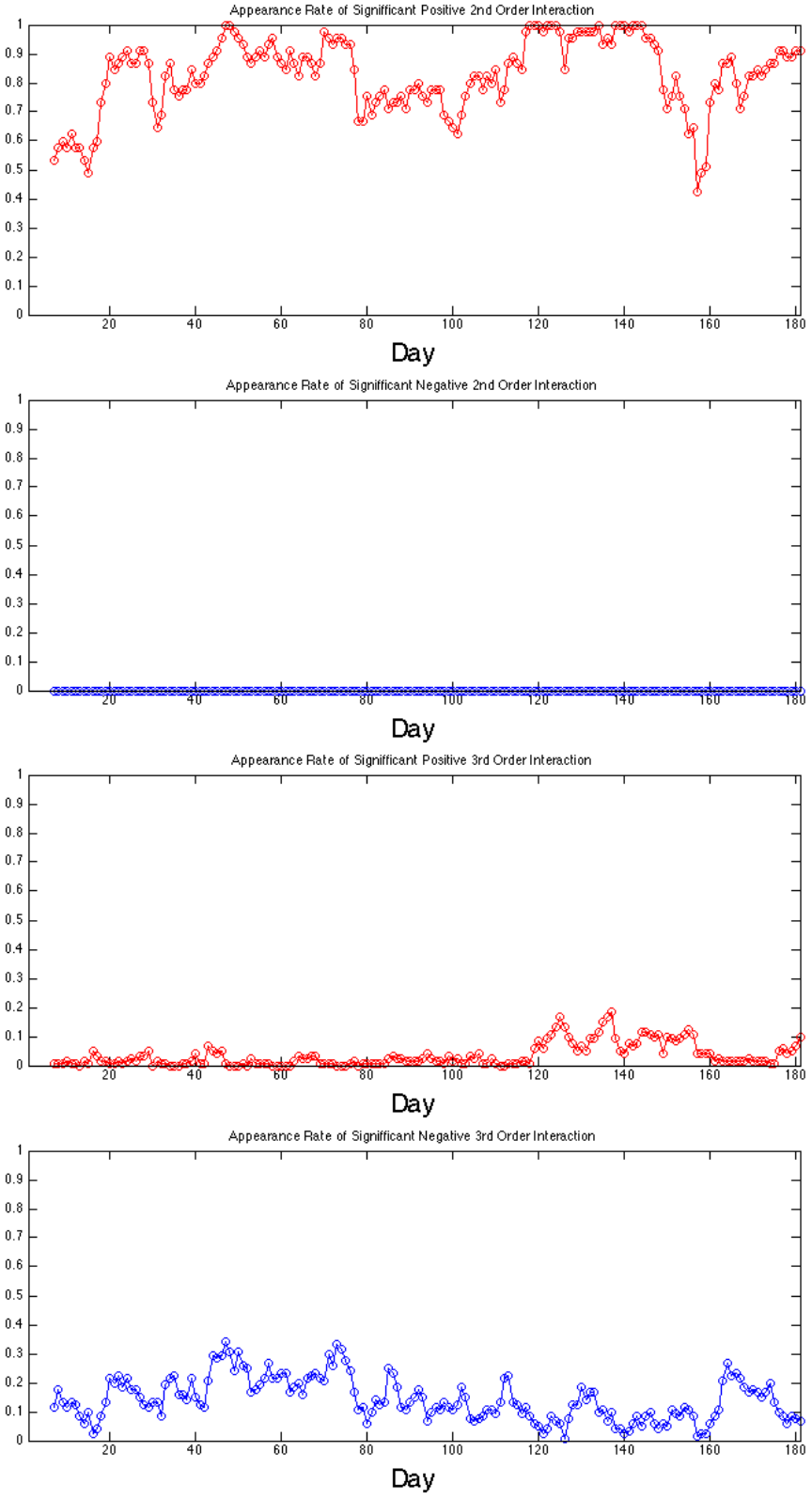


Figure 11.4: **Appearance rate of significant order-wise interactions between most-cited 10 weblogs during french presidential 2007.** Top Up: Appearance rate of significant positive 2nd-order interaction. Top Down: Appearance rate of significant negative 2nd-order interaction. Bottom Up: Appearance rate of significant positive 3rd-order interaction. Bottom Down: Appearance rate of significant negative 3rd-order interaction.

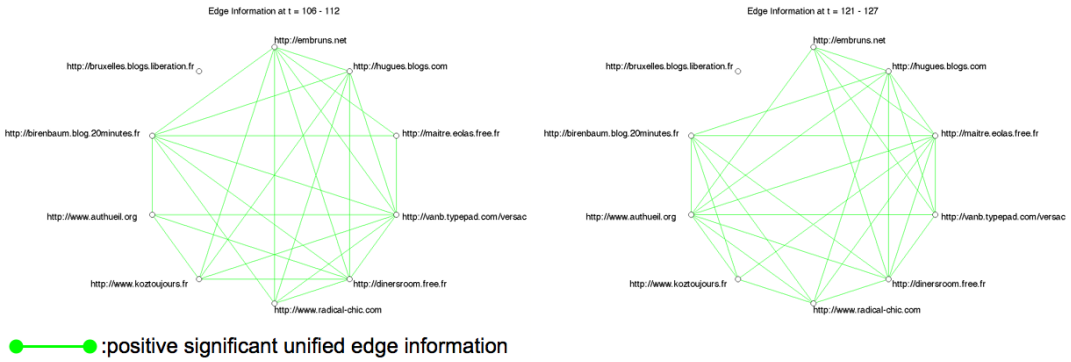


Figure 11.5: **Significant unified edge information between most-cited 10 weblogs during the preceding week of the 1st (112th day) and 2nd round (127th day) in french presidential 2007.** Left: Graph representation of positive(green) and negative(no exist) significant unified edge information during the week of the 1st round. Right: Graph representation of positive(green) and negative(no exist) significant unified edge information during the week of the 2nd round.

information. The higher-order edge information is also unpredictable from corresponding higher-order interaction, since for edge information we always consider the exact information carried by an edge, while latter consider for instance a triplet of edges for the 3rd order. The peak during 2nd round (113th-127th day) is more distinctive than the order-wise interaction.

Figs. 11.7 show the appearance rate of the significant positive and negative unified edge information with respect to the 5% threshold 5.991 of χ^2 test with degree of freedom 2. Compared to the augmentation of unified edge information dynamics during the 2nd round (113th-127th day) in Fig.11.6 (Bottom), the appearance rate of positive significance is rather modest and even decreasing in both end of the 2nd round. This fact implies that although there exist remarkable augmentation of unified edge information during the 2nd round, such edges are limited in relatively small number with respect to the total possible combinations between vertices. This means that some characteristic subgroup of weblogs emerged, which are highly correlated inside by themselves but relatively less with others. Semantically, this represents the relative localization of topics between different subgroups. This notion seems intuitively contradicting with the result of augmentation of significant 3rd-order interaction, but is actually additive and complementary. Unified information also include the effect of 3rd-order negative interaction, which shows rather constant dynamics when combined with positive one. Comprehensive arguments do occur as the positive 3rd-order interaction augment, on which there are additional diversification into further strongly correlated subcommunity with respect to the unified edge information. Such strong interaction are spatially localized, which are depicted in Fig.11.8 for the period of 113th-119th day. In this period, edge information shows strongly biased localization than the order-wise interactions.

We finally insist that the unified edge information is compatible to further clustering using generally edge weight, which largely concerns complex network subjects. Unified edge information contains more than the simple correlation, since it includes order-wise components up to the highest substantive order. The observed community structure in Fig. 11.8 (Bottom) may be clearly identified by further clustering using unified edge information. Further application to empirical range of binary occurrence data and the clustering of the derived results should be realized to further investigate the utility of this method.

11.6 Conclusion

We theoretically developed the way to decompose edge information into different orders with positive/negative signs of interaction, and defined an unified measure, namely unified edge information which is accessible for further clustering.

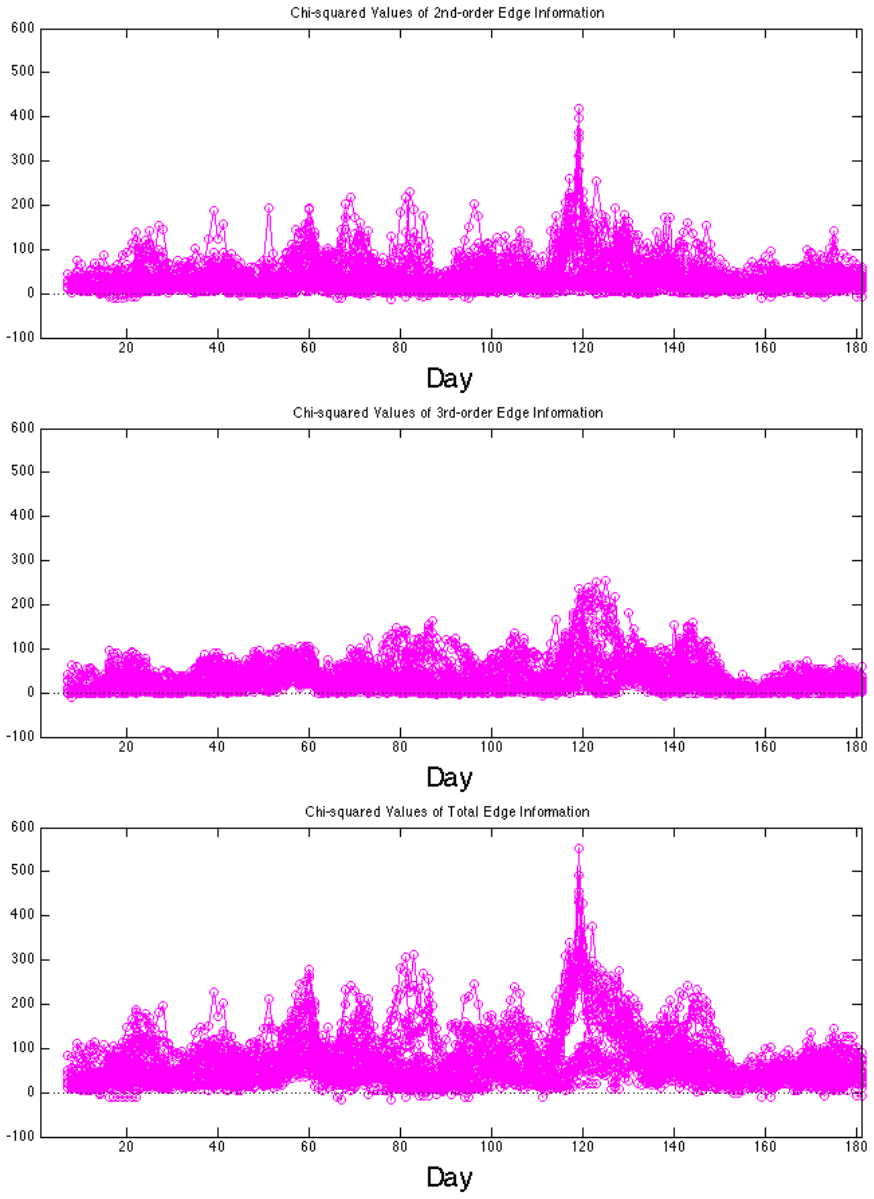


Figure 11.6: **Dynamics of order-wise and unified edge information between most-cited 10 weblogs during french presidential 2007.** Top: Dynamics of χ^2 values of 2nd-order edge information $2N \cdot s_3 \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^3)]$. Values of all edges are superimposed. Middle: Dynamics χ^2 values of 3rd-order edge information $2N \cdot s_2 \cdot D[p(\mathbf{x}, \xi^3) : p(\mathbf{x}, \xi^{23})]$. Values of all edges are superimposed. Bottom: Dynamics of χ^2 values of unified edge information $2N \cdot s_{23} \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{23})]$. Values of all edges are superimposed. Note that Pythagorean relation $2N \cdot s_{23} \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^{23})] = 2N \cdot s_3 \cdot D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi^3)] + 2N \cdot s_2 \cdot D[p(\mathbf{x}, \xi^3) : p(\mathbf{x}, \xi^{23})]$ holds.

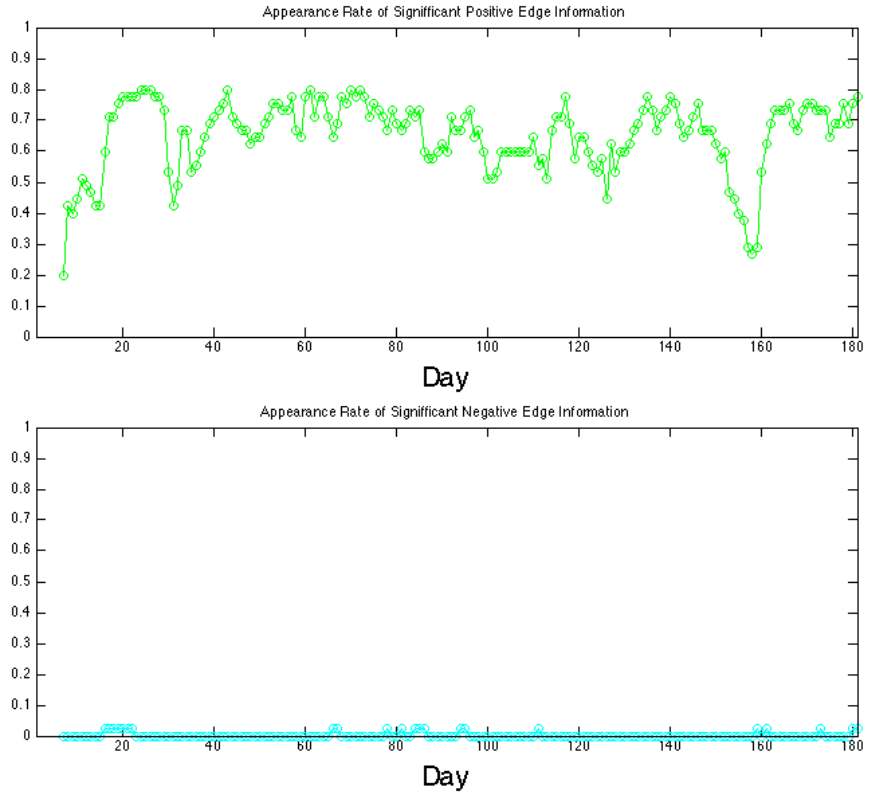


Figure 11.7: **Appearance rate of significant unified edge information between most-cited 10 weblogs during french presidential 2007.** Top: Appearance rate of significant positive unified edge information. Bottom: Appearance rate of significant negative unified edge information.

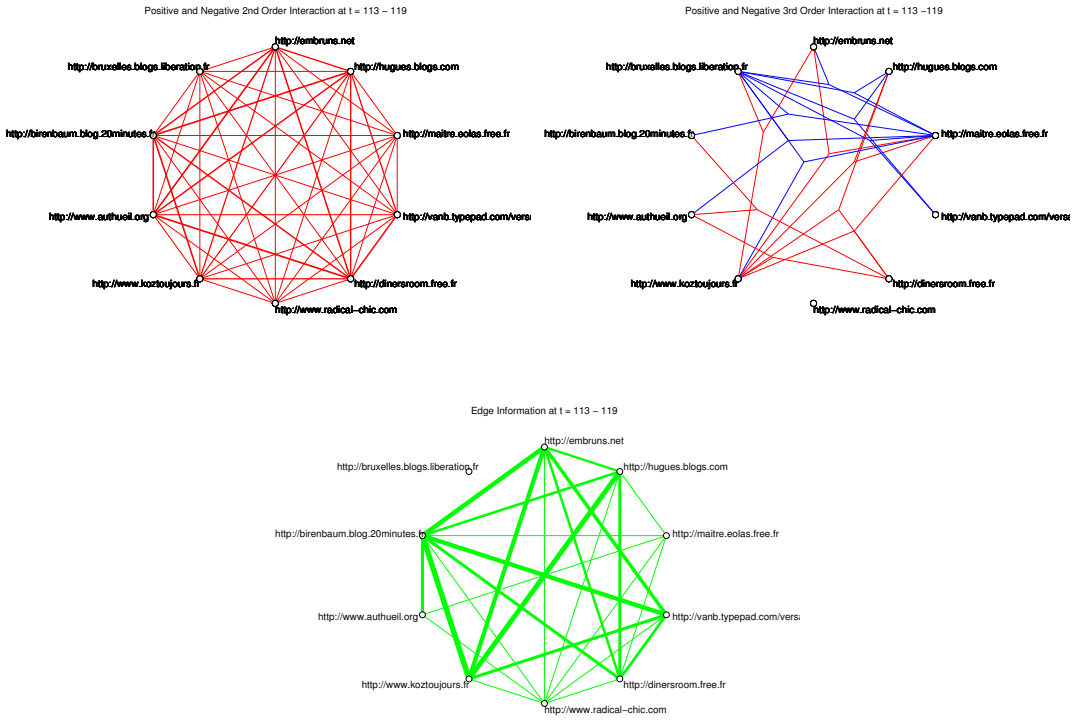


Figure 11.8: **Significant unified edge information and 2nd- and 3rd-order interaction between most-cited 10 weblogs during the 113th-119th day in french presidential 2007.** Top Left: Positive(red) and negative(no exist) significant 2nd-order interaction. Top Right: Positive(red) and negative(blue) significant 3rd-order interaction. Bottom: Positive(green) and negative(no exist) significant unified edge information. Unified edge information show strong localization. Edge width are proportional to the actual χ^2 values in these figures.

We also developed a practical framework to calculate a series of defined edge informations in large scale network by reducing parameters without losing substantive information.

Order-wise interaction analysis on political weblog data during french presidential 2007 proved the plausibility of this parameter reduction and revealed the order-wise structure related to semantic variation.

Analysis with unified edge information revealed the formation of strongly correlated community in the phase of comprehensive argument during the 2nd round of the presidential.

Chapter 12

Measuring Complexity with Respect to System Decompositionability

Abstract

Many complexity measures have been proposed from information theoretical point of view, all having the same feature based on arithmetic product of information. We consider the novel measures of complexity with respect to the system decompositionability, by introducing the geometric product of KL divergence. The novel complexity measures satisfy the boundary condition of vanishing at the limit of completely random and ordered state, and also with the existence of independent subsystem of any size. Analysis of social network data revealed the subsystem-wise dynamics with respect to the global behavior of the system during the augmentation of total complexity. The dynamics of subsystem-wise decompositionability was further examined with hypothetical view of geodesic flow with respect to the complexity measures as potential function on statistical manifold.

Keywords: Complexity measure, Boundary condition, System decompositionability, Geometric mean

Methodology: Definition of novel measures of complexity with respect to the system's decompositionability based on information geometry → Trend analysis of political weblog data as dynamical systems → Theoretical consideration on geodesic flow based on the novel measures of complexity with dynamical system perspective

12.1 Introduction: Network Characterization with the Use of Complexity Measure

There has been many attempts to define the complexity in stochastic model [83]. The complexity measure is usually the projection from system's variables to one-dimensional quantity, which is composed to express the degree of characteristic that we define to be important in what means *"complexity"*. Since the complexity measure is always a many-to-one association, it has both aspects of compressing information to classify the system from simple to complex, and losing resolution of the system's phase space. If the system has n variables, we generally need n independent complexity measures to completely characterize the system with real-value resolution. The problematics of defining a complexity measure is situated on the edge of balancing the information compression on system's complexity with theoretical support, and the resolution of the system identification to be maintained high enough to avoid trivial classification. The latter criterion increases its importance as the system size becomes larger, which we will see in later section. The better complexity measure is therefore a set of indices, with as less number as possible, which

characterizes major features related to the complexity of the system. In this sense, the ensemble of complexity measures is also analogous to the feature space of support vector machine.

Complexity measure on stochastic model has been mostly studied with information theoretical perspective. Computation-oriented definition of complexity takes deterministic formalization and measures the necessary information to reproduce a given symbolic sequence exactly, which is classified with the name of *algorithmic complexity*. [84] [85] [86]

On the other hand, statistical approach to complexity, namely *statistical complexity*, assumes some stochastic model as theoretical basis, and refers to the structure of information source on it in measure-theoretic way [87] [88] [89]. One of the most classical statistical complexity is the mutual information between two stochastic variables, and its generalized form measuring dependence between n variables are proposed by several authors [80] [38] [31]. We should also recall that complexity is not necessary conditioned only by information theory, but rather motivated from the organization of living system such as brain activity. The TSE complexity shows further extension of generalized mutual information into biological context, where complexity exists as the heterogeneity between different system hierarchy [37]. These statistical complexities are all based on the boundary condition of vanishing at the limit of completely random and ordered state [90].

In this chapter, we particularly focus on the generalized mutual information as a start point of the argument, and further consider to incorporate network heterogeneity with respect to the system's decompositionability, in a way still missing in TSE complexity.

12.2 Generalized Mutual Information as Complexity with Respect to the Total System Decomposition

In previous chapter, we have introduced a measure of complexity in terms of system decomposition, by measuring the KL-divergence between a given system and its independently decomposed subsystems. We consider here the total system decomposition, and measure the informational distance I between the system and the totally decomposed system where each element are independent.

$$I = \sum_{i=1}^{n} H(x_i) - H(x_1, \cdots, x_n), \tag{12.1}$$

This quantity is the generalization of mutual information, and is named in various ways such as generalized mutual information, integration, complexity, multi-information, etc., according to different authors. We call the I as "*multi-information* taking after [38]. This quantity can be interpreted as a measure of complexity that sums up the order-wise interactions existing in each subset of components with information geometrical formalization [80]

For simplicity, we denote the multi-information I of n-dimensional stochastic binary variables as follows, using the notation of the system decomposition:

$$I = D[< 111 \cdots 1 >:< 123 \cdots n >], \tag{12.2}$$

where $D[\cdot : \cdot]$ is the KL-divergence from the first system to the second one.

12.3 Rectangle-Bias Complexity

The multi-information contains some degrees of freedom in case $n > 2$. That is, we can define a set of distributions $\{p(\mathbf{x})|I = const.\}$ with different parameters but the same I value. This fact can be clearly explained with the use of information geometry. From the Pythagorean relation, we obtain the followings in case of $n = 3$:

$$D[< 111 >:< 113 >] + D[< 113 >:< 123 >] = D[< 111 >:< 123 >], \tag{12.3}$$
$$D[< 111 >:< 121 >] + D[< 121 >:< 123 >] = D[< 111 >:< 123 >], \tag{12.4}$$
$$D[< 111 >:< 122 >] + D[< 122 >:< 123 >] = D[< 111 >:< 123 >]. \tag{12.5}$$

Using these relations, we can schematically represent the decomposed systems on a circle diagram with diameter $\sqrt{I}$. This representation is based on the analogous algebra between Pythagorean relation of KL divergence, and that of Euclidian geometry where the circumferential angle of a semi-circular arc is always $\frac{\pi}{2}$.

Fig.12.1 represents two different cases with the same I value in case $n = 3$. The distance between two systems in the same diagram corresponds to the square root value of KL-divergence between them. Clearly the left and right figures represents different interaction between nodes, although they both have the same I value. Such geometrical variation is possible by the abundance of degree of freedom in dual coordinates compared to the given constraint. There exist 7 degrees of freedom in η or θ coordinates for $n = 3$, while the only constraint is the invariance of I value, which only reduce 1 degree of freedom. The remaining 6 degrees of freedom can then be deployed to produce geometrical variation in the circle diagram. As for considering system decomposition, the left figure is difficult to obtain decomposed systems without losing much information. While in the right figure there exists relatively easy decomposition $< 122 >$, which loses less information than any decompositions in the left figure. We call such degree of facility of decomposition with respect to the losing information as *system decompositionability*. In this sense, the left system is more complex although the 2 systems both have the same I value. Especially, in case $D[< 111 >:< 122 >] = D[< 111 >:< 113 >] = D[< 111 >:< 121 >]$, the system does not have any easiest way of decomposition, and any isolation of a node loses significant amount of information.

To further incorporate such geometrical structure reflecting system decompositionability into a measure of complexity, we consider a mathematical way to distinguish between these two figures. Although the total sum of KL divergence along the sequence of system decomposition is always identical to I by Pythagorean relation, their product can vary according to the geometrical composition in the circle diagram. This is analogous to the isoperimetric inequality of rectangle, where regular tetragon gives the maximum dimensions amongst constant perimeter rectangles.

We propose provisionary a new measure of complexity as follows, namely *rectangle-bias complexity* C_r:

$$C_r = \frac{1}{|SD| - 2} \sum_{<\cdots>\in SD} D[< 11\cdots1 >:< \cdots >] \cdot D[< \cdots >:< 12\cdots n >], \tag{12.6}$$

where SD is the set of possible system decomposition in n binary variables, and $|SD|$ is the element number of SD. For example, $SD = \{< 111 >, < 122 >, < 121 >, < 113 >, < 123 >\}$ and $|SD| = 5$ for $n = 3$. This measure distinguishes between the two systems in Fig.12.1, and gives larger value for the left figure. It also gives maximum value in case $D[< 111 >:< 122 >] = D[< 111 >:< 113 >] = D[< 111 >:< 121 >]$.

12.4 Complementarity between Complexities Defined with Arithmetic and Geometric Means

We evaluate the possibility and the limit of rectangle-bias complexity C_r comparing with other proposed measures of complexity.

Interests on measuring network heterogeneity has been developed toward the incorporation of multi-scale characteristics into complexity measures. The TSE complexity is motivated from the structure of the functional differentiation of brain activity, which measures the difference of neural integration between all sizes of subsystems and the whole system [37]. Biologically motivated TSE complexity is also investigated from theoretical point of view, to further attribute desirable property as an universal complexity measure independent of system size [91]. The hierarchical structure of the exponential family in information geometry also leads to the order-wise description of interaction, which can be regarded as a multi-scale complexity measure [80]. The relation between the order-wise interaction and the TSE complexity is theoretically investigated to establish the order-wise component correspondence between them [38].

These indices of network heterogeneity, however, all depend on the arithmetic mean of the component-wise information theoretical measure. We show that these arithmetic means still miss to measure certain modularity based on the statistical independence between subsystems.

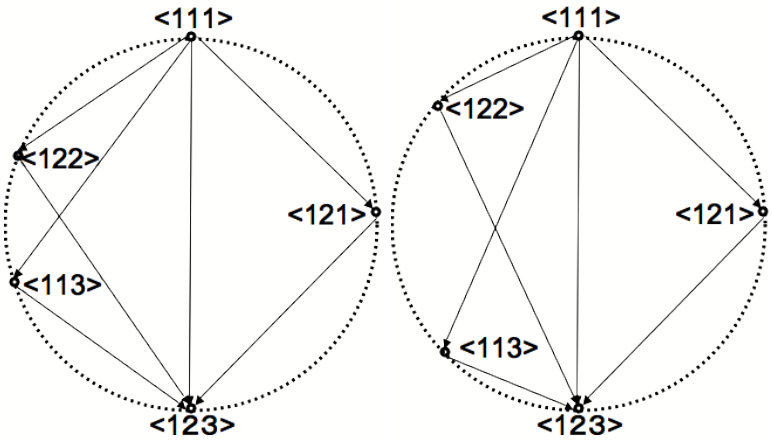


Figure 12.1: **Circle diagrams of system decomposition in 3-node network.** Both systems have the same value of multi-information I that is expressed as the identical diameter length of the circles. 2 variations are shown, where the left system is more complex (C_r high) in a sense any system decomposition requires to lose more information than the easiest one ($< 122 >$) in the right figure (C_r low).

Figs. 12.2 present the simplified cases where complexity measures with arithmetic means fail to distinguish. We consider the two systems with different heterogeneity but identical multi-information I. Here, the multi-information can not reflect the network heterogeneity. The TSE complexity and its information geometrical correspondence in **??** has a sensitivity to measure the network heterogeneity, but since the arithmetic mean is taken over all subsystems, they do not distinguish the component-wise break of symmetry between different scales. The renormalized TSE complexity with respect to the multi-information I still has the same insensitivity. Even by incorporating the information of each subsystem scale, the arithmetic mean can balance out between the scale-wise variations, and a large range of the heterogeneity in different scale can realize the same value of these complexities. For the application in neuroscience, the assumption of a model with simple parametric heterogeneity and the comparison of TSE complexity between different I values alleviate this limitation [37].

In contrast to complexities with arithmetic mean, the rectangle-bias complexity C_r is related to the geometrical mean. The C_r can distinguish the two systems in Figs. 12.2, giving relatively high C_r value to the left system and low value to the right one.

This does not mean , however, that the C_r has a finer resolution than other complexity measures. The constant conditions of complexity measures are the constraint on $\sum_{k=1}^{n} {}_nC_k$ degrees of freedom in model parameter space, which define different geometrical composition of corresponding submanifolds. We basically need $\sum_{k=1}^{n} {}_nC_k$ independent measures to assure the real-value resolution of network feature characterization. Complexities with arithmetic and geometric means are just giving complementary information on network heterogeneity, or different constant-complexity submanifolds structure in statistical manifold as depicted in Fig. 12.3. Therefore, it is also possible to construct a class of systems that has identical I and C_r values but different TSE complexity. Complexity measures should be utilized in combination, with respect to the non-linear separation capacity of network features of interest.

12.5 Cuboid-Bias Complexity with Respect to System Decompositionability

We consider the expansion of C_r into general system size n. The $n \leq 4$ situation is different from $n = 3$ and less in the existence of a hierarchical structure between system decompositions.

Fig. 12.4 shows the hierarchy of the system decompositions in case $n = 4$. Such hierarchical structure between system decompositions are not homogeneous with respect to the subsystems

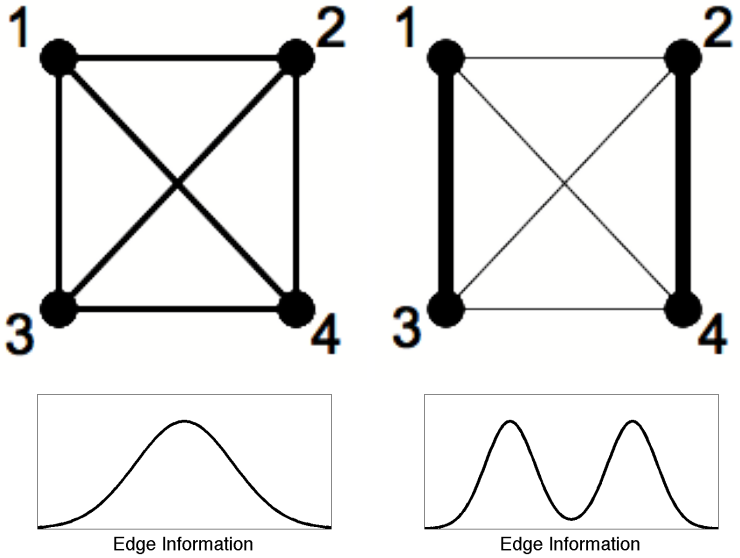


Figure 12.2: **Schematic examples of stochastic systems with identical multi-information I where complexity measures with arithmetic mean fail to distinguish.** Top Left: Example 1 of stochastic system with homogeneous mean of edge information and symmetric fluctuation of its heterogeneity. Top Right: Example 2 of heterogeneous stochastic system with bimodal edge information distribution and identical multi-information I and complexity based on arithmetic mean as example 1. Bottom Left: schematic representation of the distribution of interaction (edge information) in upper network. Bottom Right: schematic representation of the distribution of interaction (edge information) in upper network.

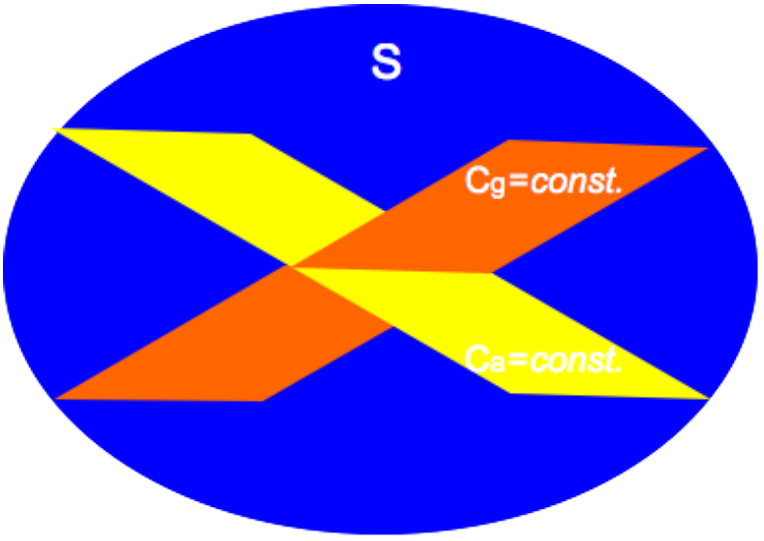


Figure 12.3: **Schematic representation of complementarity between complexity measures based on arithmetic mean (C_a) and geometric mean (C_g) of informational distance.** An example of the $n-1$ dimensional constant-complexity submanifolds with respect to $C_a = const.$ and $C_g = const.$ conditions are depicted with yellow and orange surface, respectively. The dimension of the whole statistical manifold S is the parameter number n.

numbers, and depends on the isomorphic types of decomposed systems. This fact produces certain difference of meaning in complexity between each KL divergences when considering the system decompositionability.

A simple example in 4 nodes network is shown in Fig. 12.5. We consider the modification of 2 KL divergences in the figure, $D[< 1111 >:< 1222 >]$ and $D[< 1111 >:< 1133 >]$ from the diameter of green dotted circle to the dashed one.

The joint distribution $P(x_1, x_2, x_3, x_4)$ of a discrete distribution with 4 binary variables (x_1, x_2, x_3, x_4) $(x_1, x_2, x_3, x_4 \in \{0, 1\})$ have $2^4 - 1 = 15$ parameters, which define the dual-flat coordinates of statistical manifold in information geometry.

On the other hand, the possible system decompositions exist as the followings in $n = 4$:

$$\begin{aligned} SD \quad := \quad & \{< 1111 >, < 1114 >, < 1131 >, < 1211 >, < 1222 >, \\ & < 1133 >, < 1212 >, < 1221 >, < 1134 >, < 1214 >, \\ & < 1231 >, < 1224 >, < 1232 >, < 1233 >, < 1234 >\}. \end{aligned} \tag{12.7}$$

Since the number of possible system decompositions is 15, and each is associated with the modification of different sets of $P(x_1, x_2, x_3, x_4)$ parameters, the system decompositions and KL divergences between them can be defined independently. This also holds even under the constant condition of I value or other complexity measures except the ones imposing dependency between system decompositions.

This means that we can independently modify the diameter of green dotted circle in Fig. 12.5, without changing the diameters of the red and blue circles, which define the system decompositions $< 1233 >$ and $< 1134 >$ in the sub-hierarchy of $< 1222 >$ and $< 1133 >$, respectively. Other KL divergences can also be maintained as given constant values for the same reason.

The rectangle-biased complexity C_r increases its value with such modification, but does not reflect the heterogeneity of KL divergences according to the hierarchy of system decompositions. If we consider the system decompositionability as the mean facility to decompose the given system into its finest components with respect to the "all" possible system decompositions, such hierarchical difference also has a meaning in the definition of complexity.

The effect of modifying the diameter of the green dotted circle is different between the decomposition sequences $< 1111 >\rightarrow< 1222 >\rightarrow< 1233 >\rightarrow< 1234 >$ and $< 1111 >\rightarrow< 1133 >\rightarrow< 1134 >\rightarrow< 1234 >$. The decrease of the KL divergence $D[< 1222 >:< 1233 >]$ is less than $D[< 1133 >:< 1134 >]$ since the diameter of the red dotted circle is larger than the blue one in Fig. 12.5. This means that the effect of changing the same amount of KL divergences in $D[< 1111 >:< 1222 >]$ and $D[< 1111 >:< 1133 >]$ produces larger effect on the sequence $< 1111 >\rightarrow< 1133 >\rightarrow< 1134 >\rightarrow< 1234 >$ than $< 1111 >\rightarrow< 1222 >\rightarrow< 1233 >\rightarrow< 1234 >$, when compared at the sequence level. The rectangle-biased complexity C_r does not reflect such characteristics since it does not distinguish between the hierarchical structure between the diameters of the green, red and blue dotted circles.

To incorporate such hierarchical effect in a complexity measure with geometric mean, we have the natural expansion of the rectangle-biased complexity C_r as the *cuboid-bias complexity* C_c, which is defined as follows:

$$C_c = \frac{1}{|Seq|} \sum_{i_s=1}^{|Seq|} \prod_{i=1}^{n-1} D[SD_i(i_s) : SD_{i+1}(i_s)], \tag{12.8}$$

Where Seq represents the possible sequences of hierarchical system decompositions as follows:

$$Seq = \{SD_1(i_s) \rightarrow SD_2(i_s) \rightarrow \cdots SD_i(i_s) \cdots \rightarrow SD_n(i_s) | 1 \leq i_s \leq |Seq|\}. \tag{12.9}$$

The elements $SD_i(i_s)$ of Seq corresponds to the system decomposition, which is aligned according to the hierarchy with the following algorithmic procedure:

1. Initialization: Set the initial sets of system decomposition of all sequences in Seq as the whole system $SD_1(i_s) :=< 111\cdots1 > (1 \leq i_s \leq |Seq|)$.

2. step $i \to i+1$: If the system decomposition is the total system decomposition ($SD_i(i_s) :=< 123\cdots n >$), then stop. Otherwise, choose a non-decomposed subsystem $SS_i(i_s)$ of the system decomposition $SD_i(i_s)$ different for each i_s, and further divide it into two independent subsystems $SS_i^1(i_s)$ and $SS_i^2(i_s)$. $SD_{i+1}(i_s)$ is then defined as the independent union of the $SS_i^1(i_s)$, $SS_i^2(i_s)$, and the $SD_i(i_s)$ without $SS_i(i_s)$.

3. Go to the next step $i+1 \to i+2$.

The value of $|Seq|$ corresponds to the number of different sequences generated by this algorithm. For example, $|Seq| = 3$ and $|Seq| = 18$ holds for $n = 3$ and $n = 4$, respectively.

The products of KL divergences according to the hierarchical sequences of system decompositions in equation (12.8) is related to the volume of $n-1$-dimensional cuboids in the circle diagram. An example in case of $n = 4$ is presented in Fig. 12.5, where two cuboids with 3 orthogonal edges of the different decomposition sequences $< 1111 >\to< 1222 >\to< 1233 >\to< 1234 >$ and $< 1111 >\to< 1133 >\to< 1134 >\to< 1234 >$ are depicted, whose cuboid volumes are

$$\sqrt{D[< 1111 >:< 1222 >]D[< 1222 >:< 1233 >]D[< 1233 >:< 1234 >]}, \tag{12.10}$$

and

$$\sqrt{D[< 1111 >:< 1133 >]D[< 1133 >:< 1134 >]D[< 1134 >:< 1234 >]}, \tag{12.11}$$

respectively.

In the same way as C_r, we took in the definition of C_c the arithmetic average of cuboid volumes so that to renormalize the combinatorial increase of the decomposition paths ($|Seq|$) according to the system size n.

Note that on the other hand we did not renormalize the rectangle-bias complexity C_r and the cuboid-bias complexity C_c by taking the exact geometrical mean of each product of KL divergences like $\sqrt[n-1]{\prod_{i=1}^{n-1} D[SD_i(i_s) : SD_{i+1}(i_s)]}$. This is for further accessibility to theoretical analysis such as variational method (see "Further Consideration" section), and does not change qualitative behavior of C_r and C_c since the power root is a monotonically increasing function. This treatment can be interpreted as taking the $(n-1)$-th power of the geometric means for the hierarchical sequences of KL divergences.

A more comprehensive example on the utility of the cuboid-bias complexity C_c with respect to the rectangle-biased one C_r is shown in Fig. 12.6. We consider the 6 nodes networks ($n = 6$) with the same I and C_r values but different heterogeneity. The system in the top left figure has a circularly connected structure with medium intensity, while that of the top right figure has strongly connected 3 subsystems. These systems have qualitatively five different ways of system decomposition that are the basic generators of all hierarchical sequences $Seq = \{SD_1(i_s) \to \cdots \to SD_5(i_s) | 1 \leq i_s \leq |Seq|\}$ for these networks. The five basial system decompositions are shown with the number ①, ②, ②′, ③ and ④ in top figures.

The circle diagrams of these systems are depicted in the middle figures. To suppose the same constant value of C_r in both systems, the following condition is satisfied in the middle right figure: $D[< 111111 >: ②] < D[< 111111 >: ①$in Middle Left figure$] < D[< 111111 >: ①] < D[< 111111 >: ②$in Middle Left figure$] < D[< 111111 >: ③] < D[< 111111 >: ④]$. Furthermore, the total surface of right triangles sharing the circle diameter as hypotenuse in the middle left and the middle right figures are conditioned to be identical, therefore the rectangle-bias complexity C_r fails to distinguish.

On the other hand, under the same condition, the cuboid-bias complexity C_c distinguishes between these two systems and gives higher value to the left one. The volume of 5-dimensional cuboids of the decomposition sequence $< 111111 > \xrightarrow{①②②'③④} < 123456 >$ are schematically shown in the bottom figures, maintaining the qualitative difference between KL divergences.

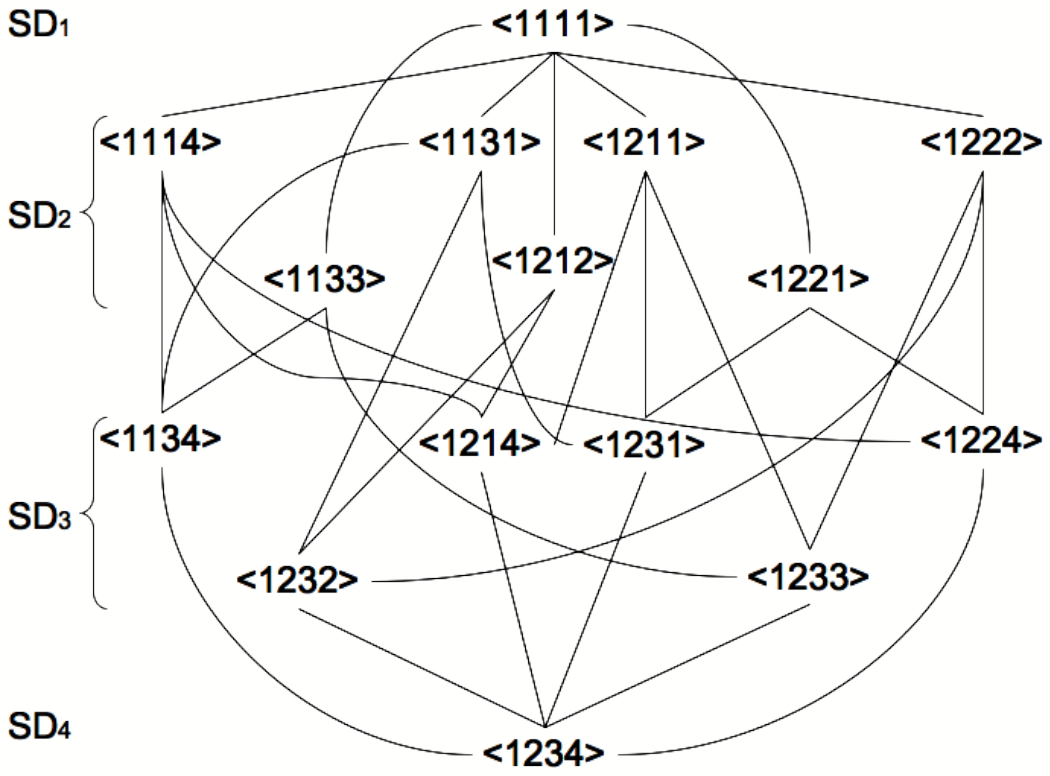


Figure 12.4: **Hierarchy of system decomposition for 4 nodes network ($n = 4$).** Possible sequences of $Seq = \{SD_1(i_s) \rightarrow SD_2(i_s) \rightarrow SD_3(i_s) \rightarrow SD_4(i_s) | 1 \leq i_s \leq |Seq| = 18\}$ are connected with the lines.

Since the multi-information I is identical between the two systems, so is the values of KL divergence $D[< 111111 >:< 123456 >]$, which is the sum of the KL divergences along the sequence $< 111111 > \xrightarrow{①②②'③④} < 123456 >$ from the Pythagorean theorem. This means that the inequality between the cuboid volumes can be represented as the isoperimetric inequality of high-dimensional cuboid. As a consequence, the left system has qualitatively higher value of C_c than the right one. The cuboid-bias complexity C_c is also sensitive to such heterogeneity.

12.6 Regularized Cuboid-Bias Complexity with Respect to Generalized Mutual Information

We further consider the geometrical composition of system decompositions in the circle diagram and insist the necessity of renormalizing the cuboid-bias complexity C_c with the multi-information I, which gives another measure of complexity namely "*regularized cuboid-bias complexity* C_c^R."

We consider the situation in actual data where the multi-information I varies. Fig. 12.7 shows the $n = 3$ cases where the C_c fails to distinguish. Both the blue and red systems are supposed to have the same C_c value by adjusting the red system to have relatively smaller values of KL divergences $D[< 111 >:< 122 >]$ and $D[< 113 >:< 123 >]$ than the blue one. Such conditioning is possible since the KL divergences are independent parameters with each other.

Although the C_c value is identical, the two systems have different geometrical composition of system decompositions in the circle diagram. The red system has relatively easier way of decomposition $< 111 > \rightarrow < 122 >$ if renormalized with the total system decomposition $< 111 > \rightarrow < 123 >$. This relative decompositionability with respect to the renormalization with the multi-information I can be clearly understood by superimposing the circle diagram of the two systems and comparing the angles between each and total decomposition paths (bottom figure). The red system

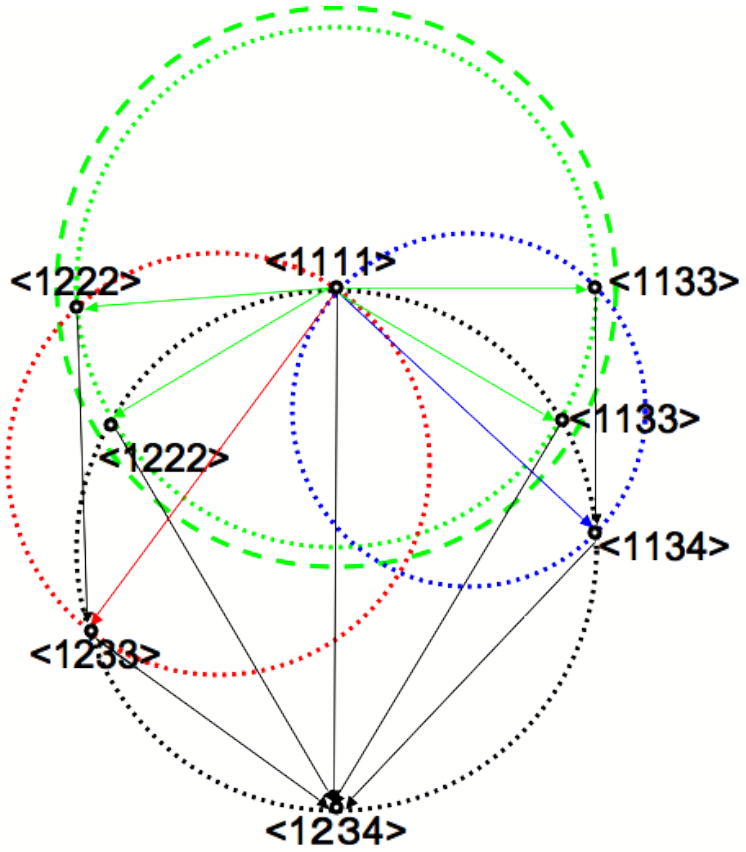


Figure 12.5: **Hierarchical effect of sequential system decomposition on cuboid volume and rectangle surface.** We consider to increase the diameter of the green circle from dotted to dashed one without changing those of the red and blue circles, which gives different effect on the change of $D[< 1222 >:< 1233 >]$ and $D[< 1133 >:< 1134 >]$ according to the hierarchical structure of the decomposition sequences.

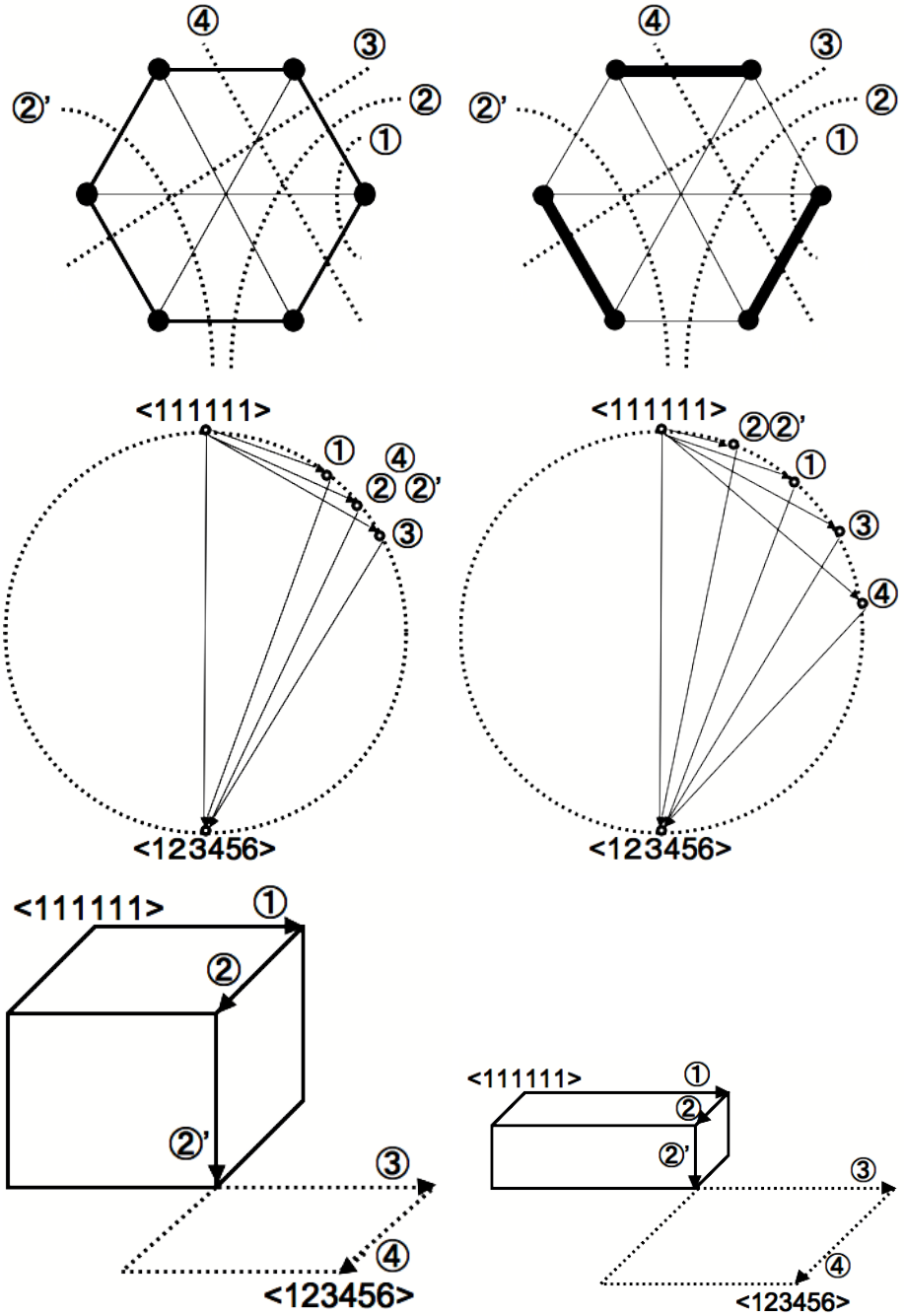


Figure 12.6: **Meaning of taking geometric mean over the sequence of system decomposition in cuboid-bias complexity C_c.** Top Left: Example of 6-node network with circularly connected structure with medium intensity. Edge width is proportional to edge information. Top Right: Example of 6-node network with strongly connected 3 subsystems. Edge width is proportional to edge information. The multi-information I of the two systems in Top figures are conditioned to be identical. The dotted line schematically represent possible system decompositions. Middle Left and Right: Circle diagrams of each system decomposition in upper networks. The total surface of right triangles sharing the circle diameter as hypotenuse in Middle Left and Middle Right figures are conditioned to be identical, therefore the rectangle-bias complexity C_r fails to distinguish. Bottom Left and Right: 5-dimensional cuboids of upper networks(Top figures) whose edges are the root of KL divergences for the strain of system decomposition $< 111111 > \xrightarrow{①②②'③④} < 123456 >$. Only the first 3-dimensional part is shown with solid line, and the remaining 2-dimensional part is represented with dotted line. The volume of Bottom Left cuboid is larger than that of Bottom Right one, according to the isoperimetric inequality of high-dimensional cuboid. The total squared length of each side is identical between two cuboids, which represents multi-information $I = D[< 111111 >:< 123456 >]$.

has larger angle between the decomposition paths $< 111 > \rightarrow < 122 >$ and $< 111 > \rightarrow < 123 >$ than any others in the blue system, which represents the relative facility of the decomposition under renormalization with I. In this term, the paths $< 111 > \rightarrow < 121 >$ in the red and blue system does not change its relative facility, and the paths $< 111 > \rightarrow < 113 >$ are easier in the blue system.

To express the system decompositionability based on these geometrical compositions in a comprehensive manner, we define the *regularized cuboid-bias complexity* C_c^R as follows:

$$\begin{aligned} C_c^R &= \frac{1}{|Seq|}\sum_{i_s=1}^{|Seq|}\prod_{i=1}^{n-1}\frac{D[SD_i(i_s):SD_{i+1}(i_s)]}{D[< 11\cdots 1 >:< 12\cdots n >]} \\ &= \frac{C_c}{D[< 11\cdots 1 >:< 12\cdots n >]^{n-1}} \\ &= \frac{C_c}{I^{n-1}}. \end{aligned} \tag{12.12}$$

The red system then has qualitatively smaller C_c^R value than the blue system in Fig. 12.7.

12.7 Modular Complexity with Respect to the Easiest System Decomposition Path

We have considered so far the system decompositionability with respect to the all possible decomposition sequences. This was also a way to avoid the local fluctuation of the network heterogeneity to be reflected in some specific decomposition paths. On the other hand, the easiest decomposition is particularly important when considering the modularity of the system. If there exists hierarchical structure of modularity in different scales with different coherence of the system, the KL divergence and the sequence of the easiest decomposition gives much information.

Fig. 12.8 schematically shows a typical example where there exist two levels of modularity. Such structure with different scales of statistical coherence appears as functional segregation in neural systems [37], and is expected to be observed widely in complex systems. The hierarchical topology of the easiest decomposition path reflect these structures. For example, in the system of Fig. 12.8, the decompositions between $< 1\ 1\ \cdots\ 1 >$ and $< 1\ 1\ 1\ 1\ 5\ 5\ 5\ 5\ 9\ 9\ 9\ 9\ 13\ 13\ 13\ 13 >$ are easier than those inside of the 4-node subsystems. The value of KL divergences also reflect the hierarchy, giving relatively low values for the decomposition between the 4-node subsystems, and high values inside of them. By examining the shortest decomposition path and associated KL divergences in possible Seq, one can project the hierarchical structure of the modularity existing in the system.

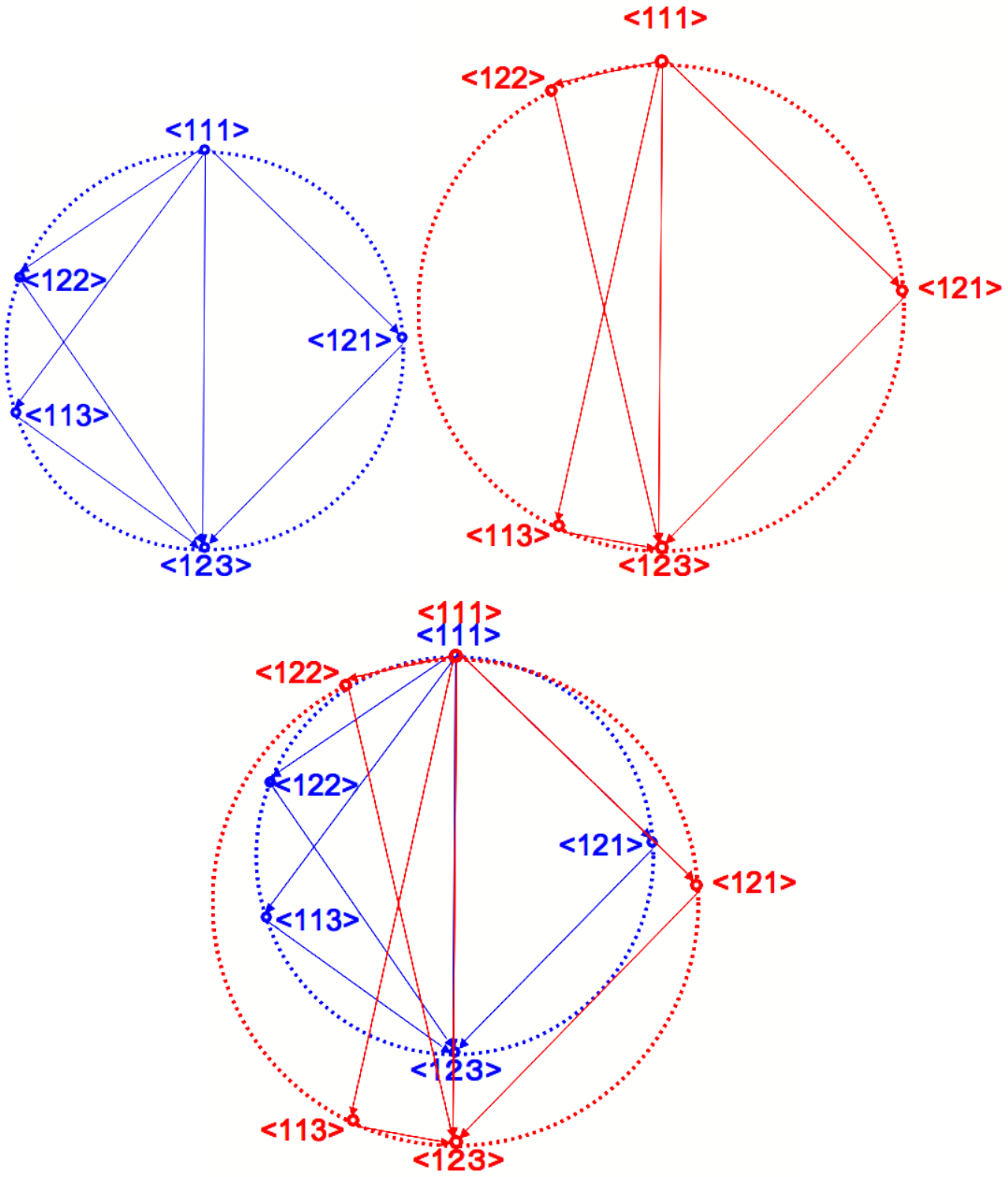


Figure 12.7: **Examples of the 3-node systems with identical cuboid-bias complexity C_c but different mutual-information I.** Top Left: System with smaller I but larger C_c^R. Top Right: System with larger I but smaller C_c^R. Bottom: Superposition of the above two systems. The regularized cuboid-bias complexity C_c^R distinguishes between the blue and red systems.

For this reason, we define the *modular complexity* C_m as follows, which is the shortest path component of the cuboid-bias complexity C_c:

$$C_m = \prod_{i=1}^{n-1} D[SD_i(i_{min}) : SD_{i+1}(i_{min})], \tag{12.13}$$

where the index i_{min} of the sequence $SD_1(i_{min}) \rightarrow SD_2(i_{min}) \rightarrow \cdots \rightarrow SD_n(i_{min})$ is chosen as follows:

$$i_{min} = \{i_1\} \cap \{i_2\} \cap \cdots \cap \{i_{n-1}\}, \tag{12.14}$$

where

$$\{i_1\} = \underset{i_s}{\text{argmin}}\{D[SD_1(i_s) : SD_2(i_s)] | 1 \leq i_s \leq |Seq|\}, \tag{12.15}$$

$$\{i_2\} = \underset{i_1}{\text{argmin}}\{D[SD_2(i_1) : SD_3(i_1)] | i_1 \in \{i_1\}\}, \tag{12.16}$$

$$\cdots$$

$$\{i_{n-1}\} = \underset{i_{n-2}}{\text{argmin}}\{D[SD_{n-1}(i_{n-2}) : SD_n(i_{n-2})] | i_{n-1} \in \{i_{n-1}\}\}, \tag{12.17}$$

which gives eventually

$$i_{min} = i_{n-1}. \tag{12.18}$$

This means that beginning from the undecomposed state $< 11 \cdots 1 >$, we continue to choose the shortest decomposition path in the next hierarchy of system decomposition. Besides its value, the modular complexity C_m should be utilized with the sequence information of the shortest decomposition path to evaluate the modularity structure of a system.

The cases where C_m are identical but C_c are different can be composed by varying the system decompositions other than in the shortest path $SD_1(i_{min}) \rightarrow SD_2(i_{min}) \rightarrow \cdots \rightarrow SD_n(i_{min})$ without modifying the index $i_m in$. There exist also inverse examples with identical C_c and different C_m, due to the complementarity between C_m and C_c.

We finally define the *regularized modular complexity* C_m^R as follows, for the same reason as defining C_c^R from C_c;

$$\begin{aligned} C_m^R &= \prod_{i=1}^{n-1} \frac{D[SD_i(i_{min}) : SD_{i+1}(i_{min})]}{D[< 11 \cdots 1 >:< 12 \cdots n >]} \\ &= \frac{C_m}{D[< 11 \cdots 1 >:< 12 \cdots n >]^{n-1}} \\ &= \frac{C_m}{I^{n-1}}. \end{aligned} \tag{12.19}$$

The cuboid-bias complexities C_c and C_c^R are bounded by the modular complexities C_m and C_m^R respectively:

$$C_c \leq C_m, \tag{12.20}$$

$$C_c^R \leq C_m^R. \tag{12.21}$$

And they coincide at the maximum values under the given multi-information I:

$$\max\{C_m | I = const.\} = \max\{\mathrm{C_c} | \mathrm{I} = const.\}, \tag{12.22}$$

$$\max\{C_m^R\} = \max\{C_c^R\}. \tag{12.23}$$

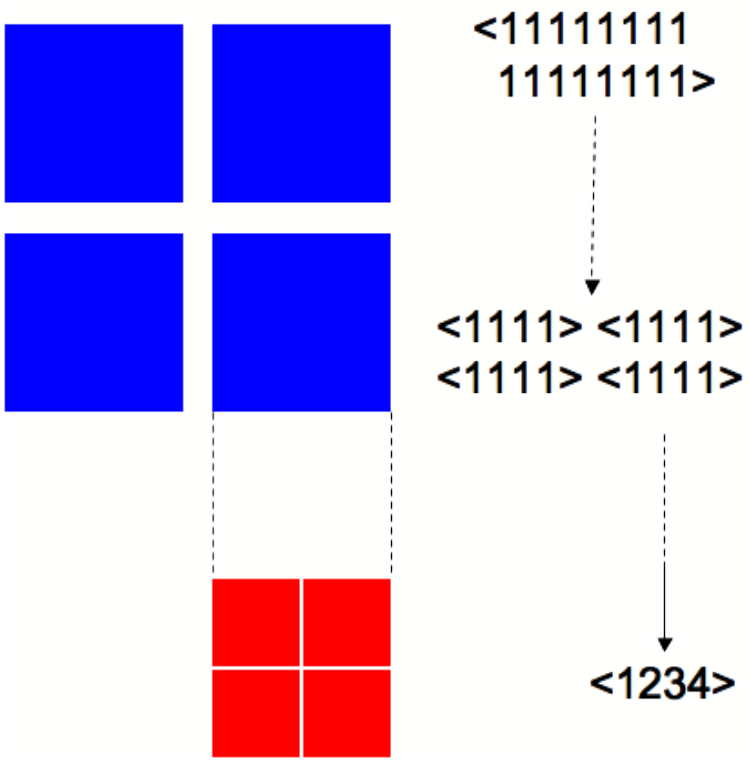


Figure 12.8: **Example of 16-node system $< 11\cdots1 >$ that has different levels of modularity.** The four 4-node subsystems $< 1111 >$ (blue blocks) are loosely connected and easy to be decomposed, while inside each component (red blocks) is tightly connected. The degree of connection represents statistical dependency or edge information between subsystems. Such hierarchical structure can be detected by observing the decomposition path of the modular complexity C_m.

These relations (12.20)-(12.23) are numerically shown in the "Numerical Comparison" section.

The superiority of the modular complexities is due to the hierarchical dependency of KL divergence value in decomposition paths. In the shortest decomposition path defining modular complexities, the easier system decomposition relatively increase its value since they incorporate more number of edge cutting. Since we eventually cut all edges to obtain $< 12\cdots n >$ at the end of the decomposition sequence, collecting the edges with relatively weak edge information and cutting them together augment the value of the product of KL divergences. The modular complexities are then the maximum value components among the possible decomposition paths calculated in cuboid-bias complexities:

$$C_m = \max\left\{ \prod_{i=1}^{n-1} D[SD_i(i_s) : SD_{i+1}(i_s)] \,\middle|\, 1 \le i_s \le |Seq| \right\}, \tag{12.24}$$

$$C_m^R = \max\left\{ \prod_{i=1}^{n-1} \frac{D[SD_i(i_s) : SD_{i+1}(i_s)]}{D[< 11\cdots1 >:< 12\cdots n >]^{n-1}} \,\middle|\, 1 \le i_s \le |Seq| \right\}. \tag{12.25}$$

The difference between the cuboid-bias complexities and the modular complexities is an index of the geometrical variation of decomposed systems in the circle graph, which reflects the fluctuation of the sequence-wise system decompositionability. If the variation of the system decompositionability for each system decomposition is large, accordingly the modular complexities tend to give higher values than the cuboid-bias complexities.

12.8 Numerical Comparison

Compare between 4 complexity measure and generalized mutual information with identical eta^{1-} and varying eta^{1+} axis1: random - max interaction axis2: $D[< 1111 >:< 1133 >]$ min - max

We numerically investigate the complementarity between the proposed complexities, C_c, C_c^R, C_m, and C_m^R. Since the minimum node number giving non-trivial meaning to these measures is $n = 4$, the corresponding dimension of parameter space is $\sum_{k=1}^{n} {}_nC_k = 15$. The constant-complexity submanifolds are therefore difficult to visualize due to the high dimensionality. For simplicity, we focus on the 2-dimensional subspace of this parameter space whose first axis ranging from random to maximum interaction of the system, and the second one representing the system decompositionability of $< 1133 >$.

For this purpose, we introduce the following parameters α and β ($0 \leq \alpha, \beta \leq 1$) in the η-coordinates of the discrete distribution with 4-dimensional binary stochastic variable:

$$\begin{aligned}
\eta_1 &= \eta_0, && (12.26)\\
\eta_2 &= \eta_0, && (12.27)\\
\eta_3 &= \eta_0, && (12.28)\\
\eta_4 &= \eta_0, && (12.29)\\
\eta_{12} &= \eta_1\eta_2 + \alpha(\eta_0 - \epsilon - \eta_1\eta_2), && (12.30)\\
\eta_{34} &= \eta_3\eta_4 + \alpha(\eta_0 - \epsilon - \eta_3\eta_4), && (12.31)\\
\eta_{13} &= \eta_1\eta_3 + \alpha\beta(\eta_0 - \epsilon - \eta_1\eta_3), && (12.32)\\
\eta_{14} &= \eta_1\eta_4 + \alpha\beta(\eta_0 - \epsilon - \eta_1\eta_4), && (12.33)\\
\eta_{23} &= \eta_2\eta_3 + \alpha\beta(\eta_0 - \epsilon - \eta_2\eta_3), && (12.34)\\
\eta_{24} &= \eta_2\eta_4 + \alpha\beta(\eta_0 - \epsilon - \eta_2\eta_4), && (12.35)\\
\eta_{123} &= \eta_{12}\eta_3 + \alpha\beta(\eta_0 - 2\epsilon - \eta_{12}\eta_3), && (12.36)\\
\eta_{124} &= \eta_{12}\eta_4 + \alpha\beta(\eta_0 - 2\epsilon - \eta_{12}\eta_4), && (12.37)\\
\eta_{134} &= \eta_1\eta_{34} + \alpha\beta(\eta_0 - 2\epsilon - \eta_1\eta_{34}), && (12.38)\\
\eta_{234} &= \eta_2\eta_{34} + \alpha\beta(\eta_0 - 2\epsilon - \eta_2\eta_{34}), && (12.39)\\
\eta_{1234} &= \eta_{12}\eta_{34} + \alpha\beta(\eta_0 - 3\epsilon - \eta_{12}\eta_{34}). && (12.40)
\end{aligned}$$

Where α represents the degree of interaction from random ($\alpha = 0$) to maximum ($\alpha = 1$), and β control the system decompositionability of $< 1133 >$. If $\beta = 1$, the system has the maximum KL divergence $D[< 1111 >:< 1133 >]$ under the constraint of α parameter, and $\beta = 0$ gives $D[< 1111 >:< 1133 >] = 0$.

ϵ is the minimum value of the joint distribution of 4-dimensional variable, which is defined to be more than 0 to avoid singularity in the dual-flat coordinates of statistical manifold. $\epsilon = 1.0e - 10$ and $\eta_0 = 0.5$ was chosen for the calculation.

The system with maximum interaction under given η_0 corresponds to the $\alpha = \beta = 1$ condition in given parameters, whose η-coordinates become as follows:

$$\begin{aligned}
\eta_1 &= \eta_0, && (12.41)\\
&\cdots &&\\
\eta_4 &= \eta_0, && (12.42)\\
\eta_{12} &= \eta_0 - \epsilon, && (12.43)\\
&\cdots &&\\
\eta_{34} &= \eta_0 - \epsilon, && (12.44)\\
\eta_{123} &= \eta_0 - 2\epsilon, && (12.45)\\
&\cdots &&\\
\eta_{234} &= \eta_0 - 2\epsilon, && (12.46)\\
\eta_{1234} &= \eta_0 - 3\epsilon, . && (12.47)
\end{aligned}$$

On the other hand, the totally decomposed system corresponds to the $\alpha = 0$ condition, and the η-coordinates are:

$$\begin{aligned}
\eta_1 &= \eta_0, && (12.48)\\
&\cdots \\
\eta_4 &= \eta_0, && (12.49)\\
\eta_{12} &= \eta_0\eta_0, && (12.50)\\
&\cdots \\
\eta_{34} &= \eta_0\eta_0, && (12.51)\\
\eta_{123} &= \eta_0\eta_0\eta_0, && (12.52)\\
&\cdots \\
\eta_{234} &= \eta_0\eta_0\eta_0, && (12.53)\\
\eta_{1234} &= \eta_0\eta_0\eta_0\eta_0. && (12.54)
\end{aligned}$$

Note that the completely deterministic case $\eta_0 = 1.0$ and $\alpha = \beta = 1$ gives $I = 0$.

The intuitive meaning of these parameters α and β are also schematically depicted in Fig. 12.10 bottom right.

Figs. 12.9 show the landscape of the proposed complexities on the α - β plane. Their contour plots are depicted in Figs. 12.10. The proposed complexities each differs from others in almost everywhere points on α - β plane except at the intersection lines. Therefore, these measures serve as the independent features of the system, each has its specific meaning with respect to the system decompositionability. The α - β plane shows a section of the actual structure of the complementarity expressed in Fig. 12.3 between the proposed complexity measures.

The relations between the cuboid-bias complexities and modular complexities in equations (12.20)-(12.23) are also numerically confirmed. The modular complexities are superior than the corresponding cuboid-bias complexities, and coincide at the parameter $\alpha = \beta = 1$ giving maximum values and interaction in this parameterization.

In general case without the parameterization with α, β and η_0, the boundary conditions of C_c, C_c^R, C_m and C_m^R include that of the multi-information I, which vanish at the completely random or ordered state. This is common to other complexity measures such as the LMC complexity, and fit to the basic intuition on the concept of complexity situated equivalently far from the completely predictable and disordered states [92] [93].

The proposed complexities further incorporate boundary conditions that vanish with the existence of a completely independent subsystem of any size. This means that the C_c, C_c^R, C_m and C_m^R of a system become 0 if we add another independent variable. This property does not reflect the intuition of complexity defined by the arithmetic average of statistical measures. The proposed complexity can better find its meaning in comparison to other complexity measures such as the multi-information I, and by interactively changing the system scale to avoid trivial results with small independent subsystem. For example, the proposed complexities could be utilized as the information criteria for the model selection problems, especially with an approximative modular structure based on the statistical independency of data between subsystems. We insist that the complementarity principle between plural complexity measures of different foundation is the key to understand the complexity in a comprehensive manner.

12.9 Analysis of Social Network Data

We analyze the social network data in the previous chapter with the use of the complexity measures C_c, C_c^R, C_m, C_m^R, and I. Figs. 12.11 and 12.12 show the results in biweekly scale. The complexities of each day were calculated using the data of 14 preceding days along the time axis.

The log-scale dynamics in the top figure of Fig. 12.11 show the strong synchronization between C_c and C_m as well as C_c^R and C_m^R. The observation in linear scale show quantitative difference between the cuboid-bias complexities and modular complexities in the middle and bottom figures. The difference between C_c and C_m tends to augment as the C_c becomes larger, while the difference

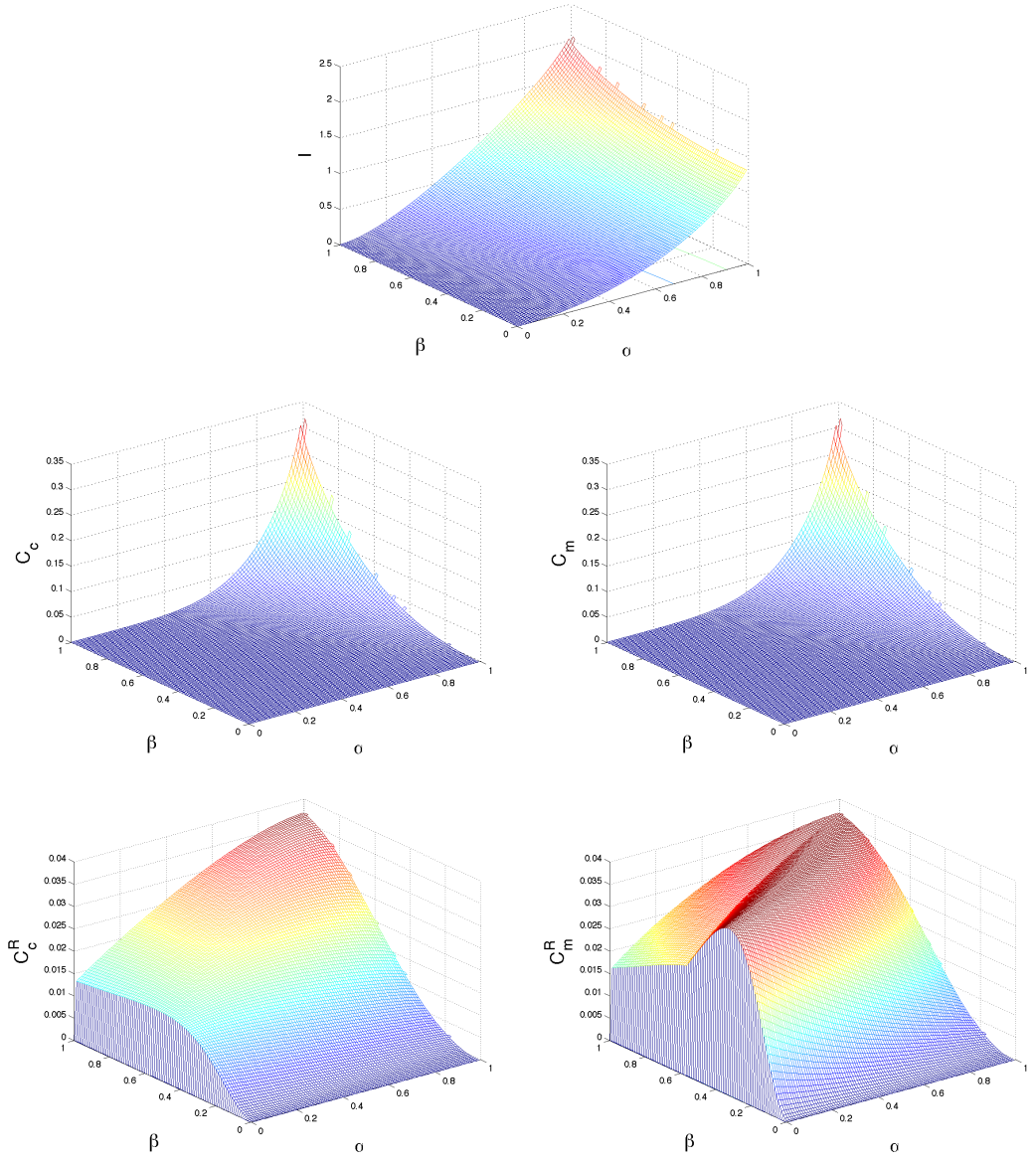


Figure 12.9: **Landscape of complexities I, C_c, C_m, C_c^R, and C_m^R on α-β plane.** Top: Mutual-information I. Middle Left: Cuboid-bias complexity C_c. Middle Right: Modular complexity C_m. Bottom Left: Regularized cuboid-bias complexity C_c^R. Bottom Right: Regularized modular complexity C_m^R. All complexity measures show the complementarity intersecting with each other, satisfying the boundary conditions vanishing at $\alpha = 0$ and $\beta = 0$ except the multi-information I.

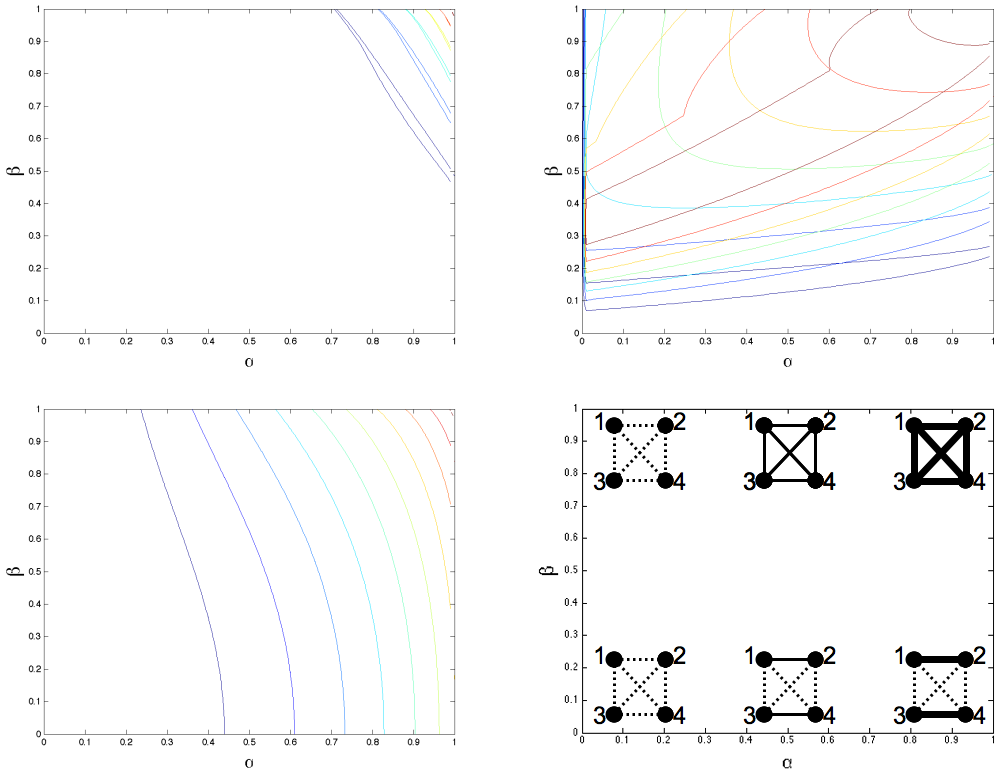


Figure 12.10: **Contour plot of the complexity landscape of I, C_c, C_m, C_c^R, and C_m^R on α-β plane.** Top Left: Contour plot superposition of C_c and C_m. Top Right: Contour plot superposition of C_c^R and C_m^R. Bottom Left: Contour plot of I. Bottom Right: Schematic representation of the system in different regions of α-β plane. Edge width represents the degree of edge information, and independence is depicted with dotted line.

between C_c^R and C_m^R does not have simple proportional relation and varies independently from other complexities. These differences show especially high value for $C_m - C_c$ and low value for $C_m^R - C_c^R$ between the 1st and 2nd round (112th-127th day) of the presidential.

The quantitative difference between the cuboid-bias complexities and modular complexities are the effect of the geometric composition of the shortest decomposition path, which represents the deviation from the maximum cuboid-bias complexities. Since the multi-information I shows high value between the 1st and 2nd round, it is natural that the complexities C_c and C_m without renormalization with I become larger during this period. Though, the augmentation of difference $C_m - C_c$ larger than the simple proportional to the C_c implies the existence of a specific subsystem contributing more than other interactions to the increase of I and C_c. This also fits to the dynamics in the top figure of Fig. 12.11, where the dynamics after the 1st round (112th day) is decreasing in C_c^R and C_m^R, increasing but decrease faster than I in C_c and C_m: Despite the increase of the I, the geometric decomposition of certain subsystem changes to reduce the complexities related to geometrical mean.

Such assumption can be verified by examining the shortest decomposition path and its KL divergence values in C_m, which are depicted in the top figure of Figs. 12.12. During the period between the 1st and 2nd round, the decomposition represented with $D[< 1214 >:< 1234 >]$ (white bar) continues to augment as other $D[< 1111 >:< 1114 >]$ (black bar) decreases and $D[< 1114 >:< 1214 >]$ (magenta bar) remains within a bounded range. The decomposition $D[< 1214 >:< 1234 >]$ (white bar) corresponds to the interaction only between R and S, who actually fought the 2nd round to eventually decide the next president. The dynamics of C_c and C_m then reflected the stronger increase of interaction between R and S than between other candidates.

On the other hand, the values of the regularized complexities C_c^R and C_m^R tends to decrease between the 1st and 2nd round. The theoretical maximum of the C_c^R and C_m^R is $\left(\frac{1}{n-1}\right)^3 = 0.037$ in case $n = 4$. The maximum regularized complexities correspond to the lowest system decompositionability. The temporal maximum of the C_c^R and C_m^R appears around 105th day, followed by the decrease after the 1st round (112th day).

This dynamics is consistent to the analysis of C_c and C_m, since as the interaction between R and S increases, the relative strength of other interactions under regularization with I decreases, which results in lower value of C_c^R and C_m^R. The relative decompositionability between the decompositions in the shortest path of the C_m^R is depicted in the bottom figure in Figs. 12.12, corresponding to the dynamics ofC_m.

The difference $C_m^R - C_c^R$ remains almost constant during the 1st to 2nd round. This can be interpreted as the setoff between the proportional decrease of the difference with respect to the C_c^R value, and the increase originated from the compositional change of system decompositions.

The shortest decomposition path in Figs. 12.12 also implies the modularity of the topics in discussion. As the first decomposition is always $< 1114 >$ (black bar), the extreme-rightist L is assumed to be the most marginal in semantic context. The second decomposition is dominantly $< 1214 >$ (magenta bar) but varies between $< 1134 >$ (yellow bar) before the 1st round (112th day) and $< 1224 >$ (red bar) after the 2nd round (127th day). This means there was a period when people considered as the most important comparison between R and B before the 1st round, and between B and S after the 2nd round in a collective level.

The proposed measures all have different dynamics from the multi-information I, implying different geometrical information of the system decompositionability. Actual data analysis would be better performed with the use of such complementary complexity measures according to one's interest and necessity of modeling.

12.10 Further Consideration

We further derive some analytical sketch on the proposed complexity measures.

12.10.1 Preparation for Extreme Value Analysis

We consider the extreme value analysis of the regularized cuboid-bias complexity C_c^R on the parameter space defined by the KL divergence of system decompositions.

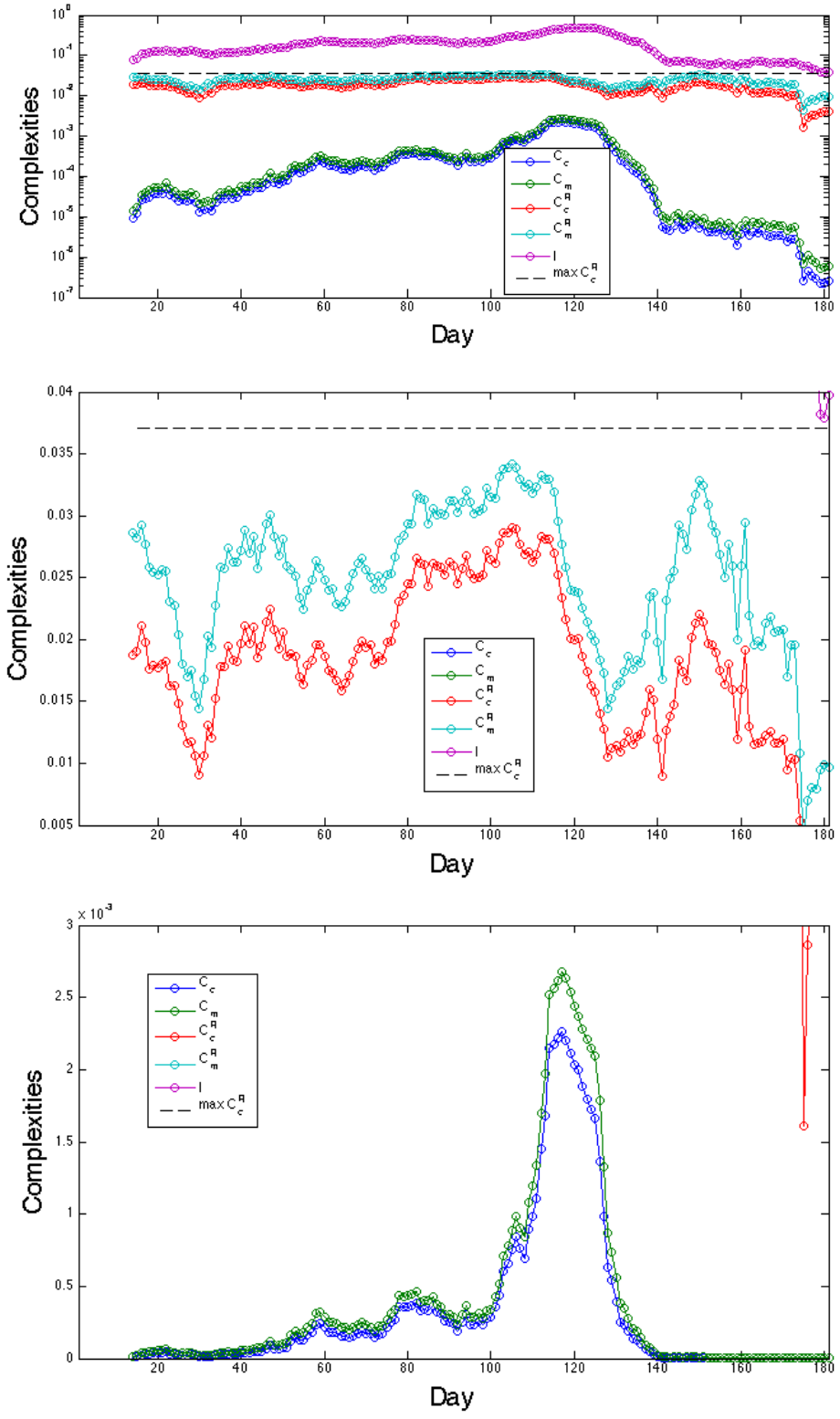


Figure 12.11: **Complexities dynamics during the french presidential in 2007.** Top: Log plot of C_c, C_c^R, C_m, C_m^R, and I in biweekly scale. Middle and Bottom: Linear plot of C_c, C_c^R, C_m, C_m^R in biweekly scale. Note that the linear scale of vertical axes is different between the middle and bottom figures.

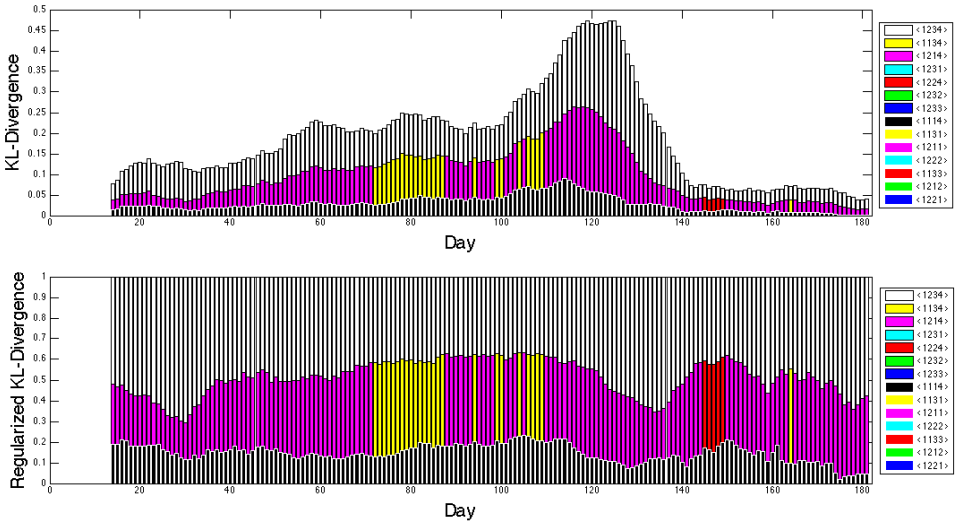


Figure 12.12: **Dynamics of modular complexities during the french presidential in 2007.** Top: Temporal dynamics of the modular complexity C_m in biweekly scale. Bottom: Temporal dynamics of the regularized modular complexity C_m^R in biweekly scale.

For preparation, we recall that each system decomposition is defined with the modification of the different set of **p** parameters defined as follows:

$$\mathbf{p} = \left\{ p(x_1, x_2, \cdots, x_n) \left| x_1, x_2, \cdots, x_n \in \{0,1\}, \sum_{x_1,x_2,\cdots,x_n \in \{0,1\}} p(x_1, x_2, \cdots, x_n) = 1 \right. \right\}. \quad (12.55)$$

Since the number of different system decompositions m is less than the number of **p** parameters having $\sum_{k=1}^{n} {}_nC_k$ degrees of freedom with respect to the system size $n \geq 3$, the following renormalized KL divergences of the system decompositions form another coordinates as dimension reduction projection:

$$\mathbf{D_{SD}} = \left\{ \frac{D[< 11\cdots1 >: SD_i(i_s)]}{D[< 11\cdots1 >:< 12\cdots n >]} \middle| 1 \leq i \leq n-1, 1 \leq i_s \leq |Seq| \right\}. \quad (12.56)$$

These coordinates are incomplete to specify a distribution in case $n = 3, 4$, but is sufficient to determine the values of complexities C_c, C_c^R, C_m, C_m^R, and I. As for $n \geq 5$, the number of system decompositions m becomes larger than the model dimension. For simplicity to discuss the property of proposed complexity measures with respect to the system decompositionability, we denote these coordinates with KL divergences as $\mathbf{D_{SD}} = \{D_{SD_j} | 1 \leq j \leq m\}$ and investigate their independency.

We consider the general case where there exist 3 different systems namely i, j, k, and its decompositions SD_i, SD_j and SD_k, as represented in the top left figure of Fig. 12.13. We consider the system decompositions as the sum of edge information EI compatible to Pythagorean relation, as defined in the "Further Consideration: Pythagorean Relations in System Decomposition and Edge Cutting" section of previous "Network Decomposition" chapter:

$$D_{SD_i} = EI_{ij}^1 + EI_{ik}^2, \quad (12.57)$$
$$D_{SD_j} = EI_{ji}^1 + EI_{jk}^2, \quad (12.58)$$
$$D_{SD_k} = EI_{ki}^1 + EI_{kj)}^2, \quad (12.59)$$

where for example, EI^l_{ij} is the l-th term of edge information in the Pythagorean relation cutting the links between subsystems i and j. For the convenience of analysis, we symmetrize the edge information as follows, for all combinations of subscripts i, j, and k.

$$EI = \frac{EI^1 + EI^2}{2}. \tag{12.60}$$

Then we have the following integrated relations between system decompositions and symmetrized edge informations:

$$D_{SD_i} = EI_{ij} + EI_{ik}, \tag{12.61}$$
$$D_{SD_j} = EI_{ji} + EI_{jk}, \tag{12.62}$$
$$D_{SD_k} = EI_{ki} + EI_{kj}. \tag{12.63}$$

Note the symmetrized edge information is not symmetric with respect to the order of subscript, such as $EI_{ij} \neq EI_{ji}$.

For later analysis, we also define the total edge information EI_{kk}, necessary to cut arbitrary set of edges inside of the subsystem k as an additional part of the system decomposition SD_k:

$$D_{SD_k} = EI_{ki} + EI_{kj} + EI_{kk}. \tag{12.64}$$

Since the transformation between $\mathbf{p}$ parameters and $\mathbf{D_{SD}}$ coordinates are complex, the independency inside of $\mathbf{D_{SD}}$ need to be investigated. The following describes the outline of analysis for the general case of $n \geq 3$, which is schematized in the top left figure of Fig. 12.13.

1. From the symmetry of system, it is sufficient to show in case we fix the edge information inside of subsystems i, j to focus on the relation between the system decompositionability D_{SD_i} and D_{SD_j}, while SD_k representing other arbitrary variable system decompositions.

2. In case we only have edge information between subsystems i and j, meaning $D_{SD_i} = D_{SD_j} > 0$ and $EI_{ik} = EI_{ki} = EI_{jk} = EI_{kj} = EI_{kk} = 0$, naturally the constraint is $D_{SD_i} = D_{SD_j}$. The two coordinates D_{SD_i} and D_{SD_j} are completely dependent.

3. Otherwise, $D_{SD_k} > 0$ holds in the definition (12.64). In this case, there exists independent region between D_{SD_i} and D_{SD_j}, under the constraint of $D_{SD_j} = D_{SD_i} - (D_{SD_k} - EI_{kk})$ and $D_{SD_i} = D_{SD_j} - (D_{SD_k} - EI_{kk})$ as border lines. It is because we can change the value of D_{SD_i} and D_{SD_j} by distributing the edge information $D_{SD_k} - EI_{kk}$ differently between EI_{ki} and EI_{kj}, maintaining the total value D_{SD_k}. The amount of $D_{SD_k} - EI_{kk}$ therefore defines the limit of independency between D_{SD_i} and D_{SD_j}.

4. Another border line exists as $D_{SD_i} + D_{SD_j} = D_{SD_k} - EI_{kk}$ since this coincides with the case $EI_{ij} = EI_{ji} = 0$. Since $EI_{ij}, EI_{ji} \geq 0$, the relation $D_{SD_i} + D_{SD_j} \geq D_{SD_k} - EI_{kk}$ gives another condition of independent region of D_{SD_i} and D_{SD_j} .

5. As for changing the edge information inside of the subsystems i or j, it suffices to change their definition so that to treat these edges with EI_{kk}.

The independent region between two arbitrary system decompositionability D_{SD_i} and D_{SD_j} are depicted as white region in the top right figure of Fig. 12.13. The gray region is not realizable depending on the rest of the edge information $D_{SD_k} - EI_{kk}$.

12.10.2 Extreme Value Analysis

We now consider the extreme value analysis of the regularized cuboid-bias complexity C^R_c in the independent region of $\mathbf{D_{SD}}$ coordinates. The extreme values of a function is usually defined as points where all partial derivatives are 0. However, such point does not necessary have geometric

meaning other than global maximum in cuboid-bias complexities. Instead, we consider the points where at least one partial derivative is 0, namely "*partial extreme values*".

The partial extreme values assure the existence of decomposition paths giving the maximum value of its cuboid volume in circular diagram. This also means that the corresponding sequences in Seq are the most difficult paths, impossible to find easier decomposition. If we count the number of decomposition sequences giving maximum regularized cuboid volume $(\frac{1}{n-1})^{n-1}$ with s_{max} $(0 \leq s_{max} \leq |Seq|)$, there exists the following hierarchical relation between the regularized cuboid-bias complexity $C_c^R(s_{max})$ conditioned with s_{max} value:

$$C_c^R(0) < C_c^R(1) < \cdots < C_c^R(s_{max}) < \cdots < C_c^R(|Seq|). \tag{12.65}$$

And the corresponding sets on the $\mathbf{D_{SD}}$ coordinates also follow geometrical hierarchy, as depicted in the bottom figure of Fig. 12.13. The hierarchy corresponds to add the following constraints on the parameter space $\mathbf{D_{SD}}$, according to the value of s_{max}.

$$D_{SD_i} = \frac{1}{n-1}, \tag{12.66}$$

where

$$i \in U \subset \{1, 2, \cdots, m\},\ |U| = s_{max}. \tag{12.67}$$

The corresponding sets in $\mathbf{D_{SD}}$ space of each $C_c^R(s_{max})$ give partial extreme values with respect to the D_{SD_i} axis. This fact can also be verified with the use of variational method considering the perturbation expansion on $\mathbf{D_{SD}}$. In any case, the proof is reduced to simple isoperimetric inequality of $n-1$-dimensional cuboid.

We can also consider the decomposition of C_c^R with respect to the value of s_{max}. Beginning from the C_c^R value of interest, for example calculation result from a data set, we consider the path toward the maximum value of $C_c^R = C_c^R(|Seq|)$, following the geometrical structure induced by partial extreme values. If the data set belongs to the $C_c^R(s_{data})$ region with concrete value of s_{data}, it is possible to consider the following shortest path to $C_c^R(|Seq|)$ by changing only one element of $\mathbf{D_{SD}}$ in each step:

$$C_c^R(s_{data}) \to C_c^R(s_{data}+1) \to \cdots \to C_c^R(|Seq|). \tag{12.68}$$

For example, the decomposition path $C_c^R(0) \to C_c^R(1) \to C_c^R(2)$ is depicted with dashed line in the bottom figure of Fig. 12.13. Note that we can not necessary choose one element to follow this decomposition, since the independence inside of $\mathbf{D_{SD}}$ is not complete. In such case, we can still find the shortest path skipping the impossible projection. The example to avoid $C_c^R(1)$ and take $C_c^R(0) \to C_c^R(2)$ is depicted with dotted line.

The decomposition path (12.68) contains system decomposition-wise information of $C_c^R(s_{data})$ with respect to $C_c^R(|Seq|)$, and the difference along this hierarchy quantify the contribution of each element in $\mathbf{D_{SD}}$ to C_c^R.

The partial extreme values are also interesting whether it appears in actual data analysis. The augmentation of C_c^R is mutually observed in social network analysis as well as the autonomous learning process of CNN in Part 2, though the augmentation route in $\mathbf{D_{SD}}$ space remains to be revealed. This leads us to hypothesize the utility of partial extreme values as a kind of potential function underlying the dynamics in complex systems, questioning whether the augmentation of C_c^R follows subsystem-wise, hierarchically and temporally ordered maximization, or rather homogeneous and simultaneous increase. In earlier case, partial extreme values play important role for the characterization of the dynamics.

Figs. 12.14 show the actual dynamics of the political weblog data on $\mathbf{D_{SD}}$ coordinates. There exist certain phases that system decompositionability approaches toward an edge of maximum cuboid volume $\frac{1}{n-1}$, uncorrelated to others. These dynamics support the notion of partial extreme

values as peaks of potential function, and explain the increase of complexity as a geodesic flow with respect to the closest partial extreme values $C_c^R(1)$. Positive correlation between two elements of $\mathbf{D_{SD}}$ means the potential of geodesic flow has its peak in $C_c^R(2)$ instead of $C_c^R(1)$. Further higher-order correlations can find appropriate potential peaks in the hierarchical structure of $C_c^R(s_{max})$ in the same manner. Strong negative correlation was also observed between $\mathbf{D_{SD}}$, meaning the subsystem-wise transition of complexity is occurring (last figure of Figs. 12.14, during the 1st and the 2nd round (112th-127th day) of presidential).

12.10.3 Relation to Algebraic Geometry

Since the landscape on $\mathbf{D_{SD}}$ is important to characterize the property of C_c^R, it is useful to consider the geometrical structure of the isometric surface with $C_c^R = const.$ condition. For that purpose, analysis with algebraic geometry can be considered as a prominent tool. Algebraic geometry investigate the geometrical property of polynomial equations. The regularized-bias complexity C_c^R on $\mathbf{D_{SD}}$ space is simply the polynomial function, therefore directly accessible to algebraic geometry.

However, if we want to investigate the isometric surface of C_c^R on the $\mathbf{p}$ parameter space, the definition of KL divergence inevitably include logarithmic function, which is a transcendental function and outreach the analytical requirement of algebraic geometry.

To introduce compatibility between the $\mathbf{p}$ parameter space of information geometry and algebraic geometry, it suffices to expand the model by replacing the logarithmic functions as another variables such as $q = \log p$, and reconsider the intersection between the result from algebraic geometry and $q = \log p$ function.

The isometric surface of C_c^R is also important to test the utility of this measure as a potential of the dynamics.

12.10.4 Complexity of the Systems with Continuous Phase Space

We have developed the concept of system decompositionability based on discrete binary variables. One can also apply the same principle to continuous variable.

For an ergodic map $G : X \to X$ in continuous space X, KS entropy $h(\mu, G)$ is defined as the maximum of entropy rate with respect to all possible system decomposition A, when the invariant measure μ exists:

$$h(\mu, G) = \sup_A h(\mu, G, A). \tag{12.69}$$

Where A is the disjoint decomposition of X that consists of non-trivial sets a_i, whose total number is $n(A)$, defined as

$$X = \bigcup_{i=1}^{n(A)} a_i, \tag{12.70}$$

$$a_i \cap a_j = \phi,\ i \neq j,\ 1 \leq i, j \leq n(A), \tag{12.71}$$

meaning the natural expansion of system decomposition into continuous space.

The entropy rate $h(\mu, G, A)$ in (12.69) is defined as

$$h(\mu, G, A) = \lim_{n\to\infty} \frac{1}{n} H(\mu, A \vee G^{-1}(A) \vee \cdots \vee G^{-n+1}(A)), \tag{12.72}$$

according to the entropy $H(\mu, A)$ based on the decomposition $A = \{a_i\}$

$$H(\mu, A) = -\sum_{i=1}^{n(A)} \mu(a_i) \ln \mu(a_i), \tag{12.73}$$

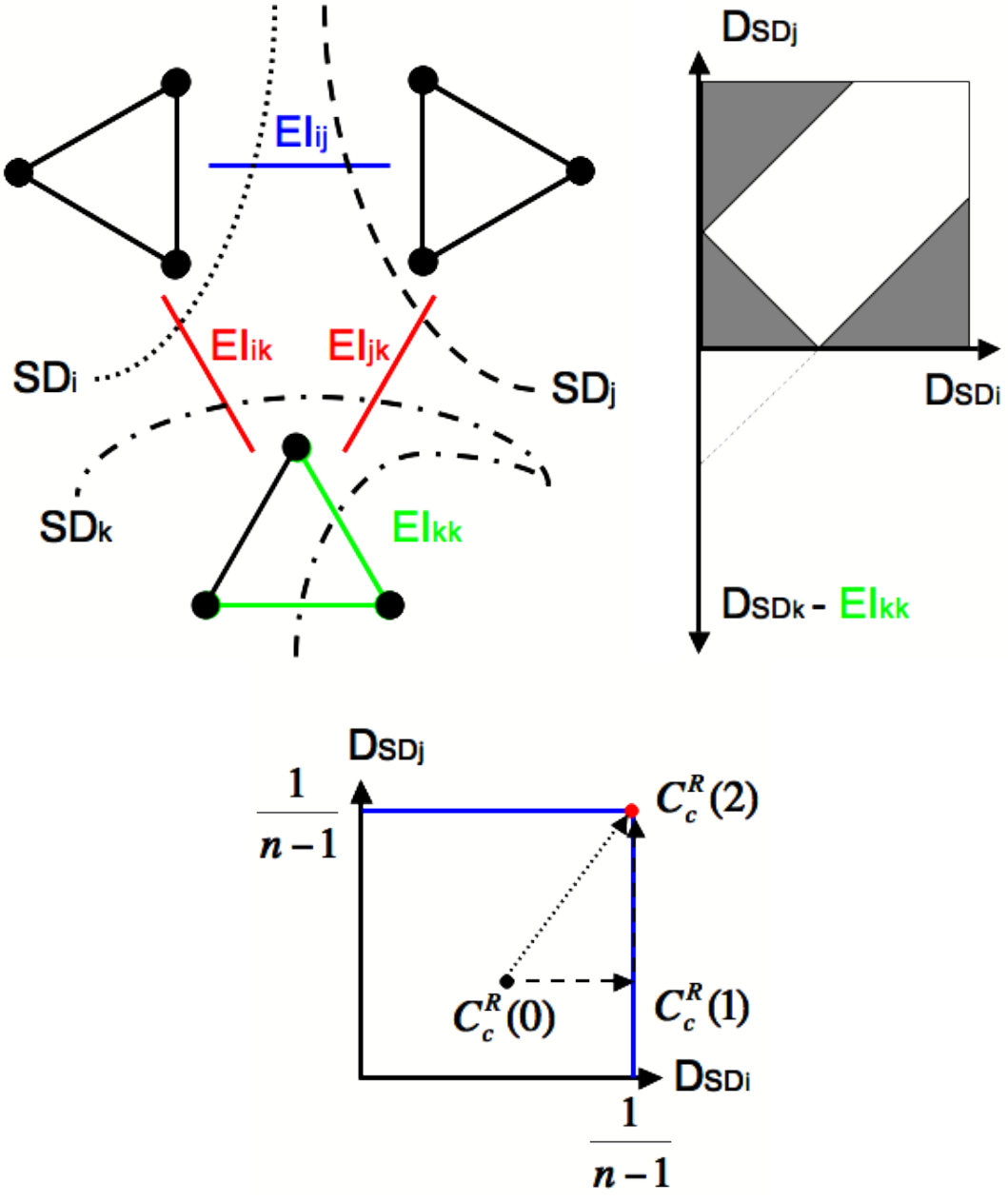
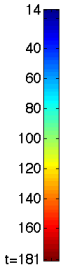
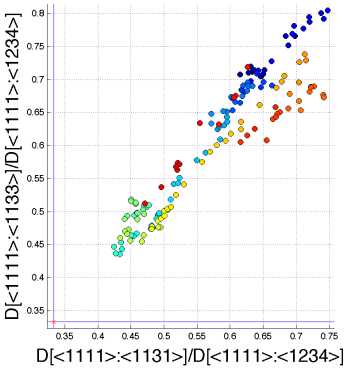
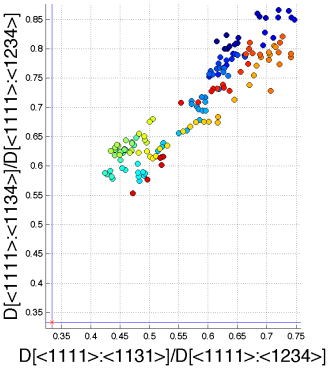
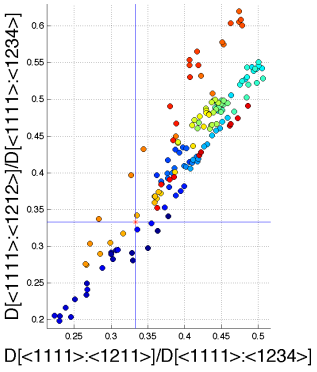


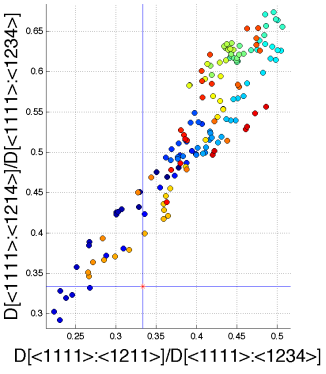
Figure 12.13: **Graphical explanation related to the extreme value analysis.** Top Left: Definition of system decompositions SD_i, SD_j, and SD_k to decompose subsystems i, j, k, and related edge information. Top Right: Independent region of the $\mathbf{D_{SD}}$ coordinates. Bottom: Geometrical hierarchy in $\mathbf{D_{SD}}$ space corresponding to the hierarchy of regularized cuboid-bias complexity $C_c^R(s_{max})$. The coordinates belonging to $C_c^R(1)$ are depicted with blue lines, while that of $C_c^R(2)$ is plotted with red point. Only D_{SD_i} - D_{SD_j} plane is shown.

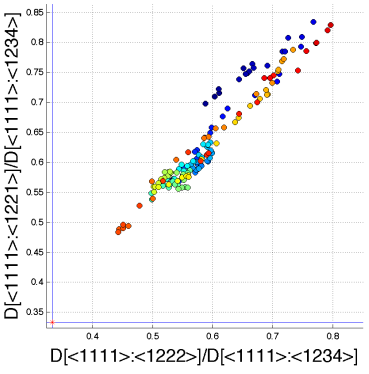


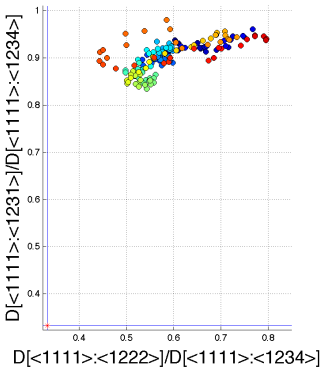


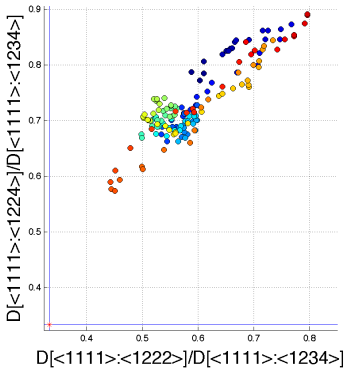


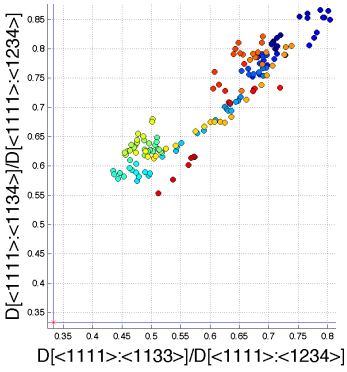


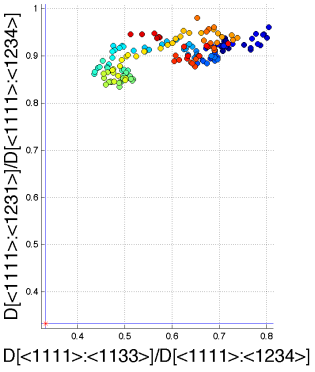


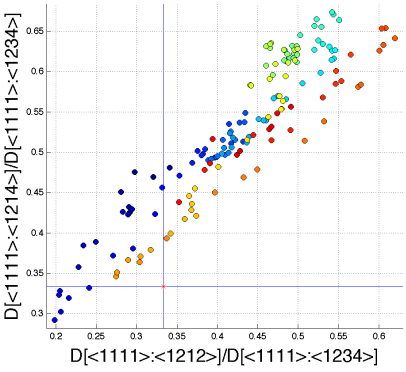


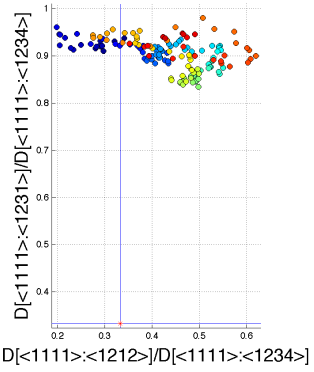


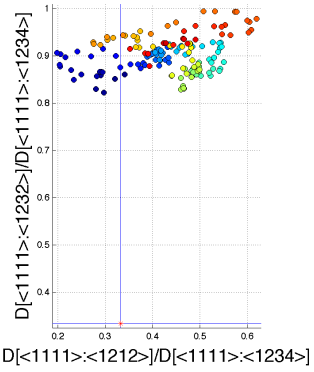


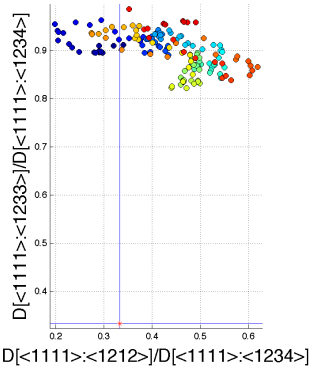












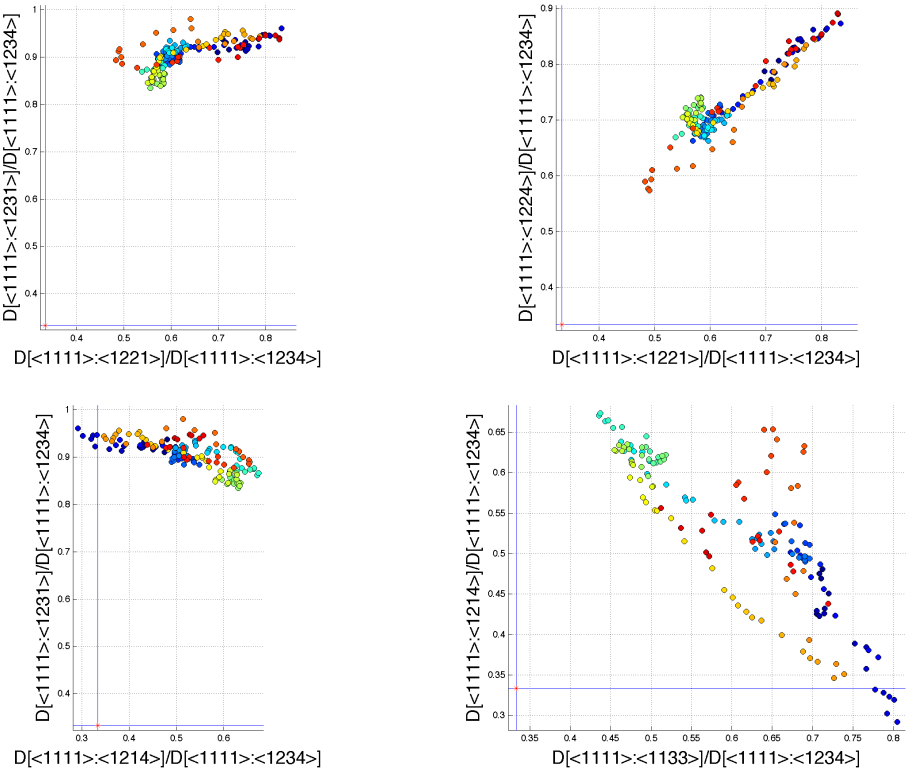


Figure 12.14: **Examples of uncorrelated and correlated dynamics of political weblog data on $\mathbf{D_{SD}}$ coordinates in biweekly scale.** The surface color of the circles represents the time axis during the presidential. The blue lines correspond to the partial extreme values $C_c^R(1)$, while the red asterisk is $C_c^R(2)$. Both uncorrelated and correlated dynamics can be approximatively explained as a geodesic flow on $\mathbf{D_{SD}}$ toward $C_c^R(1)$ and $C_c^R(2)$, respectively. The last figure shows an example of negative correlation representing the transition of complexity from one subsystem to another, during the 1st and the 2nd round (112th-127th day) of presidential.

and the product $C = A \vee B$ as

$$\begin{aligned} C &= A \vee B \\ &= \{c_i = a_j \cap b_k | 1 \leq j \leq n(A), 1 \leq k \leq n(B)\}. \end{aligned} \tag{12.74}$$

In a more general case, topological entropy $h_T(G)$ is defined simply with the number of decomposed subsystem elements by preimages as follows, without requiring ergodicity, therefore neither the existence of invariant measure μ:

$$h_T(G) = \sup_A \lim_{n\to\infty} \frac{1}{n} \ln n(A \vee G^{-1}(A) \vee \cdots \vee G^{-n+1}(A)). \tag{12.75}$$

Topological entropy take the maximum value of the possible preimages division, so that to measure the complexity in terms of the mixing degree of the orbits. For example, if the KS entropy is positive as $h(\mu, G) > 0$, the dynamics of G on an invariant set of invariant measure μ is chaotic for almost everywhere initial conditions. As for the positive topological entropy $h_T(G) > 0$, the dynamics of G contain chaotic orbits, but not necessary as attractive chaotic invariant set, since $h_T(G) \geq h(\mu, G)$ and the KS entropy can be negative.

Although these definitions are useful to characterize the existence of chaotic dynamics, the system decompositionability is another property representing different aspect complexity of the system. It is rather the matter of the existence of independent dynamics components, or the degree of orbit localization between arbitrary system decompositions. We propose the following "*geometric topological entropy*" $h_g(G)$ applying the same principle of taking geometric product between all hierarchical structure of the system decomposition A.

$$h_g(G) = \prod_{\sigma(A)>0} \lim_{n\to\infty} \frac{1}{n} \ln n(A \vee G^{-1}(A) \vee \cdots \vee G^{-n+1}(A)), \tag{12.76}$$

where $\sigma(A) > 0$ means to take all components of A having positive Lebesgue measure on X.

This gives 0 if the preimage of certain $a_i \in A$ is a_i itself, meaning there exist a subsystem a_i whose range is invariant under G, closed by itself. The system X can be completely divided into a_i and the rest. This correspond to the existence of an independent subsystem in cuboid-bias and modular complexities. In case such independent components do not exist, it still reflects the degree of orbit localization for all possible system decompositions in multiplicative manner. The condition $\sigma(A) > 0$ is to avoid trivial case such as the existence of unstable limit cycle, whose Lebesgue measure is 0.

Typical example giving $h_g(G) = 0$ is the function having independent ergodic components, such as the Chirikov-Taylor map with appropriate parameter [94].

12.11 Conclusion

We reconsidered the problem of how to define complexity measures in terms of the construction of non-linear feature space. We defined new type of complexity based on the geometrical product of KL divergence representing system decompositionability. Analysis of social network data is demonstrated. Different complexity measures as well as newly proposed ones are compared on a complementarity basis on statistical manifold. Further consideration implied a rich connection to other theoretical fields.

Part V

Reconstructing Embryogenesis

Chapter 13

Project Description

Abstract

The context of works in the following chapters of this part is introduced. The works were realized in two european projects, Embryomics and BioEMERGENCES. The aims and scope of these projects are summarized, with the global strategy that defines the standpoint of the works in the following chapters.

Keywords: Embryogenesis, Zebrafish, Phenomenological reconstruction, Theoretical reconstruction

13.1 Introduction: Context of work

The following chapters in this part was realized in two european projects, Embryomics (NEST 012916) [95] and BioEMERGENCES (NEST 028892) [96]. The members of the projects consist mainly of experimental biologists, engineers and various scientists including applied mathematicians, physicists, and computer scientists. The projects were mostly devoted to experimental measurement of animal embryos and spatio-temporal reconstruction of their morphogenesis with cell lineage tree, which will be explained in the next section.

As a free theorist, I was fully enjoying the discussion on theoretical aspects concerning scientific evaluation of obtained data, during several week-long workshops in each participating laboratory. Since the scientific evaluation with a dynamical system perspective on morphogenesis presuppose the reconstructed 4-D morphogenetic dynamics and lineage tree, my work does not directly contribute to the primal reconstruction process of cell lineage from the raw image data, but rather seek for the possible theoretical formalization that would bring scientific impact using the derived whole-embryo simulator.

In this chapter, we briefly summarize the aim and strategy of these projects to contextualize the work of the following chapters.

13.2 Project Description of Embryomics and BioEMERGENCES Projects

13.2.1 Aims and Scope

In Embryomics and BioEMERGENCES projects, we aim at providing an experimental platform to observe *in vivo* emergent patterns at various scales and measure their variability between different individuals of the same species. This is a strategy towards the measurement of the individual susceptibility to genetic diseases or response to treatments. Emergent patterns arising at all levels of living organisms are influenced both by the external environment (top-down) and by the internal environment (bottom-up). As a consequence, two living beings are different even if they are two clones of the same species because the history of their coupling with their external environment is different. For these reasons, medicine evolves towards personalized protocols. To make them

tangible, we have to be able to achieve the measurement at all organization levels of the individual response to genetic defects or xenobiotics.

The emergence of personalized medicine is a good indicator that the perception and interpretation of the individual variation has changed. Physicians dealing with so-called "genetic diseases" are left with the responsibility to anticipate whether a patient with an identified genetic defect is going to develop or not the corresponding disease. Giving an answer is a matter of measuring the impossible.

The impossible measurement of individual differences has to be tackled in an animal model suitable for high throughput *in vivo* investigations. The zebrafish *Danio rerio* gathers a number of interesting features including the accessibility of its embryo and the transparency of its tissues making it suitable for *in vivo* investigations at different scales, varying from genetic, molecular, cellular, tissularÉ to whole individual. In addition, the zebrafish has been largely validated as a powerful model for investigations related to human, and will soon become a major model organism for pre-clinical drug testing by pharmaceutical industries. This non-mammalian vertebrate animal model will allow us to tackle the measurement of the individual susceptibility to genetic diseases or individual response to treatments.

The Embryomics project is a first step towards an integrated understanding of biological processes. The Embryomics project is devoted to the morphodynamical "reconstruction" of the cell lineage tree underlying the processes of animal embryogenesis and provides a set of strategies, methods and algorithms to "sequence" the cell lineage tree as a branching process deployed in space and time. This is being investigated in the scale of a single individual of several species including zebrafish.

The BioEmergences project has a different much more general scope as it will i) address the general problem of reconstructing various types of morphodynamical patterns for a large number of individuals, and ii) cope with the very difficult problem of the individual variation, its various manifestations and the question of how to measure it. The experimental platform is also applied to a class of mutant embryos (genetic model for holoprosencephaly).

As a part of the BioEMERGENCES project, the individual response of the mutant embryos to treatments will be investigated by testing the Dbait molecules.The Dbait molecules were very recently designed as intelligent anti-cancer drugs to especially target tumors that are resistant to classical therapies. The Dbait molecules, designed to delude cancer cells, are nucleic acid derivatives thought to prevent the replication of DNA and interfere with the DNA repair process. Indeed Dbait molecules seem to have no toxic effects by themselves but kill cancer cells in conjunction with a DNA breaking agent (ionized radiation or chemical agent). Consequently, cells treated with a "Dbaits combinational therapy" should stop proliferating and undergo apoptosis or cell death rapidly and at a high rate. This should also happen even for cells resistant to classical cytotoxic agents. This project will contribute to the extensive studies actually performed to evaluate the therapeutic potential of the Dbait molecules and improve their design.

The main result expected from BioEMERGENCES is the specification of a European platform to achieve high throughput measurement of individual differences and screening of drugs combinations such as bi or tri-therapies. Such a platform will allow responding to both the unavoidable scientific question of the construction of a synthetic description of individuals and the future requirement for new drugs in the field of personalized nano-medicine.

13.2.2 Measurement and Reconstruction Methodology

For these purposes, the 4-D image data sets were obtained by *in vivo* time lapse optical sectioning of single cells, group of cells, morphogenetic fields and possibly whole organisms, to allow the qualitative and quantitative measurement of individual differences. 4-D *in vivo* imaging are produced by using the most recently developed imaging techniques including high-speed confocal laser scanning microscopy (CLSM), multiphoton laser microscopy (MLSM) and selective plane illumination microscopy (SPIM).

Embryo labeling was obtained through RNA injection performed at the one-cell stage to obtain ubiquitous expression of H2B/mcherry fusion protein and farnesylated eGFP, which stained nuclei and membranes respectively [97] [98]. While nuclear staining was instrumental to perform cell

tracking, membrane staining was essential to assess cell morphology, behavior and neighborhood, and to reveal morphological landmarks.

Our whole computational image-processing framework in Embryomics project was inspired from theoretical knowledge about mammalian vision. For each low-level image-processing task we designed specific non-linear partial differential equations (PDEs) that present the remarkable adaptive properties of the mammalian visual system and able to deal with increasing difficulty found at later developmental stages.

Image filtering was achieved with the so-called geometric mean curvature (GMC) flow in the level set formulation, belonging to the geometrical nonlinear partial differential equations family, that appeared to have the best performances in processing MLSM images [99]. The GMC contrast invariant flows performed non-linear multiscale analysis on scalar intensity functions in grey level scale. Filtered images were used in the subsequent algorithmic steps.

Nuclei centre detection was performed using the PDEs based nonlinear multiscale strategy called flux-based level set centre detection (FBLSCD). All visible objects in the image can be seen as humps of relatively higher image intensity. The multiscale FBLSCD method makes the hump decreasing until a stopping condition adapted to the rest of the algorithm chain.

Cell nuclei and cell membrane segmentation used the approximate cell centers from the above step as the gaze from which a point of view surface was constructed. The evolution of the point of view surface with respect to a metric induced by the image tended to a minimal surface in a Riemannian manifold representing the segmentation of the nucleus or membrane shape. This subjective surface technique (SST) was used for its ability to fill missing information in the image [100][101][102][103][104]. SST numerical implementation was based on co-volume methods [105]. Nuclei segmentation was then used to correct nuclei double centers that remained from the nuclei centers detection step. Finally, membranes segmentation was used to detect mitosis and this information was implemented into the tracking algorithm. The cell number in the imaged volume increased from 3291 at t=0 (sphere stage) to 11 176 at t=540 (8-somite stage) where as a rough estimate we observed about 1/3 of the embryo total cell number. The cell density obtained from the segmentation of the imaged global cell volume and plotted as a function of time provided an estimate of the average proliferation rate. As expected, the average cell density stabilized by the end of gastrulation and early somitogenesis (time steps 300 to 360).

Cell tracking in the 4D space of the segmented data was not properly achieved by classical methods using a nearest neighbor principle. Although we kept the temporal resolution as high as possible ($\triangle$t is 67 seconds in the image data set 070418a), cell displacement at the time of cell division was a major problem that the tracking algorithm had to solve. The spatio-temporal lineage tree reconstruction was best achieved with a tracking algorithm from the EM (Estimation-Maximization) procedure family [106]. Starting with a first spatiotemporal lineage tree mainly based on "a minimal deformation/ nearest neighbor principle" [107], the E-step provided the probability distribution of the cell dynamics in its features space and the M-step calculated the next spatiotemporal lineage with maximum likelihood. The iterative EM procedure corrected false negative and false positive from the nuclei centers detection step. More generally, the EM procedure is able to learn from observed errors and by iterating the two EM steps, the maximum likelihood of the lineage tree is guaranteed to be increasing and thus convergent.

The outcome of the tracking procedure was assessed by using the Embryomics visualization interface and following the path of individual cells. The visualization interface allows going back and forth between the 4D rendering and the annotated flat representation of the tree. The Embryomics algorithmic chain provided for each cell at each time step its identification number, 4D coordinates, nucleus and membrane shape, and the time of previous and next division along the tree. A number of features characteristics for cell dynamics might then be derived from theses parameters. The tracking error rate was measured with the sampling on the interface. More than 15 000 links between the t and t+1 positions of cells have been checked. For the first 300 time steps (5h34 imaging, end of gastrulation), the total error rate is lower than 2% meaning that more than 98% of the cells were correctly tracked. Tracking accuracy degraded beyond time step 400 (7h26 imaging, 3 somite stage) and by time step 500 (9h18 minutes imaging, 7 somite stage), 90% of the links were correctly found. Nuclei centers detection was not a major source of tracking errors as false positive detection was kept lower than 0,5%. About half of the tracking errors were identified has false tracks (the tracking jumped from one cell trajectory at time t to another cell trajectory at time

t+1) and the other half was false deaths (the tracking did not find the position of the cell at time t+1 and cell trajectory ends as for a dead cell). All the information provided by error checking and categorization might be used to improve the performances of the EM tracking procedure in further cycles. Thus the reconstruction accuracy is expected to improve from users annotation through the Embryomics visualization interface.

The computational resource for massive calculation is deployed from CC-IN2P3 (Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules).

13.3 Standpoint of the Work

Besides the concrete strategy of *in vivo* embryo measuring experiment and 4-D reconstruction, the project has been animated by a larger conceptual framework of P. Bourgine. His global strategy toward the understanding of embryogenesis with complex systems perspective is depicted in Fig. 13.1.

The 4-D image data obtained from *in vivo* measurement is called "raw data", and need further image processing such as detection of each cell, segmentation of cell membrane, nuclei, and other biological features. The acquisition of raw data and its phenomenological reconstruction is a major derivative of Embryomics project, as well as in BioEMERGENCES with drug treatment. The reconstructed simulator of the measured embryogenesis allows us to experience the "augmented phenomenology", where one can interactively investigate morphogenetic process in multiscale 4-D dynamics with controllable viewpoint.

The phenomenological reconstruction has a high potential as an empirical database as well as an educational interface. However, it is not sufficient for scientific understanding of embryogenesis and to simulate drug response within: it lacks in the mechanism of the morphogenetic process. To elucidate the underlying mechanism of embryogenesis, we further need a "theoretical reconstruction", which is basically an inverse problem from the recorded data, with the aid of necessary hypotheses, theoretical methods, and tools.

The work presented in the following chapters correspond to this stage. The theoretical reconstruction requires the detection of effective physical and chemical laws driving morphogenetic process, to find plausible applications of mechanical models that are consistent to the phenomenology. For this purpose, the analysis with statistical models in view of reconstruction of underlying dynamical systems is an important issue.

The projects have so far not attained the complete phenomenological reconstruction of a whole embryo, nor the candidate model of theoretical reconstruction. Still, it is important to propose possible observables on the partial simulators to assume global strategy of theoretical reconstruction. The phenomenological reconstruction should be compatible to the scientific interests, where one is required to abstract biological indices necessary for theoretical reconstruction.

Theoretical reconstruction derives simulations, which are expected to reproduce the multi-scale dynamics of embryogenesis in a way compatible to phenomenological model. The plausibility of the theoretical reconstruction is then tested by experimental validation. The future model that would satisfy this validation is expected to bring us the "augmented virtuality" of embryogenesis, where one can not only understand the micro-scale mechanism of the dynamics but can simulate the generative process of the multi-scale cell coordination as a vivid sequence in living organism.

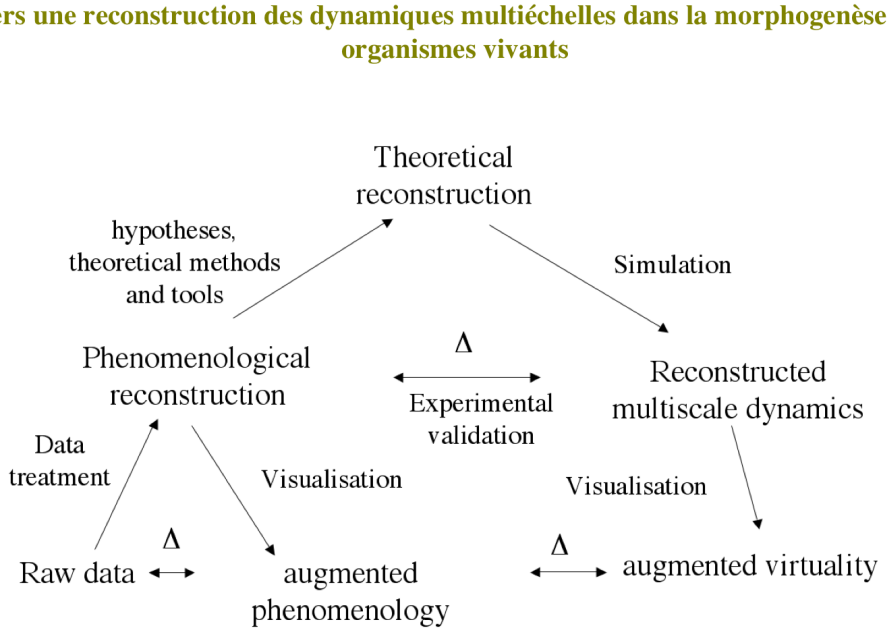


Figure 13.1: **Reconstruction strategy of embryogenesis by Paul Bourgine.** The works of the following chapters are situated at the arrow from "phenomenological reconstruction" to "theoretical reconstruction".

Chapter 14

Toward a Dynamical Definition of Tissue: Detection of Significant Correlation in Cell Movement

Abstract

We define the concept of "dynamical tissue" using the dynamical parameters of cellular resolution obtained from Embryomics and BioEMERGENCES projects, in contrast to traditional, anatomical and morphology-based definition of "static tissue". We construct a theory of statistical testing using information geometry to detect the significant correlation in cell movement vectors and its non-uniformity. Tentative results with zebrafish embryo data is shown, with the possible interpretation of deterministic constraint which defines the way of cell migration coordination and provides information to further theoretical reconstruction. Theoretical development for the measurement of the chronical sustainability of the proposed dynamical tissue is also investigated.

Keywords: Static tissue, Dynamical tissue, Cell movement, Statistical testing

Methodology: Definition of dynamical tissue with information geometry based on experimental data as dynamical systems → Analysis of experimental data → Further theoretical consideration on temporal duration of dynamical tissue with dynamical system perspective

14.1 Introduction

Bioimaging space-time resolution in Embryomics and BioEMERGENCES projects has a possibility to change the paradigm of the measurement in biology, and would lead to an empirical dynamical description of developmental process. One of remarkable things is the measurement of dynamical parameters such as cell movement vector field. To classify the types of cells, traditional biology has always been depended on morphological parameters such as cell form, cell size, gene expression, etc. These parameters can be regarded as the "static" parameters since they do not consider the "dynamical" ones arising directly from physio-chemical interactions between cells.

The dynamical parameters are closely related to the mechanical laws working among cells, and are part of important epigenetic factors. Although the mechanical constraint apparently has a substantial effect on tissue formation in developmental process, little is known on the hidden loop between mechanics and genetics. For example, the lineage tree of cell differentiation is not completely deterministic in vertebrates, and tolerate certain spatial shuffling of cells. This implies the existence of the coupling between the genetic expression and mechanical factors.

To tackle to the mechanical side of the morphogenesis, we raise the following questions:

1. What is the coordination of cell movement in developmental process ?

2. What kind of movement can we call "it is significant" ?

For that purpose, we introduce a measure of statistical testing so that to give a quantitative characterization of normal development in zebrafish embryogenesis. Besides, we establish a new concept of tissue, namely "dynamical tissue", based on the dynamical parameters such as vector fields of cell movement. This is the mechanical counterpart of the classical, anatomy-based definition of tissue, which we call as "static tissue".

14.2 Detection of Significant Correlation in Cell Movement

We develop a testing theory with the application of the information geometry so that to detect significant correlation and non-uniformity of cell movement.

14.2.1 Model Description

We established a statistical model of cell movement compatible to statistical testing. Since the inter-cellular mechanical interaction such as osmotic pressure is assumed to be much stronger than external field such as gravity, we need to set a model from inside of the cell movement dynamics without global reference. For that purpose, we formalize the cell movement with an internal measurement perspective. We consider the detection of correlation between neighboring cells movement with respect to the marginal distribution of the movement.

Let us consider the neighboring N_c cells with their movement vector field. Fig. 14.1 schematically shows the definition of the model. We consider to compare each pair of cells in the neighbor, by defining a cell as "cell 1" and the other cell as "cell 2". This means to take all ${}_{N_c}P_2$ permutation pairs as sample data of the model. We set a radius threshold r of vector field norm of cell movement, which will be defined in later section so that to maximize correlated component. If the movement vector norm of a cell is less than r, the movement vector is symbolized with the direction d_0, meaning noise. If the movement vector norm of the cell 1 is larger than r, we symbolize it with the direction d_1. As for the cell 2, if the movement vector norm is larger than r and the difference of the deflection angle between cell 1 and cell 2 with respect to the polar coordinates is less than the angle threshold $\phi = \pi/4$, the cell 2 data is symbolized with d_1. We call the area within the cone with center angle 2ϕ around the cell 1's movement direction as "syn cone", with which we define the synchronization of 2 cells' movement direction.

Then the ${}_{N_c}P_2$ sets of data can be expressed as a joint distribution $p(x_1, x_2)$ which is a statistical model with discrete variables taking $\{d_0, d_1\}$ for cell 1, and $\{d_0, d_1, d_2\}$ for cell 2:

$$\begin{aligned}
\log p(x_1, x_2) &= \log \frac{p_{d_1 d_0}}{p_{d_0 d_0}} \delta_{x_1}^{d_1} \delta_{x_2}^{d_0} && (14.1)\\
&+ \sum_{u_2 \in \{d_1, d_2\}} \log \frac{p_{d_0 u_2}}{p_{d_0 d_0}} \delta_{x_1}^{d_0} \delta_{x_2}^{u_2} && (14.2)\\
&+ \sum_{u_2 \in \{d_1, d_2\}} \log \frac{p_{d_1 u_2} p_{d_0 d_0}}{p_{d_1 d_0} p_{d_0 u_2}} \delta_{x_1}^{d_1} \delta_{x_2}^{u_2} && (14.3)\\
&- (-\log p_{d_0 d_0}). && (14.4)
\end{aligned}$$

Where x_1 and x_2 represent the symbolized vector data, and the parameters $p_{u_1 u_2}$ ($u_1 \in \{d_0, d_1\}$, $u_2 \in \{d_0, d_1, d_2\}$) are the occurrence probability of the symbolized direction pair u_1 and u_2 for cell 1 and cell 2, respectively.

The following η and θ parameters with corresponding potentials Ψ and Φ form the dual-flat coordinates of information geometry with respect to the Fisher information matrix:

$$\eta_{01} = p_{d_0 d_1} + p_{d_1 d_1}, \tag{14.5}$$
$$\eta_{02} = p_{d_0 d_2} + p_{d_1 d_2}, \tag{14.6}$$
$$\eta_{10} = p_{d_1 d_0} + p_{d_1 d_1} + p_{d_1 d_2}, \tag{14.7}$$
$$\eta_{11} = p_{d_1 d_1}, \tag{14.8}$$
$$\eta_{12} = p_{d_1 d_2}, \tag{14.9}$$
$$\Psi = -\log p_{d_0 d_0}, \tag{14.10}$$

and

$$\theta_{01} = \log \frac{p_{d_0 d_1}}{p_{d_0 d_0}}, \tag{14.11}$$
$$\theta_{02} = \log \frac{p_{d_0 d_2}}{p_{d_0 d_0}}, \tag{14.12}$$
$$\theta_{10} = \log \frac{p_{d_1 d_0}}{p_{d_0 d_0}}, \tag{14.13}$$
$$\theta_{11} = \log \frac{p_{d_1 d_1} p_{d_0 d_0}}{p_{d_1 d_0} p_{d_0 d_1}}, \tag{14.14}$$
$$\theta_{12} = \log \frac{p_{d_1 d_2} p_{d_0 d_0}}{p_{d_1 d_0} p_{d_0 d_2}}, \tag{14.15}$$
$$\Phi = \theta_{01}\eta_{01} + \theta_{10}\eta_{10} + \theta_{11}\eta_{11} + \theta_{12}\eta_{12} - \Psi. \tag{14.16}$$

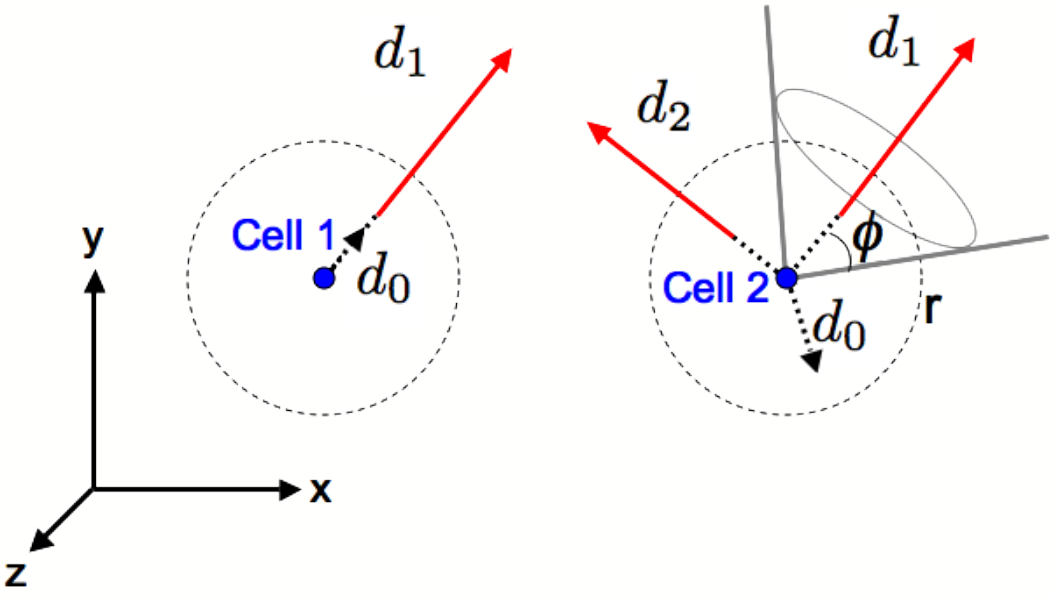


Figure 14.1: **Model definition.** Each pair of cell movement vectors is symbolized with discrete variables d_0, d_1, d_2 with respect to the radius threshold r and syn cone center angle $2\phi = \frac{\pi}{2}$.

14.2.2 Statistical Testing

Based on the model $p(x_1, x_2)$, we develop a statistical testing theory with the use of information geometry to detect the following dynamics in cell movement:

- Correlation in the same direction with respect to the syn cone definition.
- Non-uniformity of cell movement deflection angle with respect to the polar coordinates.

For that purpose, we formalize the following null hypotheses on the statistical manifold S of $p(x_1, x_2)$:

- Null Hypothesis 1: The cells moves independently following its marginal distribution η_{10}.
- Null Hypothesis 0: The distribution of cell movement deflection angle follows uniform distribution with respect to the polar coordinates.

Fig. 14.2 shows the geometrical relation between these null hypotheses. We first calculate the coordinates η^q of the following uniform distribution $q(x_1, x_2)$ from data distribution $p(x_1, x_2)$:

$$\eta^q_{01} = (\eta_{01} + \eta_{02})SR, \tag{14.17}$$
$$\eta^q_{02} = (\eta_{01} + \eta_{02})(1 - SR), \tag{14.18}$$
$$\eta^q_{10} = \eta_{10}, \tag{14.19}$$
$$\eta^q_{11} = \eta_{11}, \tag{14.20}$$
$$\eta^q_{12} = \eta_{12}. \tag{14.21}$$

Where SR is the surface ratio of the syn cone on the unit sphere. The solid angle (steradian, [sr]) of the cone with center angle 2ϕ (radian) in the unit sphere is calculated as $2\pi\left(1 - cos\left(\frac{2\phi}{2}\right)\right)$, and is equivalent to the sphere surface isolated by the cone. Since the solid angle of an unit sphere (center angle 2π) is $2\pi\left(1 - cos\left(\frac{2\pi}{2}\right)\right) = 4\pi$ which coincides with the total surface, the SR can be calculated as the ratio of solid angles between the syn cone and total sphere as follows:

$$\begin{aligned} SR &= \frac{\text{Syn cone solid angle [sr]}}{\text{Unit sphere solid angle [sr]}} \\ &= 2\pi\left(1 - cos\left(\frac{2\phi}{2}\right)\right) \cdot \frac{1}{4\pi} \\ &= 0.1464. \end{aligned} \tag{14.22}$$

We then suppose the independence between cells in $q(x_1, x_2)$, and represents the joint distribution with $q'(x_1, x_2)$ which corresponds to the independent uniform distribution with the same marginal distribution η_{10} as data $p(x_1, x_2)$. The coordinates $\eta^{q'}$ of the distribution $q'(x_1, x_2)$ become as follows:

$$\eta^{q'}_{01} = (\eta_{01} + \eta_{02})SR, \tag{14.23}$$
$$\eta^{q'}_{02} = (\eta_{01} + \eta_{02})(1 - SR), \tag{14.24}$$
$$\eta^{q'}_{10} = \eta_{10}, \tag{14.25}$$
$$\eta^{q'}_{11} = \eta_{10}(\eta_{01} + \eta_{02})SR, \tag{14.26}$$
$$\eta^{q'}_{12} = \eta_{10}(\eta_{01} + \eta_{02})(1 - SR). \tag{14.27}$$

To decompose orthogonally the effect of correlation from the non-uniformity of cell deflection angle, we consider the other dual coordinates $\theta' = (\theta'_{01}, \theta'_{02}, \theta'_{10}, \theta'_{11}, \theta'_{12})$ of $\eta^{q'}$ in the same way as the definition (14.5)-(14.16). The independent uniform distribution $q'(x_1, x_2)$ is then expressed as follows with the mixture coordinates $\zeta^{q'}$:

$$\zeta^{q'} = (\zeta^{q'}_{01}, \zeta^{q'}_{02}, \zeta^{q'}_{10}, \zeta^{q'}_{11}, \zeta^{q'}_{12}), \tag{14.28}$$
$$\zeta^{q'}_{01} = \eta^{q'}_{01}, \tag{14.29}$$
$$\zeta^{q'}_{02} = \eta^{q'}_{02}, \tag{14.30}$$
$$\zeta^{q'}_{10} = \eta^{q'}_{10}, \tag{14.31}$$
$$\zeta^{q'}_{11} = \theta'_{11}, \tag{14.32}$$
$$\zeta^{q'}_{12} = \theta'_{12}. \tag{14.33}$$

The $q'(x_1, x_2)$ on the mixture coordinates is the projection point of the null hypothesis 0. The projection origin $p'(x_1, x_2)$ of the null hypothesis 0 coincides with the projection point of null hypothesis 1 from data distribution $p(x_1, x_2)$, which is defined as follows with the mixture coordinates $\zeta^{p'}$:

$$\zeta^{p'} = (\zeta^{p'}_{01}, \zeta^{p'}_{02}, \zeta^{p'}_{10}, \zeta^{p'}_{11}, \zeta^{p'}_{12}), \tag{14.34}$$
$$\zeta^{p'}_{01} = \eta_{01}, \tag{14.35}$$
$$\zeta^{p'}_{02} = \eta_{02}, \tag{14.36}$$
$$\zeta^{p'}_{10} = \eta_{10}, \tag{14.37}$$
$$\zeta^{p'}_{11} = \theta'_{11}, \tag{14.38}$$
$$\zeta^{p'}_{12} = \theta'_{12}. \tag{14.39}$$

The distribution $p'(x_1, x_2)$ has the same marginal distribution as the data $p(x_1, x_2)$ but do not have correlation between each pair of cells. The distribution $q'(x_1, x_2)$ does not have either the non-uniformity of deflection angle, nor the correlation. The mixture coordinates separate the marginal distributions and correlation components orthogonally with respect to the Fisher information metric, which realizes independent testing of the 2 null hypotheses.

The KL divergence $D[\cdot : \cdot]$ between data P and projected null hypothesis Q of discrete distribution is known to be compatible to chi-squared test. The chi-squared value is given by

$$2ND[P : Q] = \sum \log P \frac{\log P}{\log Q} \sim \chi^2(\dim S - \dim S_{n.h.}) \tag{14.40}$$

Where N is the sample number, $\dim S$ is the dimension of statistical manifold, and $\dim S_{n.h.}$ is that of submanifold defined by the null hypothesis.

Then the null hypotheses 1 and 0 can be tested with chi-squared test by calculating the following values λ_1 and λ_0 from corresponding KL divergences:

$$\lambda_1 = 2_{N_c} P_2 D[p(x_1, x_2) : p'(x_1, x_2)] \sim \chi^2(2), \tag{14.41}$$
$$\lambda_0 = 2_{N_c} P_2 D[p'(x_1, x_2) : q'(x_1, x_2)] \sim \chi^2(4). \tag{14.42}$$

Where λ_1 asymptotically follows the chi-squared test with degrees of freedom 2, while λ_0 is that of degrees of freedom 4.

The significant correlation of the null hypothesis 1 can further be judged whether it is positive or negative correlation with the following positive/negative sign of s:

$$s = \mathrm{sgn}(\eta^{\mathrm{p}'}_{11} - \eta_{11}). \tag{14.43}$$

In case of the null hypothesis 0, the positive/negative signs become 2-dimensional and we do not have a simple word to describe the non-uniformity in a binary way.

In our model, the radius threshold r for the symbolization of vector norm in each cell can be optimized by maximizing the KL divergence of the null hypothesis 1 as follows:

$$r = \underset{r_c}{\mathrm{argmax}}(D[p(x_1, x_2; r_c) : p'(x_1, x_2; r_c)]). \tag{14.44}$$

Where $p(x_1, x_2; r_c)$ and $p'(x_1, x_2; r_c)$ is the model and its null hypothesis 1 with radius threshold r_c for the symbolization of d_0. This mean that if the r value is too large, it risks to lose the information of correlated components by judging them as noise (d_0). While if the r value is too small, the correlated components can be blurred by actual noise. This method automatically detects the threshold which gives the finest distinction between noise and correlated components under information maximization principle.

Schematic examples of the relation between the null hypotheses and the cell movement distribution is depicted in Fig. 14.3. The null hypothesis 0 assumes the uniform distribution of cell

deflection angle including noise with vector norm under threshold r. While the null hypothesis 1 measure the degree of correlation on generally non-uniform distribution of cell deflection angle. Note that this does not depend on the uniform distribution assumption. The testing is valid on any non-uniform distribution since the non-uniformity and correlated components are independent parameters separated orthogonally with the use of mixture coordinates.

This theory is valid not only between 2 cells, but also between 2 different levels of organism, such as a region of tissue and cells inside. For instance, the model can detect the correlation between tissue deformation and mitosis axis in the same setting. As long as we compare the local correlation and uniformity of vector fields, this theory is compatible to multi-scale analysis.

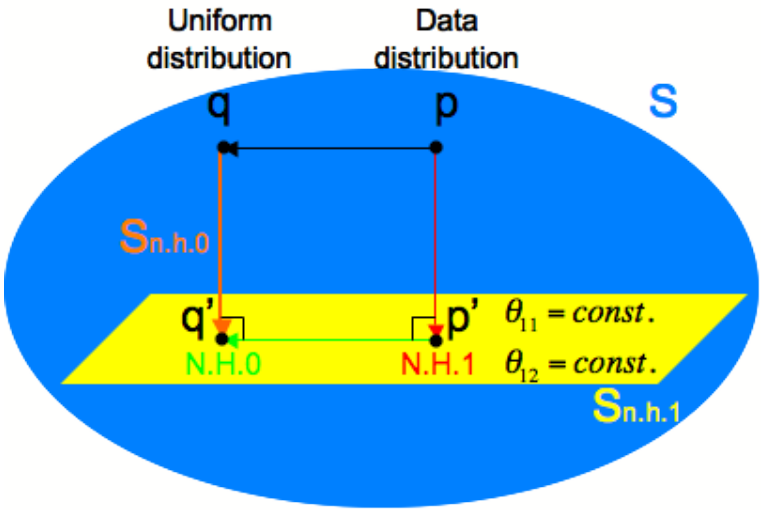


Figure 14.2: **Information geometrical structure of statistical testing for null hypotheses 1 and 0.** The submanifolds of the null hypotheses are defined by fixing corresponding η and θ parameters in the mixture coordinates. The manifold of the statistical model is described as S, while the submanifolds for the null hypotheses 1 and 0 are $S_{n.h.1}$ and $S_{n.h.0}$, respectively. The KL divergence $D[p : p']$ measures the significance with respect to the null hypothesis 1 and quantifies the correlated components in the syn cone. While $D[p' : q']$ represents the significance with respect to the null hypothesis 0 and expresses the non-uniformity of the cell deflection angle distribution including noise.

14.2.3 Test Data

We test the chi-squared values λ_1 and λ_0 to adjust the 5% significance threshold parameters with respect to the fluctuation arising from the small sample number N_c. The results are shown in Figs. 14.4. The random vector fields express less than 5 % significant cells with respect to both null hypotheses 1 and 0 (Figs. 14.4 Left column Middle and Bottom). This is the case in Figs. 14.3 Top Left. The mixture of the non-uniform small norm vectors and uniform flow with long norm shows dominant significance in both null hypotheses 1 and 0. This corresponds to the situation in Figs. 14.3 Bottom Right.

Since the model is based on the information maximization principle to define the noise radius threshold r, the results on the test data is directly compatible to other data with different quantitative variation of vector norm parameters. The algorithm automatically detects the scale of noise and take the threshold with the highest sensitivity to the correlated components. This is one of the benefits of the quantization of analog data taking a discrete distribution model, and is a strong point when investigating biological systems which we do not explicitly know about the spatio-temporal variation of noise.

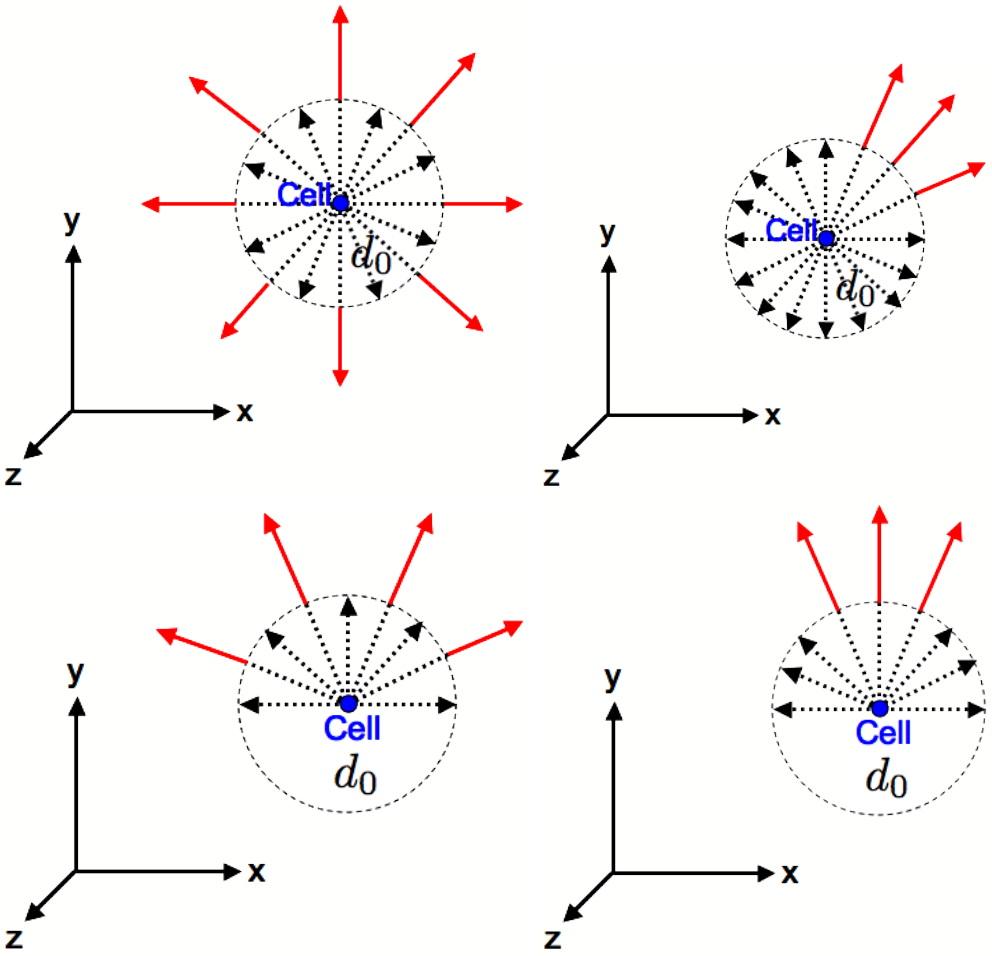


Figure 14.3: **Schematic examples of significant data with respect to the combination of null hypotheses 1 and 0.** All cell movement vectors are superimposed with the same origin, which is expressed as "cell" in each figure. Top Left: Both null hypothesis 1 and 0 can not be rejected. The deflection angle distribution of cell movement is uniform, including the vectors with norm over radius threshold r shown in red. Top Right: Null hypothesis 1 is rejected, while null hypothesis 0 is not. The non-noise red vectors are significantly localized with respect to the overall deflection angle including noise expressed with black dashed line vectors. Bottom Left: Null hypothesis 1 can not be rejected, while null hypothesis 0 is rejected. Cell deflection angle shows significant non-uniformity, though the non-noise red vectors are not localized with respect to the non-uniformity. Bottom Right: Both null hypothesis 1 and 0 are rejected. The cell deflection angle is non-uniform, and non-noise red vectors show further significant localization besides the non-uniformity

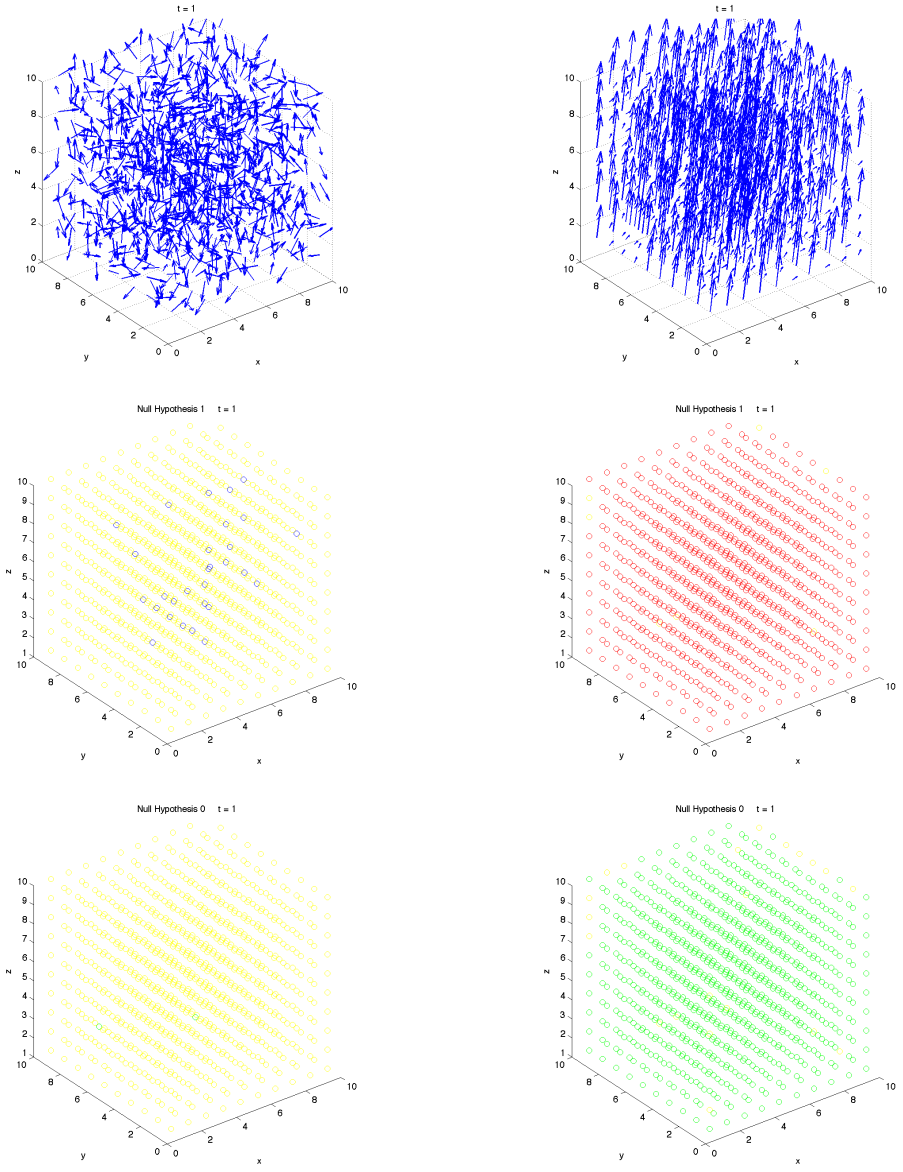


Figure 14.4: **Results of statistical testing on sample artificial vector field data.** Left column figures: Random vector field. Right column figures: Random mixture of small norm vectors with non-uniform distribution of deflection angle and large norm uniform flow. Top: Vector field visualization. Middle: Significance of the null hypothesis 1. Cells having significant correlated components with respect to the 10 neighboring cells are depicted with red (positive correlation) and blue circles (negative correlation). Yellow circles are not significant cells. Bottom: Significance of the null hypothesis 0. Cells following significant non-uniform distribution of cell movement deflection angle with respect to the 10 neighboring cells are depicted with green. Yellow circle represents cells which are not significant.

14.3 Tentative Results

We applied the above testing theory to the tentative data of cell movement vector field derived from the Embryomics project.

The vector field of cell movement was calculated according to the optical flow of image intensity using Thirion's demon algorithm [108] [1], and was averaged over 20 steps of tracking on the reconstructed cell lineage tree. Since the time interval of the observation is about 67 seconds, the vector field represents the average movement of cells during 22 minutes and 20 seconds.

The testing results are shown in Figs. 14.5 and 14.6 with respect to the null hypothesis 1 and 0, respectively.

The animal pole of zebrafish embryo shows certain global patterns of significant cells with respect to both null hypotheses. Such localization of cell movement correlation is a candidate of "dynamical tissue", which is an important feature to reveal the mechanism of morphogenesis.

The correlation with respect to the null hypothesis 1 tends to show significance first which is followed by the non-uniformity with respect to the null hypothesis 0. This means that the local synchronization of cell movement occurs in a bottom-up manner, expanding from between individual cells to larger population. This fits well to the superiority of inter-cellular mechanical interaction than the effect of external or global field.

The characteristic band pattern traversing the animal pole during the time step 201-421 in Figs. 14.5 and 14.6 represents the cell flow at the beginning of the gastrulation process.

Such dynamical constraint is a way to define the mechanical aspect of embryogenesis from a dynamical system's point of view. The detected deterministic property is essential to choose the candidate model for further theoretical reconstruction.

The evaluation of the phenomenological model with the use of dynamical tissue concept is also important in industrial application. The perturbation of the dynamical tissue with respect to the external drug treatment is essential to evaluate its epigenetic effect in terms of the disturbance in morphogenetic process.

14.4 Scarcity of Correlation Duration

We further consider the utility of the proposed statistical model on a theoretical basis. We consider how consistent the dynamical correlation persists in the same cell lineage. The following theory is inspired by the symbolic dynamics analysis of chaotic dynamical system.

The proposed dynamical tissue is based on a synchronic relation between dynamical parameters. We can further extend this definition to the chronological consistency with the use of the cell lineage tree. Suppose that the statistical significance of a cell which defines the existence of a dynamical tissue with respect to the neighboring cells is symbolized with $s(t)$ with a given time t. For example, the null hypothesis 1 is judged within 3 symbols as $s(t) \in \{+, -, 0\}$, which corresponds to the distinction of positive/negative/no significance, respectively.

We question the temporal duration of the dynamical tissue symbolized with this significance. We measure how long the series of $s(t)$ in a cell lineage follows the same symbol. For that purpose, we set the null hypothesis 2 as follows:

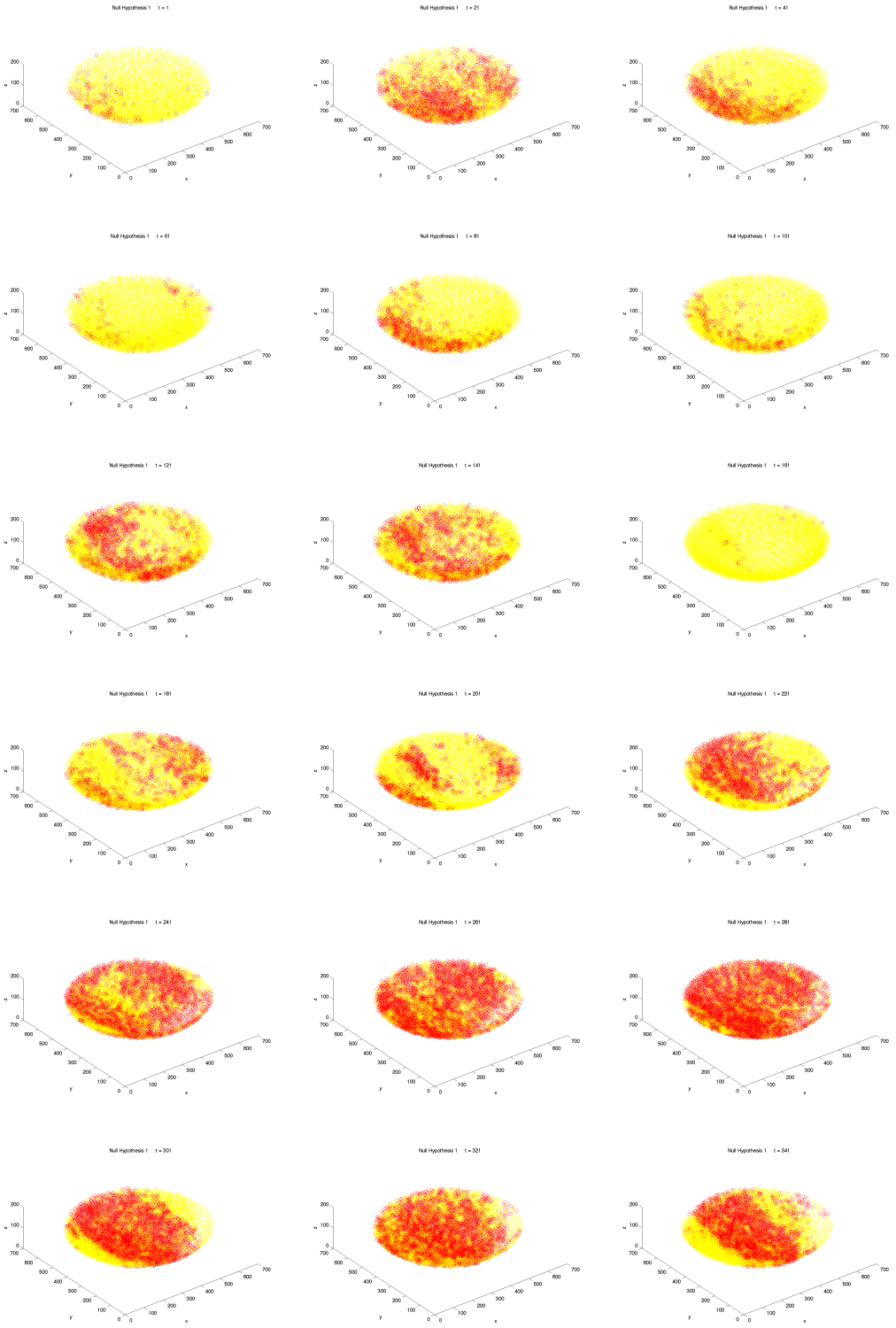
- Null hypothesis 2: The sequence $\{s(t)\}_{t=1}^{m}$ of a given cell lineage is an independent stochastic process whose appearance rate is that of the marginal distribution of 3 symbols $\{+, -, 0\}$ in the whole embryo cells at time t.

If we express the average appearance rate of the symbols $s(t) \in \{+, -, 0\}$ in all cells of an embryo at time t as $P_t(+)$, $P_t(-)$, and $P_t(0)$, respectively, then the appearance rate AR of a m-step sequence $\{s(t)\}_{t=1}^{m}$ under the null hypothesis 2 becomes as follows:

$$AR = P_1(s(1)) \cdot P_2(s(2)) \cdots P_m(s(m)). \tag{14.45}$$

This constitutes a term-wise component in trinomial theorem when considering the total AR sum $\prod_{t=1}^{m}(P_t(+) + P_t(-) + P_t(0))$ of all possible sequences.

[1] The instantaneous cell velocity was calculated by Benoit Lombardot in his Ph.D dissertation.



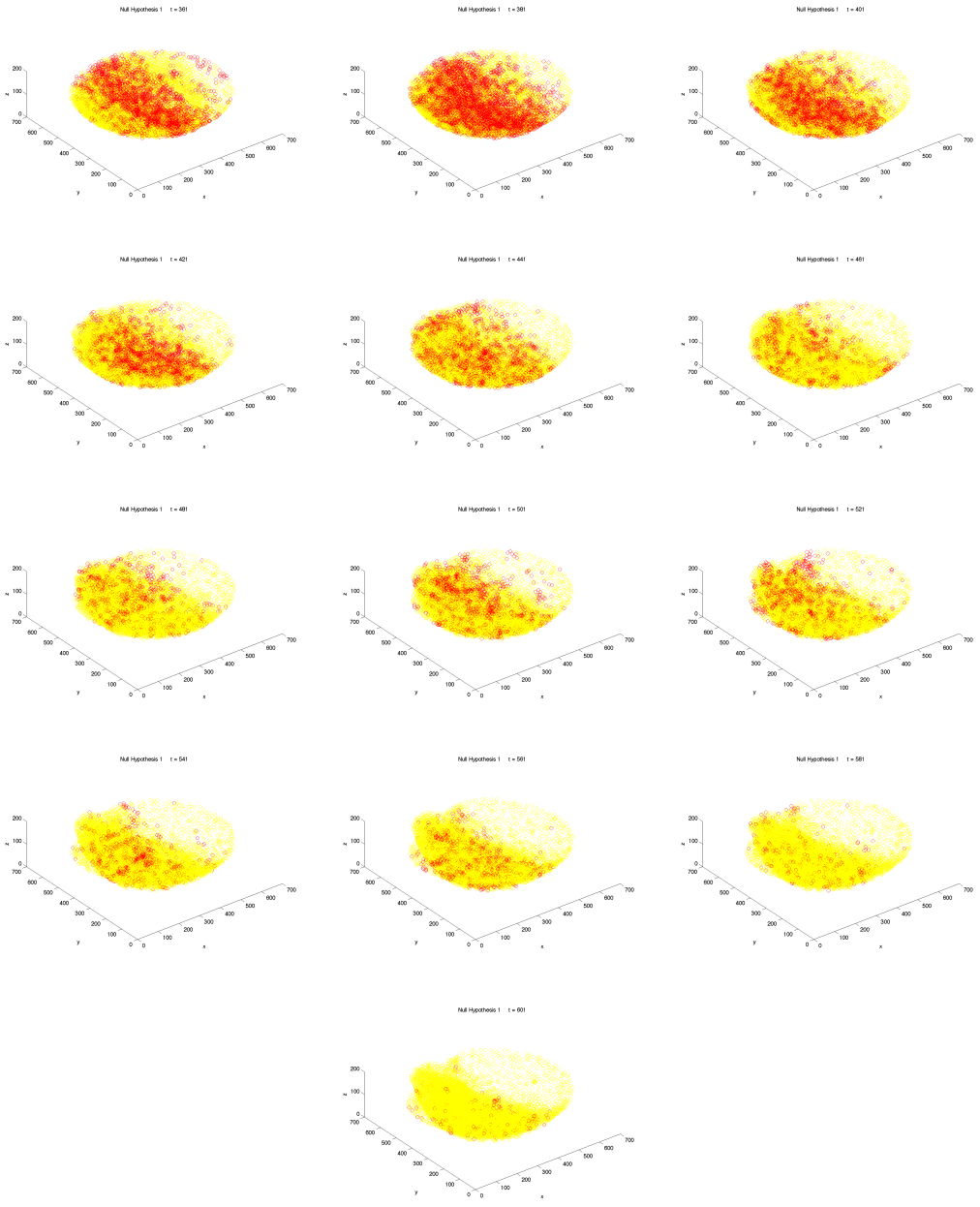
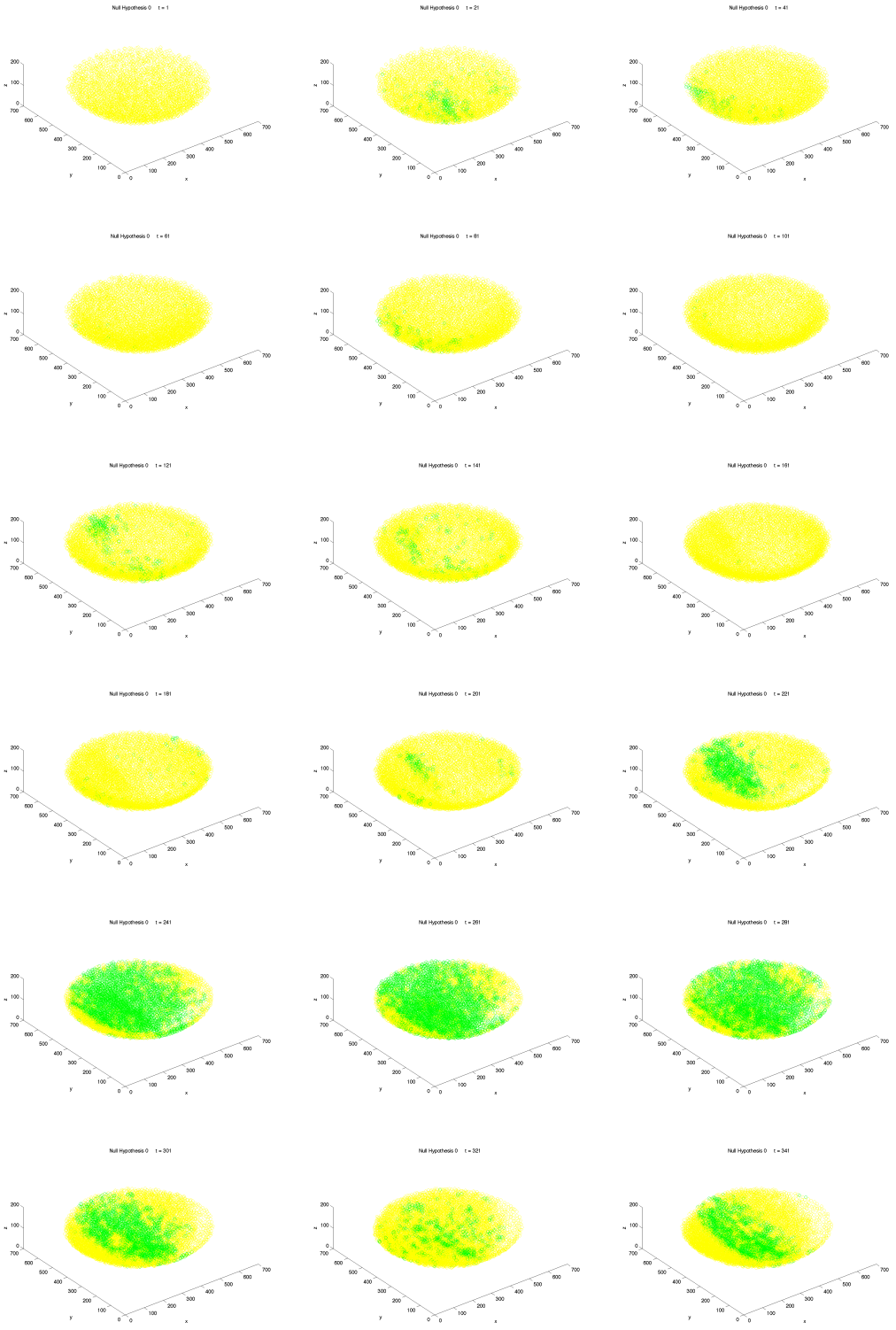


Figure 14.5: **Testing results for null hypothesis 1 on tentative vector fields data derived from Embryomics project.** Cells with significant positive correlation in movement vectors with respect to the 10 neighboring cells are depicted with red circles, while not significant cells are shown with yellow circles. No negative significant correlation was observed. Observations from the 1st to 601st steps with 20 steps interval are aligned from top left to bottom right. Data id: 070418a. Time interval $\triangle t$ between each observation: 67 seconds. Animal pole of zebrafish (danio rerio).



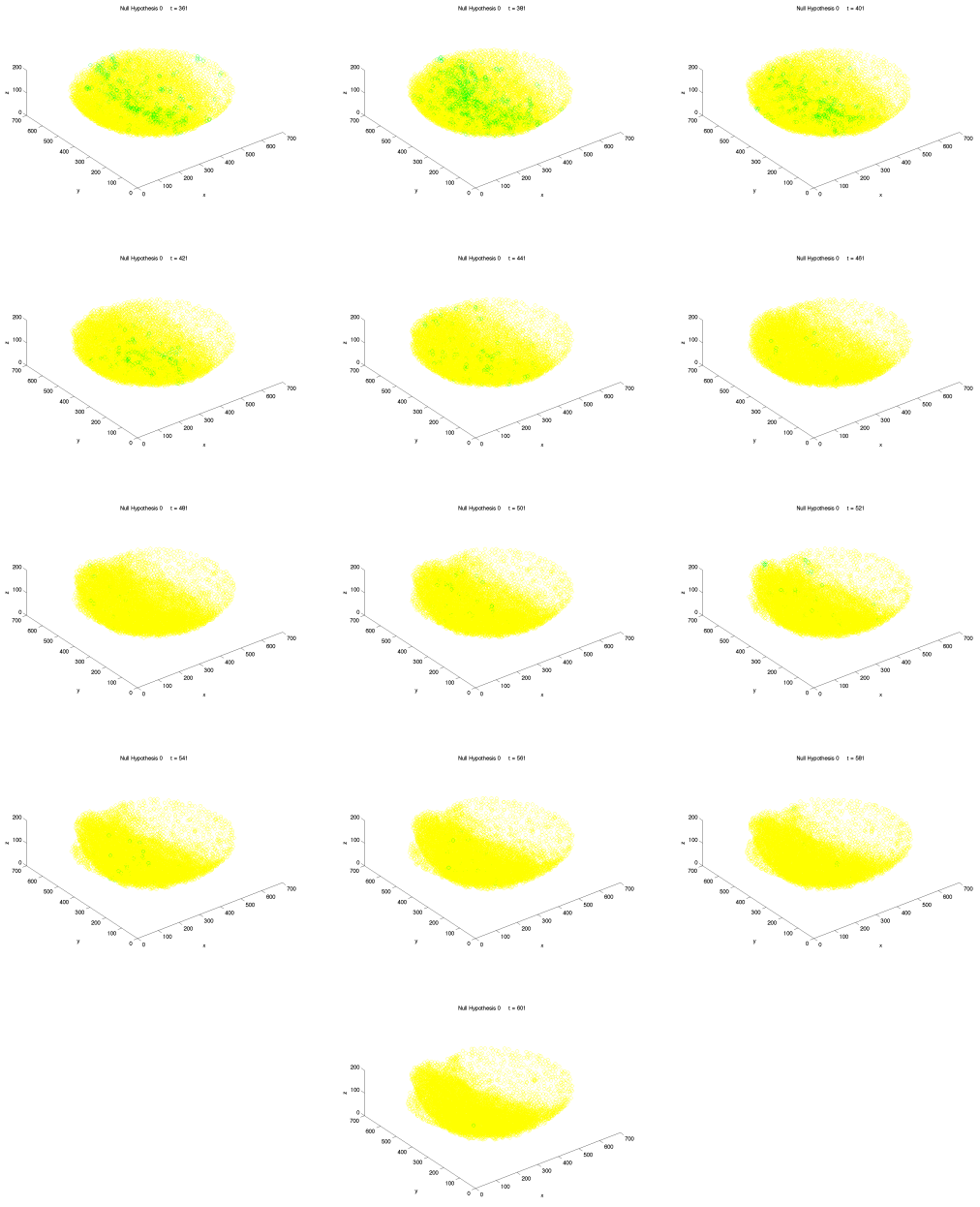


Figure 14.6: **Testing results for null hypothesis 0 on tentative vector fields data derived from Embryomics project.** Cells with significant non-uniformity of movement vector deflection angle are depicted with green circle, while not significant cells are shown with yellow circles. Observations from the 1st to 601st steps with 20 steps interval are aligned from top left to bottom right. The data are equivalent to Figs. 14.5.

The number of different m-step sequences $\{s(1)s(2)\cdots s(m)|s(t) \in \{+,-,0\}\}$ is 3^m, which is equivalent to the number of m-digit ternary number. The combination of the possible sequences are depicted in Fig. 14.7 in relation to the trinomial theorem.

Since there exists 3^m different sequences, the mean value $E[AR]$ of the appearance rate AR of a sequence is given by

$$E[AR] = \frac{1}{3^m}\prod_{t=1}^{m}(P_t(+) + P_t(-) + P_t(0)) = \frac{1}{3^m}, \tag{14.46}$$

since $P_t(+) + P_t(-) + P_t(0) = 1$ always holds for any t. Whatever the variation of $\{P_t(s(t))|s(t) \in \{+,-,0\}\}$ is, the $E[AR]$ value only depends on the number of the symbols coding the type of significance. While the appearance rate AR of each sequence depends on the value of $P_t(s(t))$ at each t.

We consider how rare a given sequence is compared to the mean appearance rate $E[AR]$, to measure the scarcity of the correlation duration. For that purpose, we measure the appearance rate AR in logarithmic scale since it is based on a multiplicative sequence, and renormalize with the mean value $E[AR]$ so that the mean appearance rate under null hypothesis 2 is represented with 0. The scarcity of correlation duration SCD is then expressed as follows:

$$SCD = \log\frac{AR}{E[AR]} \tag{14.47}$$

$$= \log 3^m \prod_{t=1}^{m} P_t(s(t)). \tag{14.48}$$

This SCD value becomes positive if the $P_t s(t)$ are more than $1/3$ in geometric average, and negative if less. If we only measure the scarcity of the sequence and do not consider the geometric average distribution of $P_t(s(t))$ of a given sequence, the SCD is defined as follows:

$$SCD = \left|\log 3^m \prod_{t=1}^{m} P_t(s(t))\right|. \tag{14.49}$$

Where $|\cdot|$ represents the absolute value. We take the latter definition as the formal SCD.

In general case where the number of symbols is represented with $\#(\{s\})$, the SCD during m steps of tracking becomes as follows:

$$SCD = \left|\log \#(\{s\})^m \prod_{t=1}^{m} P_t(s(t))\right|. \tag{14.50}$$

The simple deviation of SCD from its mean value 0, however, can not be taken directly as the degree of scarcity. Since the distribution of $\{P_t(s(t))|s(t) \in \{+,-,0\}\}$ is dependent to the time t, we need to measure the fluctuation from the mean value under null hypothesis 2 to construct a significance criterion. The 5 % significance threshold $\mu(m)$ of the m-step SCD with respect to the null hypothesis 2 is given in the following relation:

$$\int_0^{\mu(m)} P(SCD)dSCD = 0.95. \tag{14.51}$$

Where $P(SCD)$ is the density function of the SCD values of all possible m-step sequences $\{s(1)s(2)\cdots s(m)\}$.

If the SCD value is superior to $\mu(m)$ value, the sequence has 5 % significance with respect to the null hypothesis 2. This means that the consistency of the correlation duration is far from

a random stochastic process, and the corresponding dynamical tissue has a long-term persistence which belongs to a deterministic property.

The actual calculation of SCD value and the testing requires the accurate tracking of cell lineage tree during a long period. Large value of m is required to suppress the fluctuation of SCD around its mean value. Large m, however, contradicts with the accuracy of tracking. Even with 99 % accuracy of tracking, after $m = 100$ steps there remains only 36% of accuracy in each cell lineage. The multiplicative accumulation of tracking error is the major obstacle for this theory, but would be overcome by the obtention of the complete cell lineage tree in the near future.

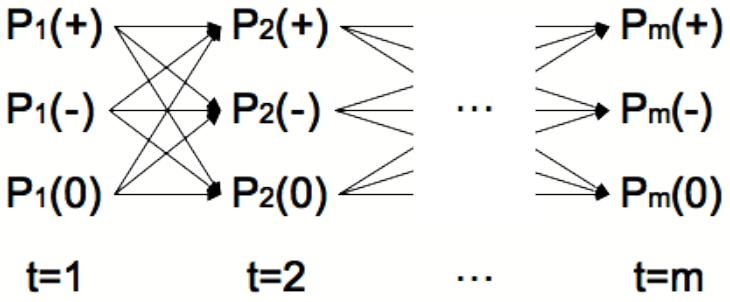


Figure 14.7: **Possible sequences of $\{s(1)s(2)\cdots s(m)|s(t) \in \{+,-,0\}\}$ in relation to trinomial theorem giving AR value for each sequence.** The mean appearance rate E[AR] is always $\frac{1}{3^m}$ in case $\#(\{s\}) = 3$ regardless of the temporal fluctuation of $P_t(s(t))$.

14.5 Conclusion

We defined the concept of "dynamical tissue" on the basis of dynamical parameters derived from the empirical measuring of embryogenesis in Embryomics and BioEMERGENCES projects. Tentative results are shown which detected significant correlation of cell movement vectors and non-uniformity of cell deflection angle, implying the bottom-up way of cell migration coordination during the gastrulation process of zebrafish. The theory was further extended to detect the significance in the temporal duration of the dynamical tissue.

Chapter 15

Toward a Multi-Scale Characterization of Embryogenesis: Incorporating Biological Variables in Semi-Parametric Clustering

Abstract

We consider the strategy to link the phenomenological and theoretical reconstruction by investigating the phenomenological model with biological variables. We propose a clustering method based on the "minimum descriptors" which incorporate specific biological interest to characterize the dynamics of embryogenesis. Tentative results are shown with theoretical strategy to measure and represent individual variations with various spatio-temporal scales in contrast to external drug treatment.

Keywords: Minimum descriptor, EM-Algorithm, Individual variation

Methodology: Definition of minimum descriptors to characterize primary data set as huge complex dynamical systems → Clustering with minimum descriptors in a way equivalent to the em-algorithm of information geometry → Theoretical proposition to measure individual variation of embryos based on information theoretical distance and dynamical system perspective

15.1 Introduction: Strategy for the multi-scale characterization of embryogenesis

The high spatiotemporal resolution of 4-D embryo measurement in Embryomics and BioEMERGENCES projects has a potential to revolutionize the state of art of embryology. However, empirical measurement does not directly lead to scientific significance without theoretical perspectives. What we can obtain as raw data from an electrical microscopy is nothing but a complex image without scientific characterization. As Albert Einstein said, theoretical framework precedes the observation and define what we can observe [109].

Fig. 15.1 shows the strategy of the phenomenological reconstruction and a proposition toward the theoretical reconstruction. Phenomenological reconstruction based on image processing is the primary task for the understanding of raw data. Embryomics project was mainly devoted to detecting important biological parameters to establish a complete cell lineage tree, such as the spatial structure of cell/nuclei membrane, occurrence of mitosis, cell movement, cell tracking between different time steps of observation, etc. These basic biological variables, or simply "primary variables," are the parameters of the phenomenological reconstruction, and are what define the basis of mathematical space for further theoretical reconstruction.

The work in this chapter is a theoretical proposition to link the phenomenological and theoretical reconstruction. We investigate on the phenomenological data to seek for a theoretical formalization of underlying mechanism in embryogenesis. For that purpose, we consider the transformation of primary variables as "minimum descriptors" for efficient and meaningful clustering of cell data. The clustering based on the minimum descriptors aims to extract important information useful to quantify individual variations. This is at the same time parameter reduction and information compression which augment the facility of analysis and the model sensitivity to a given biological interest. The minimum descriptor itself can be interactively redefined based on the results of the former clustering and its visualization. We aim to establish an integrative strategy for multi-scale characterization of embryogenesis, gathering both theoretical and biological knowledge as an interactive methodology.

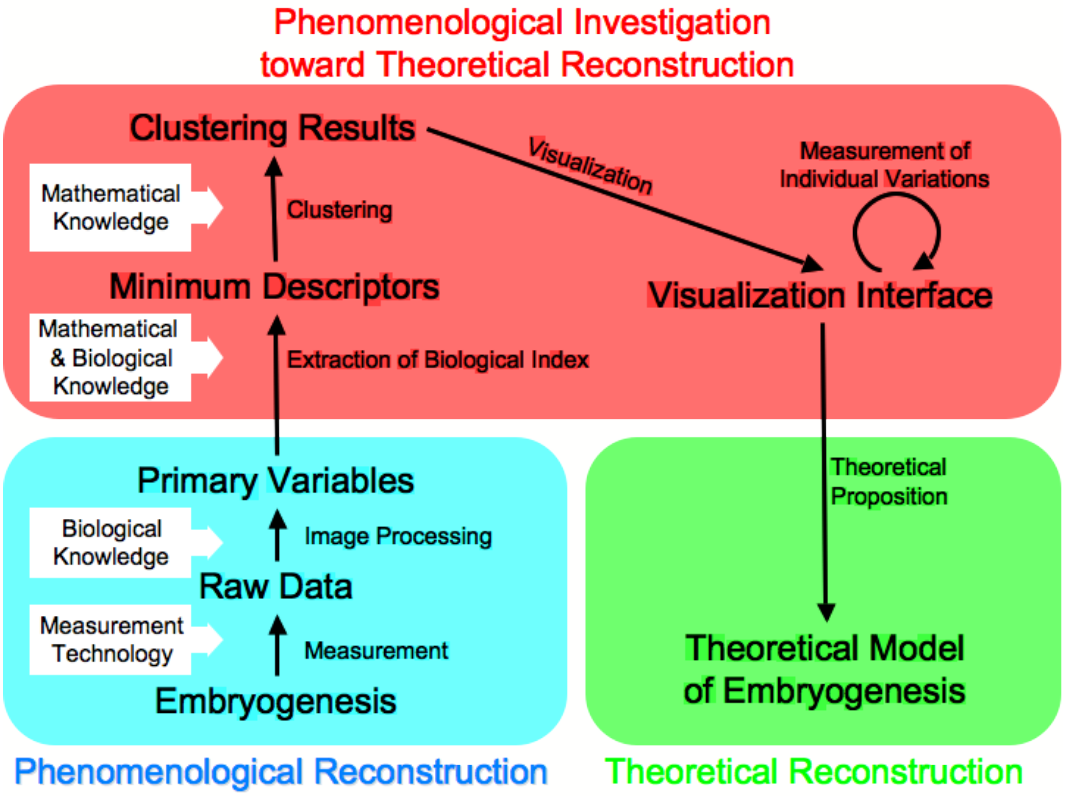


Figure 15.1: **Strategy of phenomenological investigation toward theoretical reconstruction.** The scope of the work in this chapter corresponds to the red area. Main effort in the european projects were effectively devoted to the phenomenological reconstruction, expressed as blue area.

15.2 EM-Clustering: Semi-parametric Clustering in Biologically Meaningful Space

15.3 Primary Variable Data Set

The raw data of zebrafish embryo normal development is obtained from the Embryomics project. We have imaged the prospective brain of zebrafish embryos by multiphoton microscopy from 4 hours post-fertilization through gastrulation and early neurulation stages 3. The imaged cells, whose number grows from 3,000 to more than 11,000, were tracked with less than 2% error for at least 6 hours. The tracking accuracy allows us to construct reliable lineage tree. The attempt to calculate morphological parameters of cell and nuclei membrane from raw image data is also a part

of the project. The primary variables were partially provided to test the clustering based on the minimum descriptors.

15.3.1 Minimum Descriptor

The reconstruction of early stage embryogenesis in Embryomics project established a huge data source of primary variables accessible to further characterization. The classification of cell groups based on biological features is one of the important task besides the reconstruction of lineage tree.

For that purpose, we have to established an automated way of clustering with consistent parameters. Such parameters are not necessary the primary variables itself, but generally a function of them reflecting biological knowledge to properly separate different cell group distributions. We call this biologically meaningful variables for the automated clustering as "minimum descriptor", since they compress the characteristic information of cells, and offer possibly a minimum sufficient mathematical space for cell group separation. Since it aims at offering a clear-cut aspect to see the difference between cells in low dimensional space, the meaning of "minimum" depends on the intended separability of the clustering.

For example, the following measures can be considered as minimum descriptors.

- Cell shape: The normalized product of PCA (principal component analysis) of cell membrane spatial distribution
- Instantaneous velocity of a cell: The norm of cell movement vector field.
- Global movement of a cell: The radius of long range cell movement.
- Proliferation rate: The time distance between succeeding mitosis.
- Cell density: The number of neighboring cells.

Although the minimum descriptors are based on the quantitative primary variables, they are not yet accessible in empirical time range of embryogenesis in the projects. Quantitative primary variables taking real value are difficult to detect with sufficient level of validation, due to the difficulty of establishing a control standard for the detection algorithm of image processing. On the other hand, qualitative primary variables such as cell center are easier to extract, and can be corrected by biologists with the use of an visual interface. For the sake of cell lineage tree reconstruction, qualitative primary variables have a priority in the projects. In this chapter, we are limited to test a partial data set to see the effectiveness of the clustering algorithm which is explained in the following section.

15.3.2 EM-Clustering

We consider the clustering of cells on the space defined with minimum descriptors. The form of cell distribution in the space of minimum descriptors is generally unknown. To fit the fitting function to an arbitrary multimodal distribution and separate each modal component, we choose a semiparametric clustering using Gaussian mixture distribution. The space of Gaussian mixture family is also compatible to information geometry, and can easily derive the distance between different individuals. We choose the EM-algorhithm to fit the function to data defined in minimum descriptor space. EM- algorithm is also compatible to information geometrical em-algorithm to establish its convergence [110]. This algorithm calculates the probability of each cell belonging to each Gaussian component. We also apply the AIC (Akaike's Information Criterion) to avoid overfitting in this algorithm. The AIC optimizes the number of Gaussian component for clustering in quite general assumption.

The general d-dimensional Gaussian mixture model is described as follows:

$$p(\mathbf{x}, j; \theta) = \sum_{j=1}^{J} \pi_j N(\mu_j, \mathbf{\Sigma}_j), \tag{15.1}$$

where $\mathbf{x} = (x_1, \cdots, x_d)$, J is the total number of normal distributions, whose j-th component is linearly connected with regularized weight π_j ($\sum_{j=1}^{J} \pi_j = 1$). $\theta = \{\{\mu_j\}, \{\mathbf{\Sigma}_j\}, \{\pi_j\}\}$ is the model parameters where μ_j is the mean values and $\mathbf{\Sigma}_j$ is the variance-covariance matrix of the j-th Gaussian component. Since the EM-algorithm adjusts these parameters to fit the data by the iteration of E-step and M-step, the parameters of distribution are conditioned by the iteration step k such as $\theta(k)$.

The E-step of EM-algorithm stands for the expectation process, where we calculate the probability $p(j|\mathbf{x}_n; \theta(k))$ that the element $\mathbf{x}_n$ of data $\{\mathbf{x}_1, \cdots, \mathbf{x}_n, \cdots, \mathbf{x}_N\}$ $(1 \leq n \leq N)$ was generated from the j-th distribution:

$$p(j|\mathbf{x}_n; \theta(k)) = \frac{N(\mathbf{x}_n; \mu_j(k), \mathbf{\Sigma}_j(k))\pi_j(k)}{\sum_{i=1}^{J} N(\mathbf{x}_n; \mu_i(k), \mathbf{\Sigma}_i(k))\pi_i(k)}. \tag{15.2}$$

This process corresponds to the maximum likelihood estimation of the hidden parameters $\{j\}$ with respect to the fixed parameters θ of the model.

The M-step is for the maximization process, where we maximize the logarithmic likelihood of the parameters by calculating $\theta(k+1)$ with given distribution of hidden parameters $\{p(j|\mathbf{x}_n; \theta(k))|1 \leq j \leq J\}$ as follows:

$$\mu_j(k+1) = \frac{\sum_{n=1}^{N} p(j|\mathbf{x}_n; \theta(k))\mathbf{x}_n}{\sum_{n=1}^{N} p(j|\mathbf{x}_n; \theta(k))}, \tag{15.3}$$

$$\mathbf{\Sigma}_j(k+1) = \frac{\sum_{n=1}^{N} p(j|\mathbf{x}_n; \theta(k))\mathbf{V}_{nj}(k+1)}{\sum_{n=1}^{N} p(j|\mathbf{x}_n; \theta(k))}, \tag{15.4}$$

$$\pi_j(k+1) = \frac{1}{N}\sum_{n=1}^{N} p(j|\mathbf{x}_n; \theta(k)), \tag{15.5}$$

where

$$\mathbf{V}_{nj}(k+1) = (\mathbf{x}_n - \mu_j(k+1))(\mathbf{x}_n - \mu_j(k+1))^T, \tag{15.6}$$

and N is the sample number of data $\{\mathbf{x}_1, \cdots, \mathbf{x}_n, \cdots, \mathbf{x}_N\}$.

Starting from given initial parameters, the EM-algorithm repeat the E-step and M-step by monotonically increasing the logarithmic likelihood, until it reaches to a convergence. The convergence is generally assured to be a local minimum.

To avoid over-fitting problem, we introduce the AIC (Akaike's Information criterion) to balance Gaussian components number J and logarithmic likelihood. The convergence of EM-algorithm is then judged by the AIC value. The AIC is defined as follows:

$$AIC = -2\log L + 2\#(\theta). \tag{15.7}$$

Where L is the maximum likelihood, and $\#(\theta)$ is the number of free parameters in the model.

Local minimum problem is expected to be overcome by a proper design of minimum descriptor and the compression of information which gives low dimension d of the model. We start the clustering with 2 minimum descriptors ($d = 2$) both for clarity of visualization and simplicity of analysis.

For simplicity, we call the clustering of cells with respect to the minimum descriptors by EM-algorithm as "EM-clustering".

Other generalized clustering methods using informational distance are studied on exponential family with appropriate information criterion, which includes the case of EM-clustering [111]. The expansion to exponential family allows the application of a wide range of fitting components other than Gaussian, such as Poisson, multinomial, Gamma/Beta distributions, according to the characteristics of data. Furthermore, the Bregman hierarchical clustering algorithm gives an automated way to find the dendrogram according to the informational distance between the clustering components of exponential family, which would be helpful to detect the differentiated cell groups with respect to the minimum descriptors.

15.4 Tentative Results

We tentatively applied the EM-clustering to some sample data of the projects. An example of the EM-clustering at early stage of zebrafish embryogenesis with 2 minimum descriptor is shown in Figs. 15.2. The cell shape and instantaneous cell velocity were defined as minimum descriptors. The cell shape index was calculated from a spatial distribution of cell membrane markers, by approximating it with the product of principal component analysis under hypothesis that a cell can be represented with oval shape [1]. The definition of the cell shape index CS is as follows:

$$CS = \frac{PCA1 \cdot PCA2 \cdot PCA3}{(PCA1 + PCA2 + PCA3)^3}. \tag{15.8}$$

Where $PCA1$, $PCA2$, $PCA3$ are the first 3 principal components of the spacial distribution of cell membrane. The CS therefore represents the normalized ovality of the cell, which does not depend on the cell size.

The instantaneous cell velocity was calculated according to the optical flow of image intensity using Thirion's demon algorithm [108] [2]

The side view in Figs. 15.2 shows the bimodal separation of this clustering according to the low/high cell shape index. The bimodal form can be regarded as the superposition of two different cell groups. In this case, there exist rather round cells (high cell shape index value) and flat or elongated cells (low cell shape index value). Visualization with actual spatial coordinates of the cells show that these two groups have a tendency to be spatially localized in different depth of the embryo from its surface (Figs. 15.3).

The results of clustering are also visualized with the use of interactive interface developed in Embryomics project for further biological study and hypothesis forming with the support of augmented phenomenology (Figs. 15.4).

The results imply the possibility of autonomous clustering according to the minimum descriptors reflecting important typology of the cells. Compared to the primary variables, the clustering results of minimum descriptors provide a clearer point of view to characterize the cell groups with specific biological interests. Comparison between individuals is also accessible in the model space of EM-clustering.

15.5 Toward a Measurement of Individual Variation

We consider a theoretical strategy to measure individual variation in the model space of EM-clustering. The Gaussian mixture model is known to belong to exponential family and form dual-flat manifold in information geometrical framework [112]. Therefore, the temporal evolution of the clustering results of an embryo is expressed as discrete time series trajectory in the dual-flat coordinates. In other word, we consider a dynamical system in the parameter space of information geometry.

Evolution of different embryos is comparable in the dual-flat coordinates with the use of informational distance. Comparison of two different embryos with/without drug treatment is depicted in Fig. 15.5. We consider to compare the clustering results between different embryos with respect to an appropriate information distance. KL divergence is a canonical divergence which does not depend on the transformation of observables as long as they are sufficient statistics. Since the minimum descriptors usually assume the non-linear transformation of primary variables, such characteristics of KL divergence is important to assure the objectivity of comparison using the results of EM-clustering. However, KL divergence does not hold symmetric law. To express the individual variations in a metric space, we consider the symmetrization $D_s[\cdot,\cdot]$ of KL divergence $D[\cdot,\cdot]$ as follows:

$$D_s[\mathbf{x}_1, \mathbf{x}_2] = \frac{D[\mathbf{x}_1, \mathbf{x}_2] + D[\mathbf{x}_2, \mathbf{x}_1]}{2}, \tag{15.9}$$

[1] The cell shape was tentatively calculated by Mariana Remesikova and Benoit Lombardot

[2] The instantaneous cell velocity was calculated by Benoit Lombardot in his Ph.D dissertation.

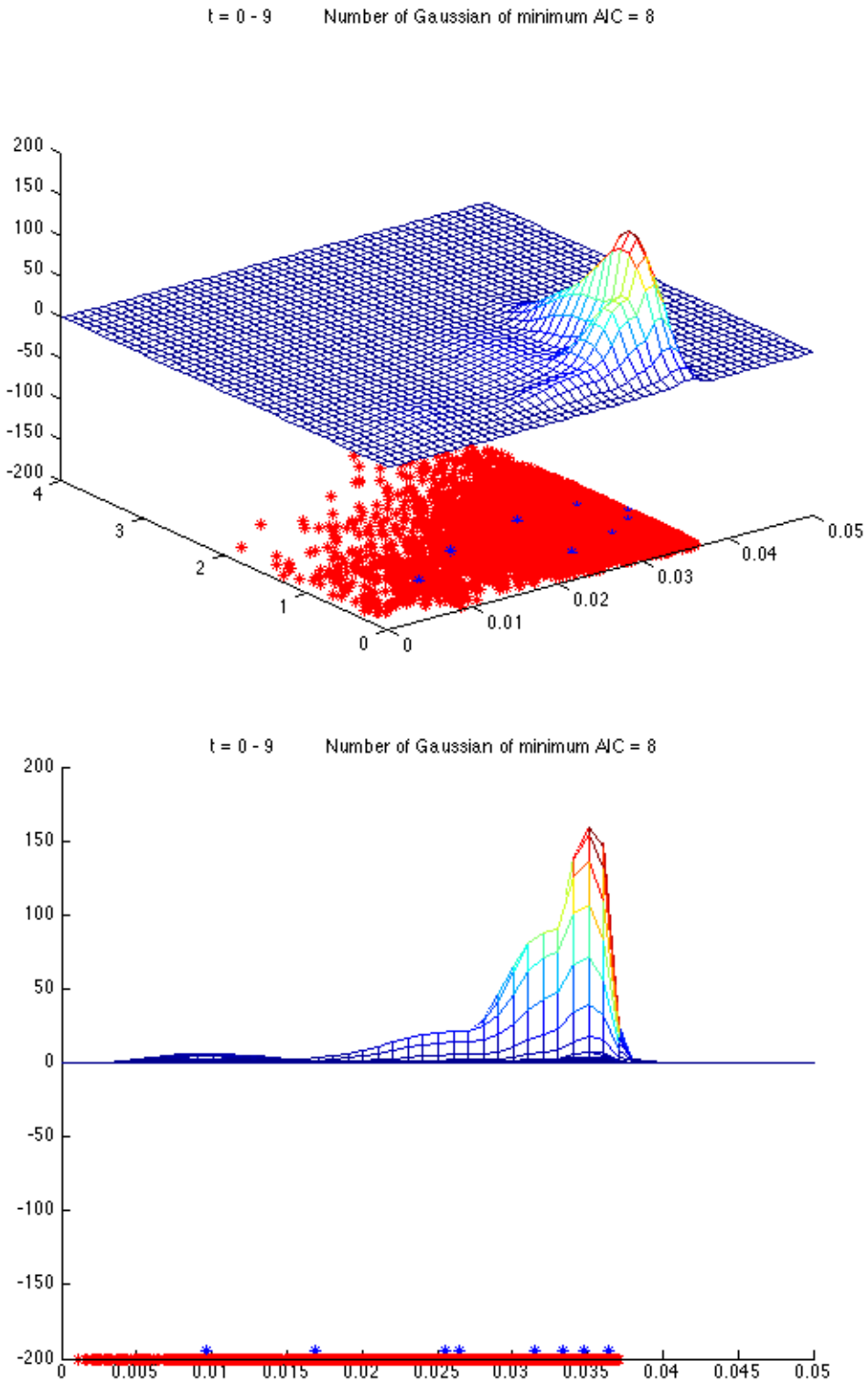


Figure 15.2: **Example of EM-clustering with two minimum descriptors.** The cell shape (X-axis) and the instantaneous cell velocity (Y-axis) were used. Up: 3D view, Down: Side view with the cell shape axis. Data id: 070420a. Animal pole of an embryo of zebrafish (danio rerio). Cells detected during 10 observations within 490 seconds are superimposed in the data. Red asterisks are the plot of cell data in 2-D plane of minimum descriptors. Blue asterisks corresponds to the mean values of each Gaussian component.

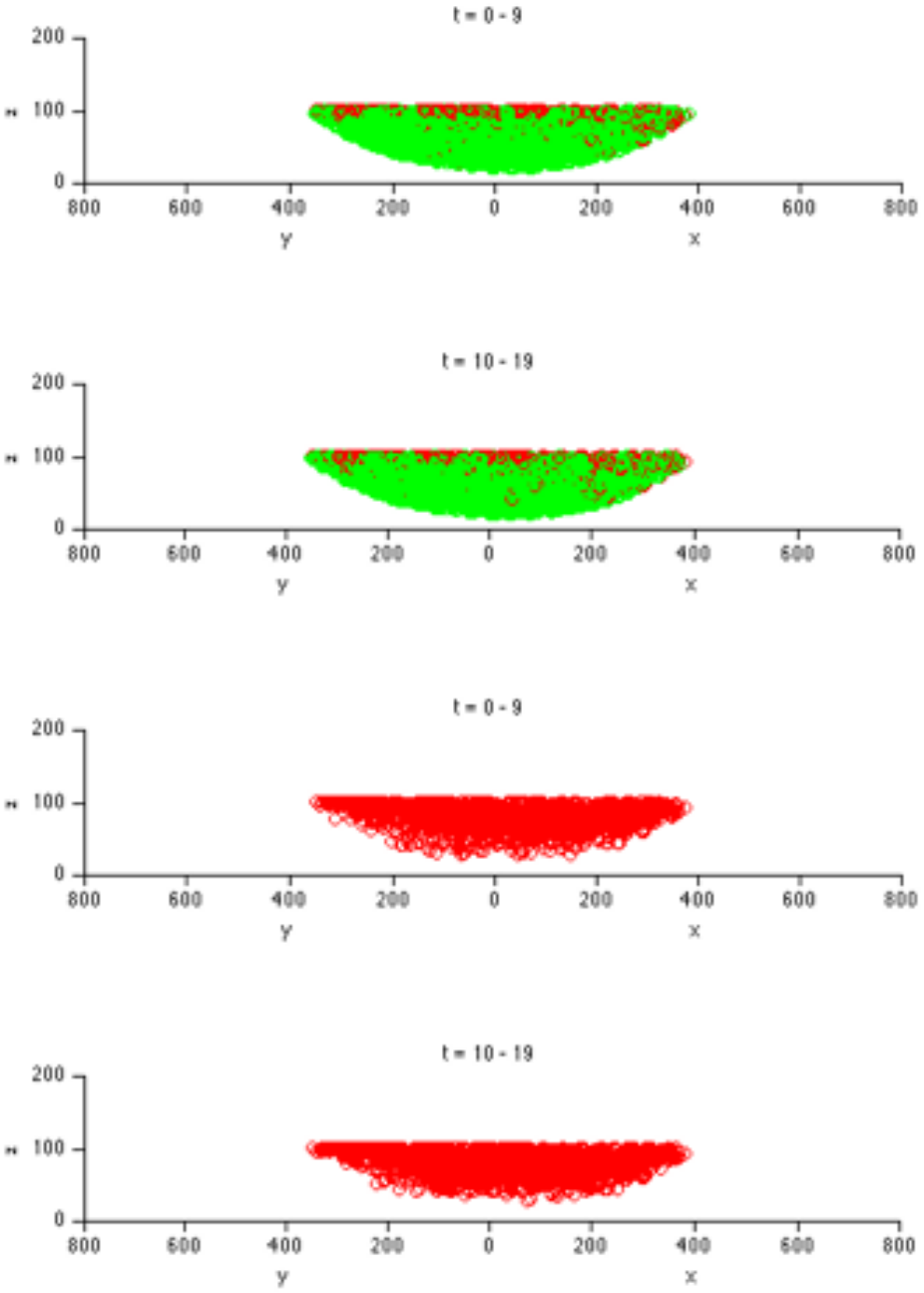


Figure 15.3: **Visualization of two cell groups separated by EM-clustering.** Cells belonging to low cell shape index distribution are depicted with red circle, while those of high cell shape index are with green circle. The top 2 figures show the superposition of all cells, while the bottom 2 figures are the visualization of low cell shape index distribution. Side view of 3-D plot is shown. Data id: 070420a. Partial imaging around the animal pole of an embryo of zebrafish (danio rerio). In each figure, cells detected during 10 observations within 490 seconds are spatially superimposed. Time interval between observations is 49 seconds.

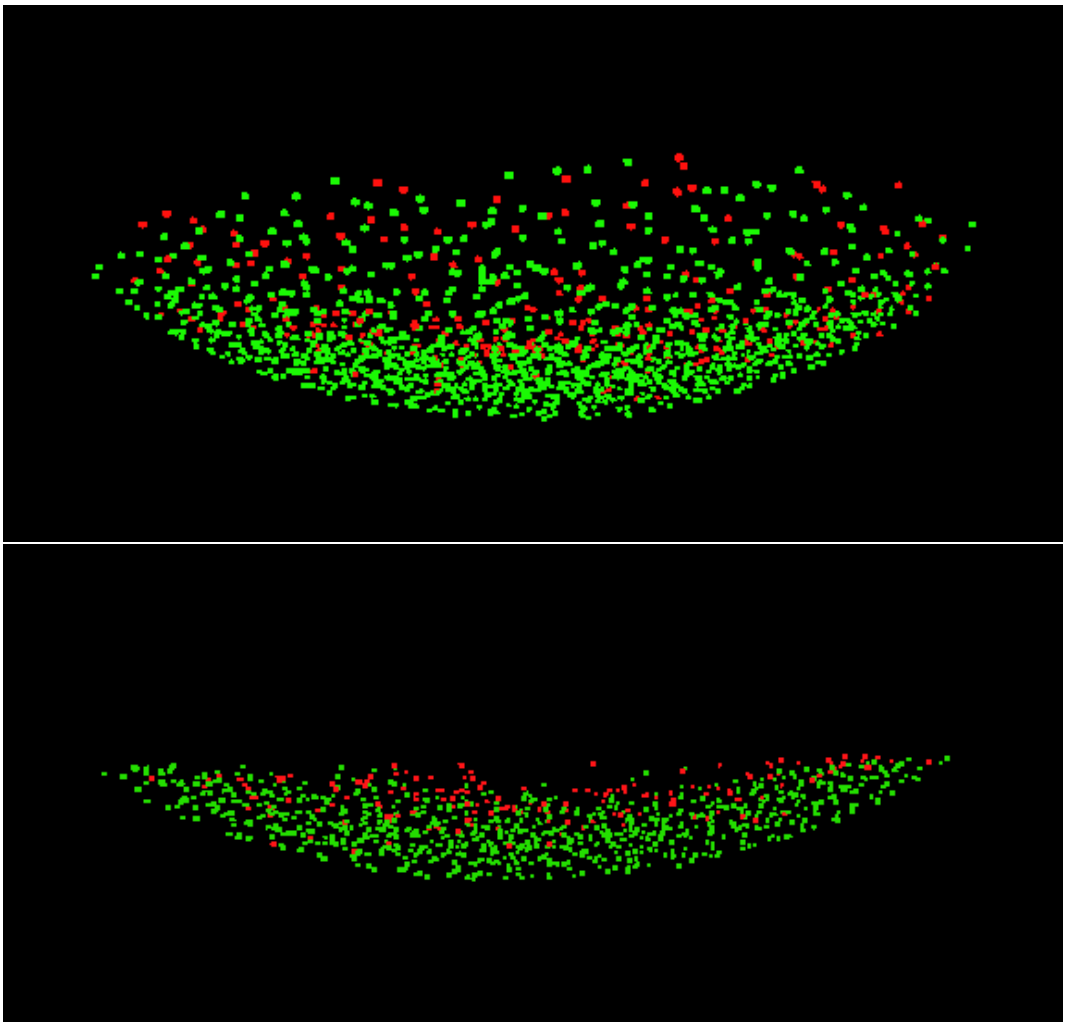


Figure 15.4: **Visualization of clustering results with Embyomics project interface for the augmented phenomenology.** Cells belonging to low cell shape index distribution are depicted with red dot, while those of high cell shape index are with green dot. The top figure shows the 3-D overview of the measured section of embryo, and the bottom figure is an example of the cross-section image to investigate spacial localization of cell groups. Data id: 070420a. Partial imaging around the animal pole of an embryo of zebrafish (danio rerio). In each figure, cells detected at an observation (time step parameter t=15) are shown.

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are a pair of different embryo data in one observation.

This symmetrization is the half of the double-sided KL divergence, and belongs to Jeffreys-Bregman divergences $D_{JB_1}[\mathbf{x}_1, \mathbf{x}_2]$ with an appropriate strictly convex and differentiable function F representing KL divergence $D[\mathbf{x}_1, \mathbf{x}_2]$ as Bregman divergence $B_F(\mathbf{x}_1, \mathbf{x}_2)$ [113]:

$$\begin{aligned} D[\mathbf{x}_1, \mathbf{x}_2] &= B_F(\mathbf{x}_1, \mathbf{x}_2) \\ &:= F(\mathbf{x}_1) - F(\mathbf{x}_2) - (\mathbf{x}_1 - \mathbf{x}_2)F'(\mathbf{x}_2), \qquad (15.10) \\ D_s[\mathbf{x}_1, \mathbf{x}_2] &= D_{JB_1}[\mathbf{x}_1, \mathbf{x}_2] \\ &:= \frac{B_F(\mathbf{x}_1, \mathbf{x}_2) + B_F(\mathbf{x}_2, \mathbf{x}_1)}{2}. \qquad (15.11) \end{aligned}$$

The Jeffreys-Bregman divergence admits a unique right-sided centroid as a minimizer independent of the generator function F. Therefore, this symmetrization is well defined with respect to the minimization operation and is also accessible to the evaluation of centrality of a given sequence.

Another known way to symmetrize KL divergence is the following Jensen-Bregman divergence, which is the generalization of celebrated Jensen-Shannon divergence [114]:

$$\begin{aligned} D_{JB_2}[\mathbf{x}_1, \mathbf{x}_2] &:= \frac{B_F(\mathbf{x}_1, \frac{\mathbf{x}_1+\mathbf{x}_1}{2}) + B_F(\mathbf{x}_2, \frac{\mathbf{x}_1+\mathbf{x}_1}{2})}{2} \\ &:= \frac{F(\mathbf{x}_1) + F(\mathbf{x}_2)}{2} - F\left(\frac{\mathbf{x}_1 + \mathbf{x}_2}{2}\right). \qquad (15.12) \end{aligned}$$

This coincides with Burbea-Rao divergence that is the information-theoretic distance for F [115] [116]. Burbea-Rao divergence is also a generalization of Jensen-Shannon divergence, and converges with respect to a weight parameter to Bregman divergence, which represents asymptotical relation between skew and symmetrized KL divergence in case of $D_{JB_2}[\cdot, \cdot]$ [113].

The correspondence between two discrete trajectories can be obtained by searching the pairs that minimize $D_s[\mathbf{x}_1, \mathbf{x}_2]$. Generally, such correspondence includes many-to-one association, but is necessary to absorb the difference of experimental parameters such as growth temperature and observation interval, which change the temporal resolution of trajectory. (For example, higher temperature realizes faster growth during the embryogenesis of zebrafish.)

The multi-scale comparison according to the specific cell group distribution are also possible. We consider two embryos of a given observation as follows:

$$p_1(\mathbf{x}, j; \theta_1) = \sum_{j=1}^{J} \pi_j^1 N(\mu_j^1, \mathbf{\Sigma}_j^1), \qquad (15.13)$$

$$p_2(\mathbf{x}, j; \theta_2) = \sum_{j=1}^{J} \pi_j^2 N(\mu_j^2, \mathbf{\Sigma}_j^2). \qquad (15.14)$$

Where $p_1(\mathbf{x}, j; \theta_1)$ and $p_2(\mathbf{x}, j; \theta_2)$ represent the density function of minimum descriptors for embryo 1 and 2, respectively.

The KL divergence between two embryo can be decomposed as follows:

$$\begin{aligned} D[p_1, p_2] &= \int d\mathbf{x} \sum_{j=1}^{J} \pi_j^1 N(\mu_j^1, \mathbf{\Sigma}_j^1) \log \frac{\pi_j^1 N(\mu_j^1, \mathbf{\Sigma}_j^1)}{\pi_j^2 N(\mu_j^2, \mathbf{\Sigma}_j^2)} \\ &= \sum_{j=1}^{J} \pi_j^1 \left(D[N(\mu_j^1, \mathbf{\Sigma}_j^1), N(\mu_j^2, \mathbf{\Sigma}_j^2)] + \log \frac{\pi_j^1}{\pi_j^2} \right). \qquad (15.15) \end{aligned}$$

The $D[p_1, p_2]$ is possible to minimize by choosing an appropriate component-wise correspondence between the 2 distributions [117]. We then formally represent a specific cell group of each

embryo with a partial mixture model as follows:

$$p'_1(\mathbf{x}, j; \theta_1) = \sum_{j=1}^{J'} \pi_j^1 N(\mu_j^1, \mathbf{\Sigma}_j^1), \tag{15.16}$$

$$p'_2(\mathbf{x}, j; \theta_2) = \sum_{j=1}^{J'} \pi_j^2 N(\mu_j^2, \mathbf{\Sigma}_j^2). \tag{15.17}$$

Here, the corresponding cell groups are represented with the same component number J' in both embryo 1 and 2. The component number can be adjusted by minimizing the AIC of one embryo.

Then the discrepancy $D'[p'_1, p'_2]$ between the specific cell groups p'_1 and p'_2 can be defined as follows:

$$D'[p'_1, p'_2] = \sum_{j=1}^{J'} \pi_j^1 \left(D[N(\mu_j^1, \mathbf{\Sigma}_j^1), N(\mu_j^2, \mathbf{\Sigma}_j^2)] + \log \frac{\pi_j^1}{\pi_j^2} \right), \tag{15.18}$$

which corresponds to the component-wise term of overall KL divergence in equation (15.15).

We then define the cell group-wise symmetrized KL divergence $D_c[\cdot, \cdot]$ as follows:

$$D_c[p'_1, p'_2] = \frac{D'[p'_1, p'_2] + D'[p'_2, p'_1]}{2}. \tag{15.19}$$

The multi-scale comparison according to the different typology of cell groups with respect to the minimum descriptors is possible using $D_c[\cdot, \cdot]$ by choosing appropriate components.

To compare between the whole trajectories of different individuals, we need to construct another metric space between individual sequences. A simple way to introduce a metric between different time series is to define the sup norm $D_e(e_1, e_2)$ between the whole sequences of embryos e_1 and e_2 as follows:

$$D_e(e_1, e_2) = \max_{x_1 \in e_1, x_2 \in e_2} D_s[\mathbf{x}_1, \mathbf{x}_2]. \tag{15.20}$$

This is equivalent to introduce a metric between different functions. We call the distance $D_e(\cdot, \cdot)$ as "inter-embryo distance".

Comparison with various temporal scales is also possible by setting a time window on embryo data. Thus the $D_e(\cdot, \cdot)$ can incorporate multi-temporal scale varying from single observation to whole process of embryogenesis.

The individual variations can then be expressed in the metric space with this norm, which is accessible to further evaluation of drug effect with respect to the range of normal fluctuation contained in normal development. Fig. 15.6 shows a schematic example of the reconstruction of inter-embryo metric space. The distribution of normal embryos and those with drug treatment are expected to form distinctive clusters. In this presentation, we aim to construct the definition of "normal" development in multi-scale resolution of both cell classification and temporal dynamics, including individual variation in contrast with other anomalies.

The dynamics of drug effect can be precisely quantified in each observation time step on the dual coordinates of Fig. 15.5. An example on the cell lineage tree with different setting will be investigated in the following chapter.

15.6 Conclusion

We developed a conceptual and theoretical framework for the phenomenological investigation of embryogenesis toward the proposition in theoretical reconstruction. Tentative results of EM-clustering showed the effectiveness of clustering with minimum descriptors both for theoretical facility and

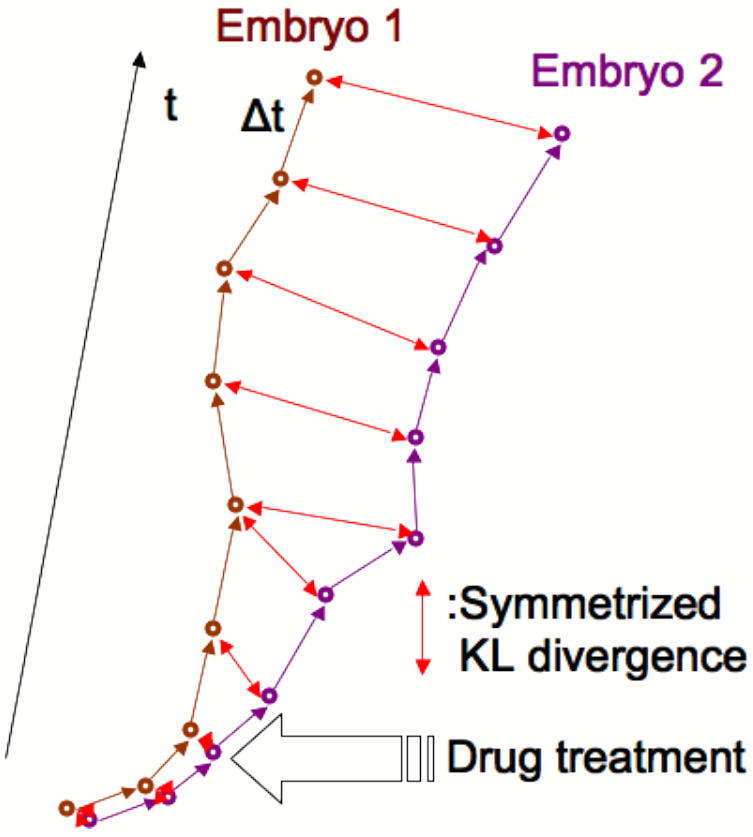


Figure 15.5: **Comparison between two individuals in the dual-flat coordinates of Gaussian mixture model.** The coordinates are the non-linear transformation of the parameter space of EM-clustering.

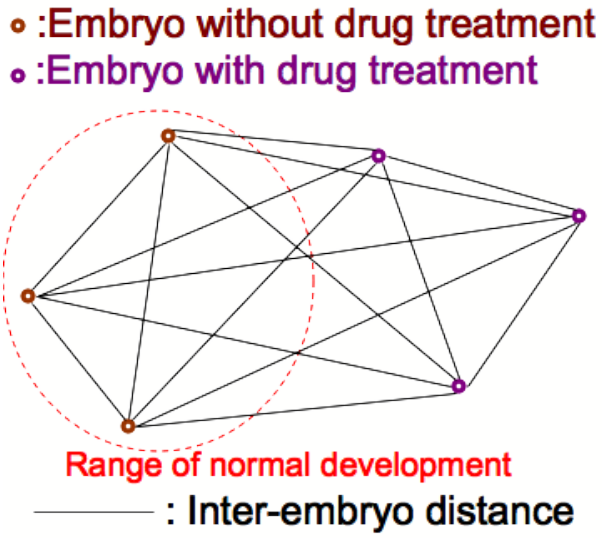


Figure 15.6: **Schematic example of reconstruction of metric space with the use of inter-embryo distance for evaluation of drug effect.** According to the distance D_e between embryos with a given time window, it is possible to present the results in 3-D metric space with principal components of data distribution.

biological clarity. Theoretical strategy toward a spatio-temporal multi-scale measurement of inter-embryo difference is constructed using informational distance. Further acquisition and treatment of quantitative primary variables are needed to derive an empirical presentation of embryogenesis with individual variations.

Chapter 16

Information Geometrical Analysis on Morphogenetic Entropy

Abstract

We consider the information geometrical analysis on the cell differentiation probability into different morphogenetic fields. This is based on the concept of "*morphogenetic entropy*" defined by Miguel A. Luengo-Oroz, which represents the amount of information that a cell requires to define the future differentiation profile. Such index is useful to define the individual difference on the basis of cell lineage tree, with different conditions such as with/without drug treatment. We theoretically show how to evaluate external effect between two groups of different treatment, and propose a series of quantity with KL divergence, namely "*relative morphogenetic entropy*", which represents the cell generation-wise propagation of the external effect. This chapter is limited to theoretical investigation.

Keywords: Morphogenetic entropy, Relative morphogenetic entropy, Cell lineage tree, Poly-drug treatment, Information geometry, KL divergence, Mutual information

Methodology: Definition of relative morphogenetic entropy on cell lineage tree as a measure of cell differentiation using information theoretical distance → Theoretical development of generation-wise decomposition of external effect with information geometry

16.1 Introduction

16.1.1 Concepts and Definitions of Morphogenetic Entropy

A lineage tree L is a binary tree associated to cells during embryo development (Fig. 16.1 [1]). The nodes of the tree correspond to cells and the links correspond to their temporal tracking. Each level of the lineage tree corresponds to a time point, going from t_{min} to t_{max}. A lineage tree can be a truncated portion of a complete binary lineage tree, so there could be several cells in the first level (t_{min}). Each cell at the level s corresponding to time t_s ($t_{min} \leq t_s \leq t_{max}$) in the lineage is linked to a (*mother*)cell corresponding to the upper level of the tree ($t_s - 1$). If a cell at t_s is the only one linked to a mother cell at $t_s - 1$, we consider that both cells are the same cell at different time points. If two cells are linked to the same cell, we consider this structure a division of the mother cell(*mitosis*). It is impossible that three cells link the same mother cell.

Let $S = \{s_1, s_2, \ldots, s_{|S|}\}$ be a set of morphogenetic regions delimited by an expert that completely covers the lineage tree L at time t_s. A cell $c_i^{t_s}$ belongs to a morphogenetic region s_j with a certain probability $p_{ij}^{t_s}$, so each cell $c_i^{t_s}$ has an associated probability distribution of belonging to the different morphogenetic regions $P_i^{t_s} = \{p_{ij}^{t_s}, s_j \in S\}$. At time t_s , the probability distributions

[1] Fig. 16.1 was offered from Miguel A. Luengo-Oroz.

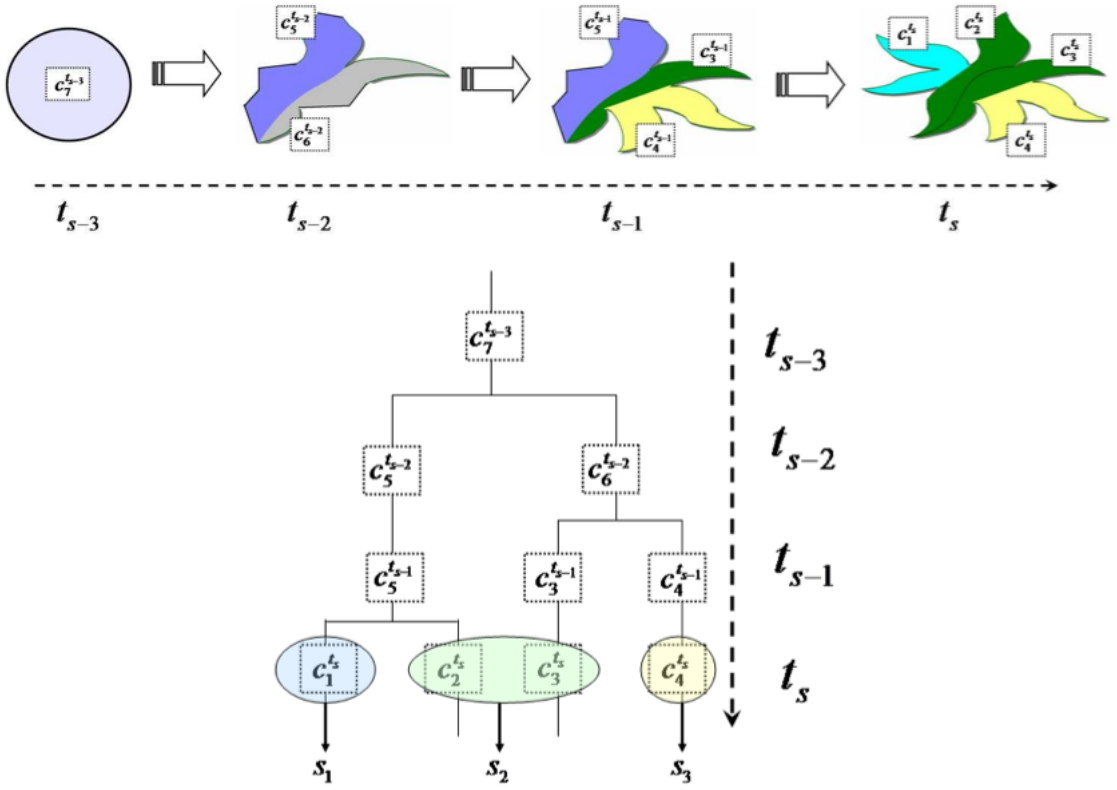


Figure 16.1: **Lineage tree $L = c_1^{t_s}, c_2^{t_s}, \ldots, c_7^{t_{s-3}}$ and morphogenetic field selection $S = \{s_1, s_2, s_3\}$.**

$P_i^{t_s}$ are completely defined by the selection S. The lineage tree topology serves to generate the probability distributions associated to the rest of the cells in the lineage: a cell c_i maintains the same probability distribution over time, so $P_i^{t_s} = P_i^{t_{s-1}}$ (at least in this first draft of the model, where there is no exchange of morphogenetic information between cells); for a cell $c_i^{t_s}$ in mitosis with two children $c_{i1}^{t_{s+1}}$ and $c_{i2}^{t_{s+1}}$, the probability distribution is the combination of the distribution of both children, so $P_i^{t_s} = \frac{1}{2}P_{i1}^{t_{s+1}} + \frac{1}{2}P_{i2}^{t_{s+1}}$.

The *morphogenetic entropy* (MH) of a cell $MH(c_i^{t_s})$ related to a morphogenetic selection S in time t_s is the entropy of the probability distribution $P_i^{t_s}$ ($0 \log 0$ is defined to be 0) [118].

$$MH(c_i^{t_s}) = -\sum_{s \in S} P_i^{t_s}(s) \log_2 P_i^{t_s}(s).$$

This measurement represents how diversified is the descendant of a cell with respect to S (this interpretation strongly reminds the Shannon's biodiversity index). When the MH is zero, this cell and all its descendants will belong to the same morphogenetic field at time s. The MH is measured in bits and can be also understand as the quantity of information needed for coding the future of all the successors of $c_i^{t_s}$ from time t_s to t_{max} in terms of lineage branchings.

This information is not the possible overall capacity of the cells. Since the MH is defined recurring from the final differentiation profile of an individual, it represents the amount of information expressed in the specific developmental process of a measured individual. It can also be interpreted as the amount of information required to code the lineage tree with a given morphogenetic field classification.

16.1.2 KL Divergence

We consider 2 cells $c_i^{t_i}$ and $c_j^{t_j}$ in the same/different lineage tree. We define the KL divergence, or the relative entropy D between the distributions of the 2 cells as follows:

$$D[P_i^{t_i} : P_j^{t_j}] = \sum_{s \in S} P_i^{t_i}(s) \log \frac{P_i^{t_i}(s)}{P_j^{t_j}(s)}. \tag{16.1}$$

This gives the amount of divergence between the 2 cells with respect to the future differentiation. In the same cell lineage, this can be naturally interpreted as the amount of differentiation with respect to the morphogenetic field definition. For example, taking $c_i^{t_i}$ a mother cell and $c_j^{t_j}$ one of its children, the KL divergence represents how much the child cell differentiated from its mother. Note this measure is asymmetric.

We formally call the KL divergence defined in equation (16.1) as "*relative morphogenetic entropy.*"

16.2 Information Geometrical Analysis

We investigate the information geometrical structure of the probability distributions $P_i^{t_s}(s)$ with the use of the relative morphogenetic entropy between two cell lineage trees with different treatment in view of the quantification of the drug effect. We derive the orthogonal decomposition of generation-wise external effect using the canonicity of the relative morphogenetic entropy. The formalization is generally compatible to the mutual information between the normal lineage tree and poly-drug treatment.

16.2.1 Orthogonal Decomposition of External Effects

We take here one specific cell $c_n^{t_n}$ in a lineage tree with morphogenetic fields $S = s_0, \cdots, s_{|S|}$, and consider associating conditional variables.

Let's take $m+1$ kinds of external effect $E = e_0, \ldots, e_m$ that we can experimentally control during the morphogenesis of an embryo. These variables can be interpreted as, for example, $e_1, \ldots, e_m$ as m kinds of different drug treatments, and e_0 as negative control. It is also compatible to poly-drug treatment if we take the element of E as the combination of plural drugs.

To see the effect of treatments E on developmental process, we define the mutual information between $S = \{s_i\}$ and $E = \{e_j\}$ as

$$I(S;E) = E_{P_n^{t_n}(S,E)} \left[\log \frac{P_n^{t_n}(S, e_j)}{P_n^{t_n}(S) P_n^{t_n}(e_j)} \right], \tag{16.2}$$

which is equivalent to

$$\begin{aligned} I(S;E) &= E_{P_n^{t_n}(E)}[D[P_n^{t_n}(S|e_j) : P_n^{t_n}(S)]], & (16.3) \\ &= E_{P_n^{t_n}(S)}[D[P_n^{t_n}(E|s_i) : P_n^{t_n}(E)]]. & (16.4) \end{aligned}$$

Here, the variables E are also included in the definition of $P_n^{t_n}(\cdot)$ and $P_n^{t_n}(\cdot,\cdot)$. $P_n^{t_n}(E)$ are simply the sample ratio of experiments among external conditions effectuated on the same cell. These distributions can be calculated from the overall results of reconstructed lineage trees with the $m+1$ different external conditions.

We define the dual affine coordinate η, θ for the distributions $P_n^{t_n}(S)$ and $P_n^{t_n}(S|e_j)$ as follows (for simplicity, we eliminated the definition of $\eta(S)$ and $\theta(S|e_j)$ that are not needed for the following development.):

$$\begin{aligned} \eta(S|e_j) &= \zeta(S|e_j) &&= (\eta_1(S|e_j), \cdots, \eta_{|S|}(S|e_j)), & (16.5) \\ \theta(S) &= \zeta(S) &&= (\theta_1(S), \cdots, \theta_{|S|}(S)), & (16.6) \end{aligned}$$

where $\eta_i(S|e_j) = P_n^{t_n}(s_i|e_j)$, and $\theta_i(S) = \log \frac{P_n^{t_n}(s_i)}{P_n^{t_n}(s_0)}$. Here, the morphogenetic field s_0 was distinguished from the others so that to give the norm of regularization factor $\psi(S) = -\log P_n^{t_n}(s_0)$ for $\theta(S)$, and $\phi(S|e_j) = \sum_{i=1}^{|S|} \theta_i(S|e_j)\eta_i(S|e_j) - \psi(S|e_j)$ for $\eta(S|e_j)$. The s_0 can be practically defined as apoptosis, but also interchangeable with any other s_i according to the focus of analysis.

We use the mixed coordinates ζ' for isolating the external effect e_j on the morphogenetic field s_i. The mixed coordinates is defined by replacing the i-th component of $\zeta(S|e_j)$ with that of $\zeta(S)$:

$$\zeta' = \zeta'(S, e_j) \quad = (\eta_1(S|e_j), \cdots, \eta_{i-1}(S|e_j), \theta_i(S), \eta_{i+1}(S|e_j), \cdots, \eta_{|S|}(S|e_j)). \tag{16.7}$$

Then from the extended Pythagorean theorem, we have

$$\begin{aligned} D[P_n^{t_n}(S|e_j) : P_n^{t_n}(S)] &= D[\zeta(S|e_j) : \zeta(S)] && (16.8)\\ &= D[\zeta(S|e_j) : \zeta'] + D[\zeta' : \zeta(S)]. && (16.9) \end{aligned}$$

The second line of this decomposition means that we are able to decompose, in terms of the KL divergence, the effect of e_j into the ith morphogenetic field (left term) and the others (right term). If we also want to measure the effect on s_0, it is simply done by exchanging $P_n^{t_n}(s_0)$ for $P_n^{t_n}(s_{i\neq 0})$ in the definition of the coordinates.

From the equation (16.3) and (16.9), we can derive the decomposition of mutual information:

$$\begin{aligned} I(S;E) &= E_{P_n^{t_n}(E)}[D[\zeta(S|e_j) : \zeta'] + D[\zeta' : \zeta(S)]], && (16.10)\\ &= E_{P_n^{t_n}(E)}[D[\zeta(S|e_j) : \zeta']] + E_{P_n^{t_n}(E)}[D[\zeta' : \zeta(S)]]. && (16.11) \end{aligned}$$

Here, the left term of (16.11) is the overall effect of E to s_i, and the right term is the effect of E to the rest $S\backslash s_i$.

16.2.2 Orthogonal Decomposition of Influences Appearing in the t_n-th Generation

We consider 2 pairs of succeeding generations in a lineage tree with different conditions, $\{P_{n_m}^{t_n-1}(S)$, $P_n^{t_n}(S)\}$ and $\{P_{n_m}^{t_n-1}(S|e_j)$, $P_n^{t_n}(S|e_j)\}$. The cell n_m is the mother cell of the cell n.

The interest here is to purify the influence of external effect on morphogenetic fields at each generation. This enable us to quantify in which stage of development the external effect can play crucial disturbance.

Let us consider the affine transformation of the dual affine coordinates on the dually flat space of this model. The degree of freedom of the coordinates $[\theta_i]$, $[\eta_i]$, and the potential ψ, ϕ, are expressed by [119]

$$\theta'_j = \sum_{i=1}^{|S|} A_i^j \theta_i + B^j, \tag{16.12}$$

$$\eta'_j = \sum_{i=1}^{|S|} C_j^i \eta_i + D^j, \tag{16.13}$$

$$\psi' = \psi + \sum_{j=1}^{|S|} D_j \theta'_j + c, \tag{16.14}$$

$$\phi' = \phi + \sum_{j=1}^{|S|} B_j \eta'_j - \sum_{j=1}^{|S|} B_j D_j - c, \tag{16.15}$$

where (A_i^j) is a regular matrix, (C_j^i) is its inverse, B_j and D_j are real-valued vectors, and c is the real number. Hence, the KL divergence $D[p:q]$ between 2 distributions p,q remains invariant under these degrees of freedom, since it is also expressed as

$$D[p:q] = \psi(q) + \phi(p) - \sum_{i=1}^{|S|} \theta_i(q)\eta_i(p), \tag{16.16}$$

which cancels completely (16.12) - (16.15). This means we can arbitrary change the coordinates to have useful decompositions of statistics among plural distributions as long as it satisfies (16.12) - (16.15) and orthogonality in the mixture coordinates.

We now aim to decompose the difference between $P_n^{t_n}(S)$ and $P_n^{t_n}(S|e_j)$, which represents the deviation from normal development at the t_n-th generation caused by external effect.

For simplicity, we formulate the cases where we can find orthogonal mixture coordinates by the linear transformation of the dual-flat coordinates. The results are possible to apply in general case by combining hierarchically plural mixture coordinates in the same manner, except the Pythagorean relation based on the orthogonality between the generation-wise effect in equations (16.26)-(16.29).

If $P_n^{t_n}(S) \neq P_{n_m}^{t_n-1}(S) \neq P_{n_m}^{t_n-1}(S|e_j)$ and $D[P_{n_m}^{t_n-1}(S|e_j) : P_n^{t_n}(S)] = D[P_{n_m}^{t_n-1}(S|e_j) : P_{n_m}^{t_n-1}(S)] + D[P_{n_m}^{t_n-1}(S) : P_n^{t_n}(S)]$ hold, there exists a set of the affine transformation of the coordinates η and θ such that

$$\begin{aligned}(\theta'_1(P_{n_m}^{t_n-1}(S)), \cdots, \theta'_{|S|-1}(P_{n_m}^{t_n-1}(S))) &= (\theta'_1(P_n^{t_n}(S)), \cdots, \theta'_{|S|-1}(P_n^{t_n}(S))), &(16.17)\\ \eta'_{|S|}(P_{n_m}^{t_n-1}(S)) &= \eta'_{|S|}(P_{n_m}^{t_n-1}(S|e_j))). &(16.18)\end{aligned}$$

This means the effect of e_j on $t_n - 1$ generation is parameterized only by $\theta'_{|S|}$.

With the new coordinates η', θ', and the potential ψ, ϕ, we can compose the mixture coordinates of $P_n^{t_n}(S)$, $P_{n_m}^{t_n-1}(S)$ and $P_{n_m}^{t_n-1}(S|e_j)$ as follows:

$$\begin{aligned}\zeta'(P_n^{t_n}(S)) &= (\eta'_1(P_n^{t_n}(S)), \cdots, \eta'_{|S|-1}(P_n^{t_n}(S)), \theta'_{|S|}(P_n^{t_n}(S))), &(16.19)\\ \zeta'(P_{n_m}^{t_n-1}(S)) &= (\eta'_1(P_{n_m}^{t_n-1}(S)), \cdots, \eta'_{|S|-1}(P_{n_m}^{t_n-1}(S)), \theta'_{|S|}(P_{n_m}^{t_n-1}(S))), &(16.20)\\ \zeta'(P_{n_m}^{t_n-1}(S|e_j)) &= (\eta'_1(P_{n_m}^{t_n-1}(S|e_j)), \cdots\\ &\quad \cdots, \eta'_{|S|-1}(P_{n_m}^{t_n-1}(S|e_j)), \theta'_{|S|}(P_{n_m}^{t_n-1}(S|e_j))). &(16.21)\end{aligned}$$

It is now natural to consider a point $Q_n^{t_n}$ with the following coordinates (Fig. 16.2):

$$\begin{aligned}(\theta'_1(Q_n^{t_n}), \cdots, \theta'_{|S|-1}(Q_n^{t_n})) &= (\theta'_1(P_{n_m}^{t_n-1}(S|e_j)), \cdots, \theta'_{|S|-1}(P_{n_m}^{t_n-1}(S|e_j))), &(16.22)\\ \eta'_{|S|}(Q_n^{t_n}) &= \eta'_{|S|}(P_n^{t_n}(S))). &(16.23)\end{aligned}$$

The coordinates θ' and η' can implement e-flat and m-flat structure, respectively, so that there exist 2 parallel e-geodesics $P_{n_m}^{t_n-1}(S)$ - $P_{n_m}^{t_n-1}(S|e_j)$, $P_n^{t_n}(S)$ - $Q_n^{t_n}$, and 2 parallel m-geodesics $P_{n_m}^{t_n-1}(S)$ - $P_n^{t_n}(S)$, $P_{n_m}^{t_n-1}(S|e_j)$ - $Q_n^{t_n}$. Here, the concept of being parallel is defined by orthogonality between the e-geodesics and m-geodesics.

Therefore, the difference between $P_n^{t_n}(S)$ and $Q_n^{t_n}$ can be considered as the combination of the deviation from normal development at $t_n - 1$ generation (difference on e-geodesic $P_{n_m}^{t_n-1}(S)$ - $P_{n_m}^{t_n-1}(S|e_j)$), and the normal development occurring in t_n-th generation (difference on m-geodesic $P_{n_m}^{t_n-1}(S)$ - $P_n^{t_n}(S)$). On the other hand, the difference between $Q_n^{t_n}$ and $P_n^{t_n}(S|e_j)$ is the new information appearing for the first time at the t_n-th generation.

We can decompose these different elements of change with the use of Phytagorean relation, by defining again the new coordinates η'', θ'', and the potential ψ'', ϕ'', for $P_n^{t_n}(S)$, $P_n^{t_n}(S|e_j)$, and $Q_n^{t_n}$, if the following conditions hold:

$$\begin{aligned}(\theta''_1(Q_n^{t_n}), \cdots, \theta''_{|S|-1}(Q_n^{t_n})) &= (\theta''_1(P_n^{t_n}(S|e_j)), \cdots, \theta''_{|S|-1}(P_n^{t_n}(Se_j))), &(16.24)\\ \eta''_{|S|}(Q_n^{t_n}) &= \eta''_{|S|}(P_n^{t_n}(S))). &(16.25)\end{aligned}$$

We then have

$$D[P_n^{t_n}(S|e_j) : P_n^{t_n}(S)] = D[Q_n^{t_n} : P_n^{t_n}(S)] + D[P_n^{t_n}(S|e_j) : Q_n^{t_n}]. \tag{16.26}$$

This decomposition can be further expanded to line up all projections of the external effect at each generation into the deviation at t_n-th generation (Fig. (16.3)):

$$\begin{aligned} D[P_n^{t_n}(S|e_j) : P_n^{t_n}(S)] = D[Q_1^{t_n} : P_n^{t_n}(S)] &+ \sum_{k=2}^{n} D[Q_k^{t_n} : Q_{k-1}^{t_n}] \\ &+ D[P_n^{t_n}(S|e_j) : Q_n^{t_n}]. \end{aligned} \tag{16.27}$$

Where $Q_k^{t_n}$ is the projection of $P_k^{t_k}(S|e_j)$ parallel to the m-geodesic $P_k^{t_k}(S)$-$P_n^{t_n}(S)$ and orthogonal to the e-geodesic $P_n^{t_n}(S)$-$P_n^{t_n}(S|e_j)$. The cell numbers in 1 lineage were aligned from 1 to n with the time from t_1 to t_n for simplicity.

Since the coordinates η and θ are dually flat, we can also have the following decomposition in the opposite sense, by exchanging the role of η'' and θ'' in (16.24) and (16.25).

$$\begin{aligned} & D[P_n^{t_n}(S) : P_n^{t_n}(S|e_j)] \\ = \; & D[P_n^{t_n}(S) : Q_n^{t_n}] + D[Q_n^{t_n} : P_n^{t_n}(S|e_j)], & (16.28) \\ = \; & D[P_n^{t_n}(S) : Q_1^{t_n}] + \sum_{k=2}^{n} D[Q_{k-1}^{t_n} : Q_k^{t_n}] + D[Q_n^{t_n} : P_n^{t_n}(S|e_j)]. & (16.29) \end{aligned}$$

Chi-squared test can be performed on each divergence if calculated with the same type of mixture coordinates. Defining 2 points p and q with the same $\eta_1, \cdots, \eta_{|S|-1}$ but different $\theta_{|S|}$ in mixture coordinates ζ', the chi-squared statistics λ of degree of freedom 1 is asymptotically defined as follows with respect to the sample number N_{sample} [77].

$$\lambda = 2N_{sample} D[p : q] \sim \chi^2(1). \tag{16.30}$$

These results are applicable to the general case where the existence of the orthogonal coordinates between $P_n^{t_n}(S)$, $P_{n_m}^{t_n-1}(S)$, and $P_{n_m}^{t_n-1}(S|e_j)$ are not assured. In such case, only $P_n^{t_n}(S) \neq P_{n_m}^{t_n-1}(S) \neq P_{n_m}^{t_n-1}(S|e_j)$ is required as the necessary condition of lineage tree. A series of quantity $\{Q_k^{t_l} | 1 \leq l \leq k \leq n\}$ can be defined based on the parallel relation between the m-geodesics $P_{n_m}^{t_n-1}(S)$-$P_n^{t_n}(S)$ and $P_{n_m}^{t_n-1}(S|e_j)$-$Q_n^{t_n}$. Although the generation-wise effect $D[P_n^{t_n}(S|e_j) : Q_n^{t_n}]$ and $D[Q_n^{t_n} : P_n^{t_n}(S|e_j)]$ are possible to calculate, the Pythagorean relation in equations (16.26)-(16.29) are lost in general case.

To obtain the Pythagorean relations in general case, we need to modify the definition of $Q_n^{t_n}$ as in Fig. 16.4. The new $Q_n^{t_n}$ is defined on the m-geodesic $P_{n_m}^{t_n-1}(S|e_j)$-$Q_n^{t_n}$ which is defined as parallel to the m-geodesic $P_{n_m}^{t_n-1}(S)$-$P_n^{t_n}(S)$, with the following condition:

$$D[P_n^{t_n}(S) : P_n^{t_n}(S|e_j)] = D[P_n^{t_n}(S) : Q_n^{t_n}] + D[Q_n^{t_n} : P_n^{t_n}(S|e_j)], \tag{16.31}$$

or

$$D[P_n^{t_n}(S|e_j) : P_n^{t_n}(S)] = D[Q_n^{t_n} : P_n^{t_n}(S)] + D[P_n^{t_n}(S|e_j) : Q_n^{t_n}]. \tag{16.32}$$

Which corresponds to the equation (16.26).

16.3 Industrial Application

With the measurement of the individual susceptibility to a genetic disease explored in a mutant fish population, this theoretical framework will allow us to tackle the complexity of the genotype-environment interaction and genotype-phenotype relationship. This will be a fundamental step towards the measurement of the individual susceptibility to genetic diseases.

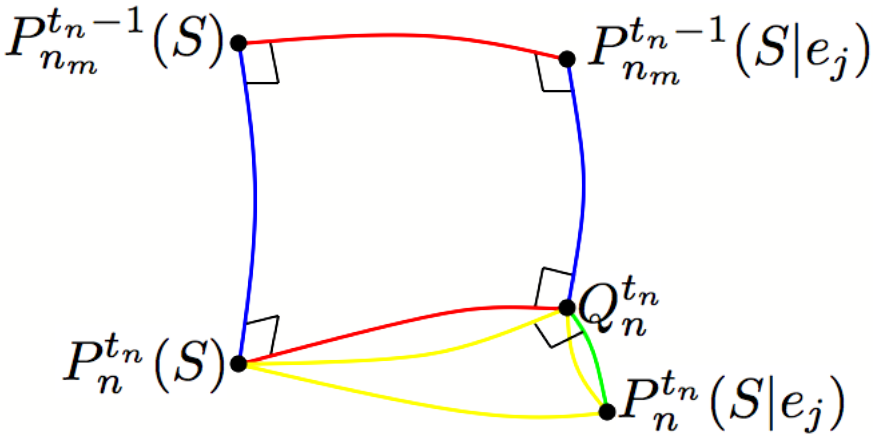


Figure 16.2: **Orthogonal Decomposition of Influences Appearing in the t_n-th Generation.** The red lines are e-geodesics on θ' coordinates, the blue lines are m-geodesics on η' coordinates, the yellow lines are e-geodesics on θ'' coordinates, and the green line is an m-geodesic on η'' coordinates. The red and blue geodesics and the yellow and green geodesics are crossing orthogonally with each other, respectively. Note that the divergence between $P_n^{t_n}(S)$ and $Q_n^{t_n}$ remains invariant under these coordinates transformations.

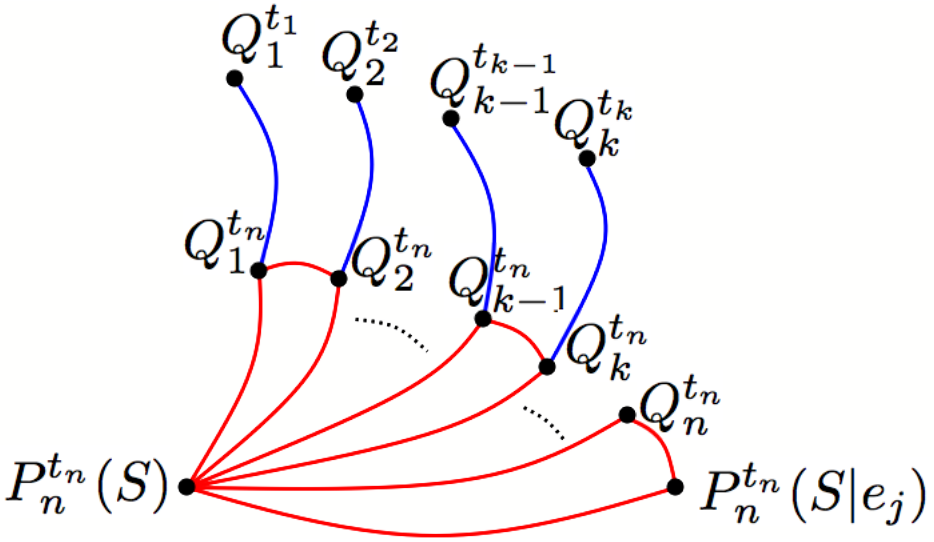


Figure 16.3: **Generation-wise Orthogonal Decomposition of Influences at the t_n-th Generation.** The red lines are e-geodesics, and the blue lines are m-geodesics. These geodesics are independently defined on different coordinates to meet the condition (16.27).

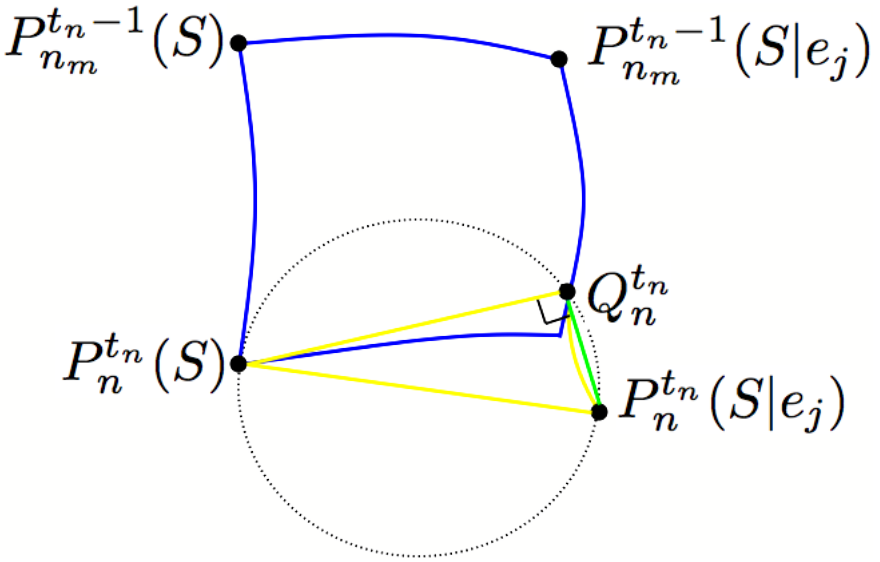


Figure 16.4: **Orthogonal decomposition of influences appearing in the t_n-th generation in general case.** The blue lines are m-geodesics defining parallel relation between $P_{n_m}^{t_n-1}(S|e_j)$-$Q_n^{t_n}$ and $P_{n_m}^{t_n-1}(S)$-$P_n^{t_n}(S)$. The yellow lines are e-geodesics, and the green line is an m-geodesic on another coordinates, which define the orthogonal decomposition in equation (16.31) or (16.32).

Our platform aims to establish a new standard for the pre-clinical trials of intelligent anti-cancer therapies. Such standard would be much more accurate and meaningful than the pre-clinical tests currently available. This does not mean that it would substitute for the tests performed in mammalian organisms. It would rather make a pre-pre-clinical strategy that would considerably minimize the extent of trials performed in mammals.

Our strategy relies on the *in vivo* observation of biological events to record data that can be further submitted to mathematical and computational treatment to achieve the automated measurement of relevant parameters. This is the kind of approach that the post genomic era requires to possibly integrate data at all levels of organization. The methodologies and tools that we aim at developing in the present project open the way to large-scale strategies to decipher the cell behaviors underlying biological processes, physiological or pathological, and investigate drug effects *in vivo* at the cellular level.

16.4 Impact on Cancerology

Our strategy is based on the hypothesis that if an anti-cancer drug is able to distinguish between cancer cells and normal cells, it will *in vivo* differentially affect the behavior of embryonic cells. This is a corollary of the current thinking that the processes of cancerogenesis and those of embryogenesis are related. More precisely, it is considered that cancerogenesis uses processes involved in some steps of normal embryonic development at the wrong place and time. Investigating and measuring, down to the cellular level, the *in vivo* and individual effects of therapeutic treatments will provide a considerable enhancement of the state of the art.

Along the lines of the same paradigm, we hypothesize that if an anti-cancer drug is able to distinguish between cancer cells and normal cells, it will differentially affect the behavior of embryonic cells *in vivo* according to their physiological state. The Dbaits paradigmatic example directly leads to the design of anti-cancer drugs screening protocols in animal models that will allow making the best possible predictions regarding drugs potential in humans.

Our strategy will allow anticipating collateral damages that are usually revealed a posteriori, when therapeutic agents are selected through a limited number of global a priori criteria. In addition, it will allow making fine adjustments and strong predictions about the therapeutic potential of combined therapies. So far, usual high cost pre-clinical tests performed using large numbers of

mammalian organisms can tell very little about the actual *in vivo* effect of the drugs at the cellular level and in a given individual. In such complex *in vivo* interaction, the theoretical reconstruction is a fundamental strategy to overcome the difficulty which can not be reduced only by the development of experimental technology.

16.5 Conclusion

We defined the concept of the relative morphogenetic entropy based on the differentiation profile of a lineage tree.With the use of information geometry, we showed that the relative morphogenetic entropy can be applied to the decomposition of the generation-wise external effect. The theory is also compatible to the evaluation of poly-drug treatment.

Part VI

Complex Systems in Linguistics

Chapter 17

Invariance in Vowel System

Abstract

We applied information geometrical setting of normal distribution to model japanese vowels based on the first and second formants. The distribution of KL divergence and its decomposed components were investigated to reveal the statistical invariance in vowel system. The result implies that although there exists significant variability in individual KL divergence distributions, the population distribution tends to converge into a specific log-normal distribution. This distribution can be considered as an invariant distribution for the standard japanese speaking population. Furthermore, it was revealed that the mean and the variance components of KL divergence is linearly related in the population distribution. Significance of these invariant features are discussed.

Keywords: Information geometry, Vowel formant, 2-Dimensional Gaussian distribution, Log-Normal distribution, KL divergence

Methodology: Information geometrical formulation of vowel system with formant frequency distributions → Analysis of information theoretical distance and its order-wise components distributions between vowels

17.1 Introduction

The sound analysis of vowel systems has been mainly focused on the time-series spectra, especially the parameters called formant. It is well known that the first and second formants give a distinctive feature of most vowels[120].

Many linguists have investigated on the difference or variation of vowel formants with respect to the speaker's physiological and social profiles such as their age, gender, occupation, living community, etc [121] [122] [123].

On the other hands, there has recently been studies on a more universalist approach with the use of various phonetic features in frequency space. The structural order of vowels composition enables us to integrate the notions of language as physical and cognitive system, and some studies refer even to the foundation of language faculty from its structural invariance [124] [125] [126].

In any case, the frequency space representing such as vowel formants or cepstrum vectors are taken *a priori* as the only mathematical space to perform the analysis. However, the coordinates transformation conserving the sufficient statistics can be considered as the change of observation method, and there exist a particular kind of coordinates that are accessible to a strong statistical theory called information geometry [16]. Although there exist applications of information geometry in speech recognition, little has been investigated on the physical law preserved in the entity of vowel system [127] [128].

In this article, we take an information geometrical formalization to the japanese five vowels formants, and analyze with the universalist point of view to find their invariant characteristics.

17.2 Modeling of Vowel System with Information Geometry

17.2.1 Sampling of Vowel Formants

The conventional five vowels of the standard japanese language, described as /a/, /e/, /i/, /o/, /u/ with Hepburn system were recorded with monotonic accent and analyzed with the use of the Praat software [129]. The first and second formants of each vowel was extracted for 500 steps with 0.01 second lapse, and 0.025 second window length for the short time Fourier transform. The data were obtained from 26 male and 29 female japanese volunteers living in or near Tokyo, whose both parents are japanese, and their age ranging from 20s to 60s. The distributions of the five vowels in the first and second formants space can each be approximated with 2-dimensional Gaussian distribution, as shown in Fig.17.1.

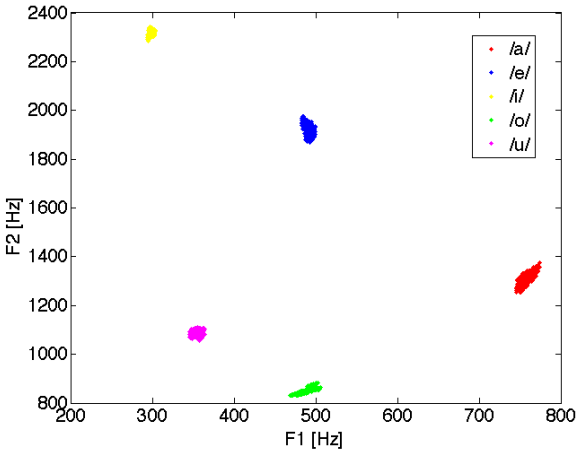


Figure 17.1: **Example of F1-F2 distribution of japanese five vowels.** Horizontal axis: first formant frequency, Vertical axis: second formant frequency. Data of only one person is depicted.

17.2.2 Model Description

Here, we derive the expression of 2-dimensional Gaussian distribution as exponential family with dual-affine coordinates $\theta = (\theta_1, \cdots, \theta_5)$ and $\eta = (\eta_1, \cdots, \eta_5)$. The Gaussian distribution $p(\mathbf{x})$ with 2-dimensional continuous variables $\mathbf{x} = (x_1, x_2)^T$ can be defined as follows:

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{|S|}} \exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T S^{-1}(\mathbf{x}-\mu)\}. \tag{17.1}$$

Where $\mu = (\mu_1, \mu_2)$ is the mean value vector, and $S = E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T] = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ is the variance-covariance matrix of the variable . The superscript T means the transpose of vectors. Note that $\sigma_{12} = \sigma_{21}$ always holds. We then need five different functions of the parameters $\mu_1, \mu_2, \sigma_{11}, \sigma_{12} = \sigma_{21}, \sigma_{22}$ as coordinates to specify a distribution.

We first have the expression of (17.1) as an exponential family of distribution by the following variable and parameter transformations:

$$p(\mathbf{x}) = \exp \sum_{i=1}^{5} \theta_i F_i(\mathbf{x}) - \Psi(\theta), \tag{17.2}$$

where $F_1(\mathbf{x}) = x_1$, $F_2(\mathbf{x}) = x_2$, $F_3(\mathbf{x}) = x_1^2$, $F_4(\mathbf{x}) = x_2^2$, $F_5(\mathbf{x}) = x_1 x_2$, $\theta_1 = A(-2\sigma_{22}\mu_1 + 2\sigma_{12}\mu_2)$, $\theta_2 = A(2\sigma_{12}\mu_1 - 2\sigma_{11}\mu_2)$, $\theta_3 = A\sigma_{22}$, $\theta_4 = A\sigma_{11}$, $\theta_5 = A(-2\sigma_{12})$, the potential $\Psi(\theta) = \log(2\pi\sqrt{|S|}) - A(\sigma_{22}\mu_1^2 - 2\sigma_{12}\mu_1\mu_2 + \sigma_{11}\mu_2^2)$, and $A = -\frac{1}{2|S|}$.

We then define the new coordinate $\eta = (\eta_1, \cdots, \eta_5)$ that is dual to $\theta = (\theta_1, \cdots, \theta_5)$ and its corresponding potential $\Phi(\eta)$ as follows:

$$\eta_i = E[F_i(\mathbf{x})], \tag{17.3}$$

$$\Phi(\eta) = \sum_{i=1}^{5} \theta_i \eta_i - \Psi(\theta). \tag{17.4}$$

Hence θ and η are dual affine coordinates. By introducing θ and η, the set of all $p(\mathbf{x})$ forms a dual-flat space with respect to the following Fisher information matrix (g_{ij}), (g^{ij}) as the Riemannian metric, and the e-, m-connection coefficients $\Gamma^{(1)}_{jik}$, $\Gamma^{(-1)}_{ijk}$, respectively [16].

$$g_{ij} = \frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \Psi(\theta), \tag{17.5}$$

$$g^{ij} = \frac{\partial}{\partial \eta_i} \frac{\partial}{\partial \eta_j} \Phi(\eta), \tag{17.6}$$

$$\Gamma^{(\alpha)}_{jik} = [ji;k] - \frac{\alpha}{2} T_{ijk}, \tag{17.7}$$

where

$$[ji;k] = \frac{1}{2}\left(\frac{\partial}{\partial \theta_i} g_{jk} + \frac{\partial}{\partial \theta_j} g_{ik} - \frac{\partial}{\partial \theta_k} g_{ij}\right), \tag{17.8}$$

$$T_{jik} = E\left[\frac{\partial}{\partial \theta_i} \log p(\mathbf{x}) \frac{\partial}{\partial \theta_j} \log p(\mathbf{x}) \frac{\partial}{\partial \theta_k} \log p(\mathbf{x})\right]. \tag{17.9}$$

17.2.3 Decomposition of KL divergence to the First- and the Second-Order Statistics

According to the definition of the Riemannian metric, information geometry gives the following theorem.

Theorem

The coordinates $\theta_{\mathbf{2}} = (\theta_3, \theta_4, \theta_5)$ are orthogonal to the coordinates $\eta_{\mathbf{1}} = (\eta_1, \eta_2)$.

Therefore, we can compose the mixed orthogonal coordinates ζ as

$$\zeta = (\eta_{\mathbf{1}}; \theta_{\mathbf{2}}) = (\eta_1, \eta_2; \theta_3, \theta_4, \theta_5). \tag{17.10}$$

Since all parameters $\mu_1, \mu_2, \sigma_{11}, \sigma_{12} = \sigma_{21}, \sigma_{22}$ are included in ζ, the mixed coordinates is sufficient to specify a probability distribution.

We use the Kullback-Leibler (KL) divergence to measure the discrepancy between two vowels $v_1, v_2 \in \{/a/, /e/, /i/, /o/, /u/\}$. Denoting the probability distribution of the vowels v_1, v_2 as $p_{v_1}(\mathbf{x})$ and $p_{v_2}(\mathbf{x})$, respectively, the KL divergence $D[p_{v_1} : p_{v_2}]$ is defined as follows:

$$D[p_{v_1} : p_{v_2}] = \int\int p_{v_1}(\mathbf{x}) \log \frac{p_{v_1}(\mathbf{x})}{p_{v_2}(\mathbf{x})} dx_1 dx_2. \tag{17.11}$$

In other way, we can also calculate $D[p_{v_1} : p_{v_2}]$ with $\theta^{v_1} = (\theta^{v_1}_1, \cdots, \theta^{v_1}_5)$ as the θ-coordinates of v_1, $\eta^{v_2} = (\eta^{v_2}_1, \cdots, \eta^{v_2}_5)$ as the η-coordinates of v_2, and their corresponding potentials $\Psi^{v_1}(\theta^{v_1})$, $\Phi^{v_2}(\eta^{v_2})$:

$$D[p_{v_1} : p_{v_2}] = -\sum_{i=1}^{5} \theta_i^{v_2} \eta_i^{v_1} + \Psi^{v_2}(\theta^{v_2}) + \Phi^{v_1}(\eta^{v_1}). \tag{17.12}$$

Using the orthogonality between θ- and η-coordinates, we have the following decomposition of $D[p_{v_1} : p_{v_2}]$.

Theorem

$$D[p_{v_1} : p_{v_2}] = D[p_{v_1} : p_{v_1 v_2}] + D[p_{v_1 v_2} : p_{v_2}], \tag{17.13}$$

where $p_{v_1 v_2}$ are given by $\zeta^{v_1 v_2} = (\eta_1^{v_1}; \theta_2^{v_2}) = (\eta_1^{v_1}, \eta_2^{v_1}; \theta_3^{v_2}, \theta_4^{v_2}, \theta_5^{v_2})$.

Since the coordinates $\theta_2^{v_2}$ include only the variance and covariance parameters, the term $D[p_{v_1} : p_{v_1 v_2}]$ represents the discrepancy in the second-order statistics of p_{v_2} from p_{v_1}, fixing the mean values μ^{v_1} as specified by p_{v_1}. Then, $D[p_{v_1 v_2} : p_{v_2}]$ represents the residual discrepancy purely in the mean values. This means that we are not only able to evaluate the discrepancy between the vowels, but to decompose its dependence into different orders of statistics.

Intuitive explanation of this theorem is shown in Fig.17.2 using one-dimensional Gaussian distributions.

For simplicity in later section, we call the logarithm of the first term as the variance component, while the logarithm of the second term as the mean value component of KL divergence. We also define the logarithmic variance/mean component ratio α as $\alpha = \frac{\log D[p_{v_1} : p_{v_1 v_2}]}{\log D[p_{v_1 v_2} : p_{v_2}]}$ for further correlational analysis. Considering the perceptual difference of these parameters, this decomposition can find phonetical meaning to study the equilibrium of vowel system (see later section). Hence, in KL divergence, the temporal fluctuation of the formants are also taken into account, while in traditional cepstrum analysis, for instance, the definition of distance between vowels are instantaneous and does not consider time-averaged higher-order statistics. In other word, the Euclidian distance between cepstrum vectors is the distance between stochastic variables, not between their probability distributions.

17.3 Result and Discussion

17.3.1 Distribution of KL Divergence between Vowels

We have calculated the KL divergence between each set of vowels for each person, and obtained the histogram summing up the sample population. This can be considered as the inter-individual distribution of japanese vowels. The log-normal distribution fits well the both male and female population, as shown in Fig.17.3 and Fig.17.4. The fitting of the histogram was performed estimating the mean value and the unbiased variance.

We have also calculated the KL divergence between each set of vowels of 1 male person, and obtained the histogram. The calculation was performed from 50 samples of the same individual. This can be considered as the intra-individual distribution of japanese vowels. The individual histogram also follows well the log-normal distribution, as shown in Fig.17.5.

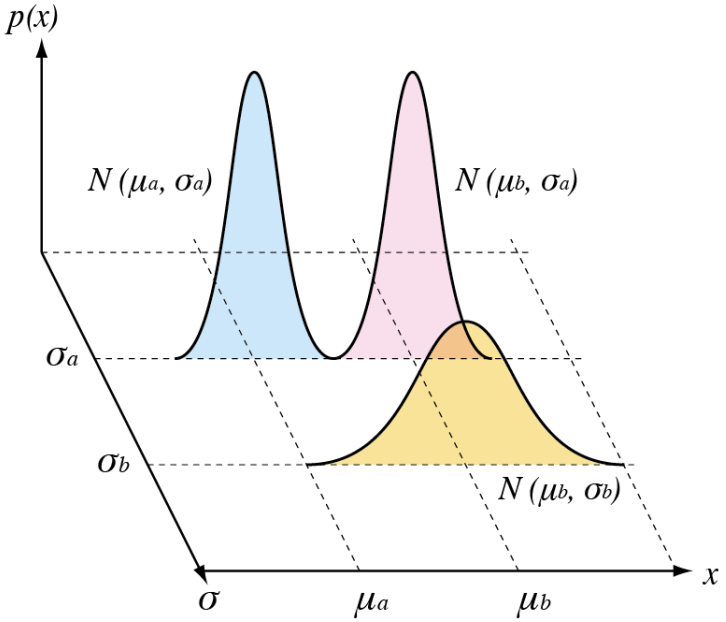


Figure 17.2: **Intuitive explanation of the Pythagorean theorem.** One-dimensional gaussian distributions $N(\mu, \sigma)$ are taken as an example. The KL divergence between $N(\mu_a, \sigma_a)$ and $N(\mu_b, \sigma_b)$ can be decomposed into the mean and variance discrepancy elements: $D[N(\mu_a, \sigma_a) : N(\mu_b, \sigma_b)] = D[(\mu_a, \sigma)_a : N(\mu_b, \sigma_a)] + D[(\mu_b, \sigma_a) : N(\mu_b, \sigma_b)]$.

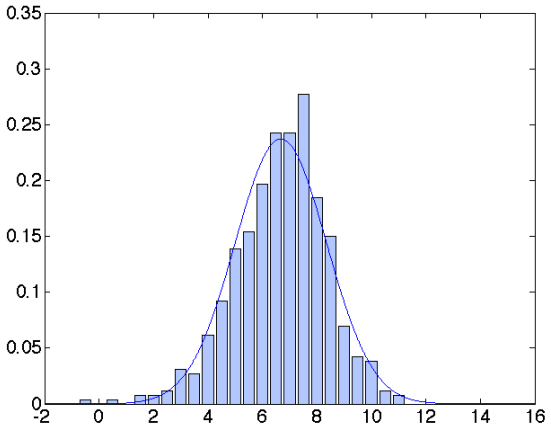


Figure 17.3: **Distribution of KL divergence between vowels in 26 male population.** Horizontal axis: logarithm of KL divergence, Vertical axis: probability density.

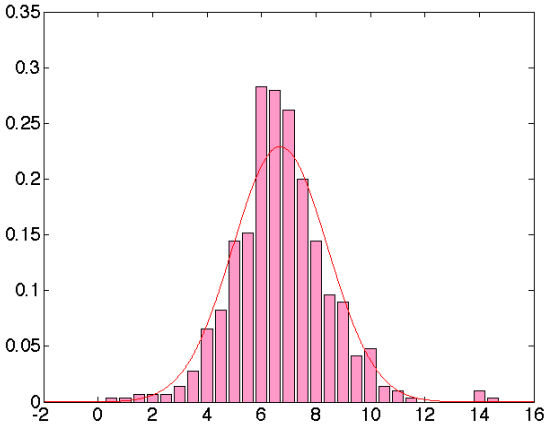


Figure 17.4: **Distribution of KL divergence between vowels in 29 female population.** Horizontal axis: logarithm of KL divergence, Vertical axis: probability density.

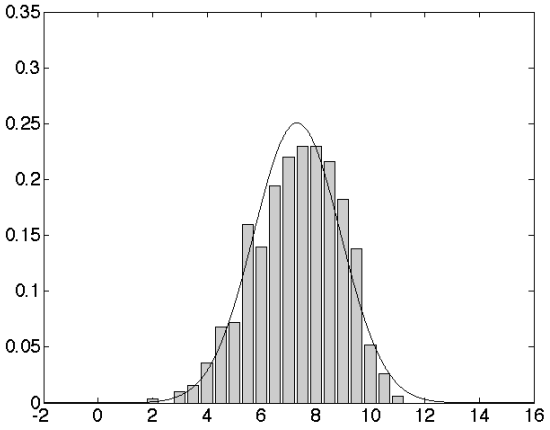


Figure 17.5: **Distribution of KL divergence between vowels in 50 samples from 1 male person (provisionally called as individual B).** Horizontal axis: logarithm of KL divergence, Vertical axis: probability density.

Next, we consider the relation between the population and the individual distribution. Taking the assumption that these distributions follow the log-normal distribution, the estimated probability density of KL divergence are depicted in Fig.17.6.

We first performed the F-test to reveal whether the variance of these distributions are significantly different or not. The two-sided significance level was set to conventional value 0.25. The result is listed in Tab.17.1. The individual distribution has significantly different variance compared to that of the male and female population distributions. On the other hand, the male and female population distribution do not have significant difference.

Based on the result of the F-test, we performed the t-test to investigate whether the mean value of these distributions are significantly different or not. The two-sided significance level was set to 0.05. The result is listed in Tab.17.2. Again, the individual distribution has significantly different mean value compared to that of the male and female population distributions, while there exists no significant difference between male and female.

These results imply that the individual distributions have statistically significant fluctuation. Though, since the population distribution is equivalent to the mixture of individual distributions, the fact that the population statistics converge to a particular distribution gives a collective limit to the individual fluctuation. This relation is not trivial considering the fact that the mixture of different normal distributions is not restricted to another normal distribution. Mathematically, it even has a capacity to approximate any continuous and differentiable function. The convergence of population distribution is a proof of collective order in individual variability. This reveals a hierarchical structure of the KL divergence distributions between individual and population. The individual distributions are distributed according to the population distribution, and despite the restriction inside of population statistics, each one is still able to express its proper variation. Therefore, the population distribution can be considered as an invariant distribution for the standard japanese speaking population.

The coincidence of male and female distributions strongly supports the notion of invariance above the superficial phonetic variation. The gender difference is often studied in order to give distinctive feature between the two groups [130] [131][132], though the converged distribution implies the existence of mutual order regardless of gender profile.

Considering the acquisition process of vowel sounds, this structure may reflect the learning process, because the individual statistics are collectively bounded by the population one.In this sense, the population distribution also makes part of the structural invariance of vowel system, which links the individual perception to the collective definition of japanese vowel system [124].

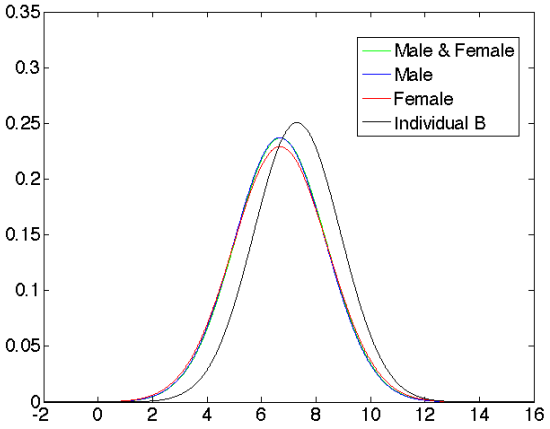


Figure 17.6: **Estimated probability density function of KL divergence between vowels.** The estimation was performed calculating the mean value and the unbiased variance of each distribution. Horizontal axis: logarithm of KL divergence, Vertical axis: probability density.

	I *vs.* M and F	I *vs.* M	I *vs.* F	M *vs.* F
F-Value	0.8935	0.8955	0.8357	0.9332
Left Critical Value	0.9312	0.9166	0.9192	0.9061
Right Critical Value	1.0736	1.0932	1.0896	1.1032
Null Hypothesis	False	False	False	True

Table 17.1: **F-Test between population and individual distribution of KL divergence.** The F-test is designed to test if two population variances are equal. The F-values, their left and right critical values of the significance level 0.25, and the truth value of the null hypothesis are listed for each combination of the distributions. I, M, F denote the individual B, the male population, and the female population, respectively.

	I *vs.* M and F	I *vs.* M	I *vs.* F	M *vs.* F
p-Value	3.0731e-17	1.9563e-12	2.6790e-12	0.8965
Null Hypothesis	False	False	False	True

Table 17.2: **t-Test between population and individual distribution of KL divergence.** Depending on the result of the F-test, the first 3 columns are the results of the Welch's t-test, while the last column is that of Student's t-test. The p-values and the truth value of the null hypothesis with respect to the significance level 0.05 are listed for each combination of the distributions. I, M, F denote the individual B, the male population, and the female population, respectively.

17.3.2 Relation between the Mean Value Component and the Variance Component of KL Divergence

It has been revealed that the perception of each vowel largely depends on the first and second formant frequencies [120] [121]. In this sense, the mean value of each vowel's formant is essential to recognize which vowel it is. Actually, the five vowels in most of our experimental data form distinctive clusters in F1-F2 space (Fig. 17.7), as reported widely in vowel systems (for example, [120] [133]).

On the other hand, the variances represent the fluctuation range of the formants, and it gives an additional phonetic feature. The difference of the formants' variances can be recognized as a part of the differences in so-called voice quality. Indeed, certain fluctuation of the formants is considered to relate the naturalness of vowel sounds [122]. The frequency variance is also associated with the naturalness of tone timbre [134]. Furthermore, distributions of vowels in variance parameter space show certain localization of each vowel, which may give another distinctive feature (Fig. 17.8).

In our setting, the differences between vowel distributions depend on these parameters, the mean values and the variance-covariance matrix, that encode qualitatively different perceptual information. Since the KL divergence between 2 vowels gives mathematical discrepancy between the 2 distributions depending on the parameters which affect our perception, it is natural to consider that it also reflects our cognitive distinctiveness between these vowels. The decomposition of the KL divergence into the mean value and the variance component enables us to investigate whether there exists a balance related to the distinctiveness between them, when comparing 2 vowels.

The relation between the mean value and the variance/mean component ratio α are plotted with linear regression in Fig.17.9. The correlation coefficients between the mean value component, the variance component, and α are listed in Tab.17.3. In each distribution, the variance component is correlated to the mean value component to certain extent. Though, if we compare the mean component to α, there exist less correlation. Especially, the sum of the male and female population shows little correlation. This fact implies that the α is the invariant ratio between the mean value and the variance component in population distribution. Indeed, α is symmetrically distributed with a sharp peak, and fit well with normal distribution, as depicted in Fig.17.10. The sharp symmetric peak supports the invariance of α in population distribution.

More precisely, the variance component consists of 2 elements, the one which is linearly proportional to the mean value component, and the other which is not linear. The linear element is canceled out when measuring the correlation with α. Although the latter exist certainly in individual distribution, it is consistent in any distribution that the correlations between the mean component and α show only weak correlation. Therefore, we deduce that the α is an invariant feature in population distribution, while it accepts certain fluctuation in each individual that cancels each other at the population level.

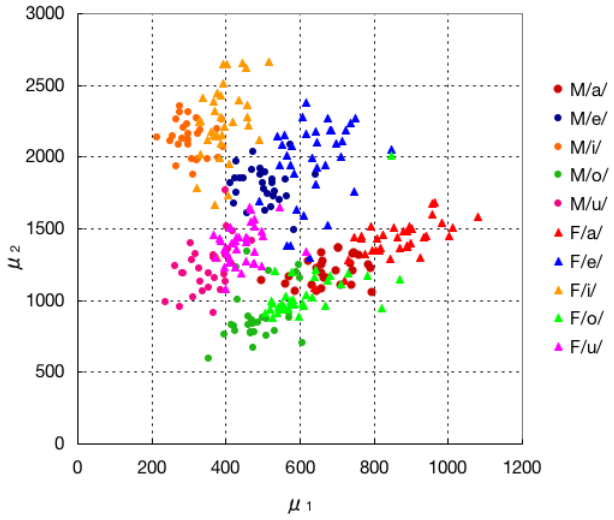


Figure 17.7: **Population distribution of five vowels' F1 F2 mean values.** M and F denote male and female, respectively. Each vowel forms cluster in its characteristic frequency region.

	M and F	M	F	I
Mean *vs.* Variance	0.3242	0.1050	0.4923	0.3986
Mean *vs.* α	0.0444	-0.0890	0.1749	0.2365

Table 17.3: **Correlation coefficients between the logarithms of the mean value component, the variance component, and the variance/mean component ratio α of KL divergence.** For simplicity, the components are referred to as Mean, Variance, and α. The results of the population and individual distributions are listed. I, M, F denote the individual B, the male population, and the female population, respectively.

17.3.3 Vowel Combination-wise Distribution of KL Divergence

We further investigate the content of lognormal distribution between vowels by decomposing it into vowel combination-wise distribution. Figs. 17.12 and 17.13 show the distribution of KL divergence for different combinations of vowels. The combination-wise distributions also show the tendency to fit lognormal distribution with various mean and variance, but contain certain fluctuation. This may be due to the combinatorial decrease of the sample number in combination-wise distribution, since the combination /a/-/e/, for example, is only 1/20 of the all possible combinations between 5 vowels. The fact that the fluctuation reduces by taking larger combinations such as /a/-/e//i//o//u/ also supports this notion. The distribution between 5 vowels of the same sample number order also show large fluctuation, as shown in Figs. 17.14. Although more accurate forms of the vowel combination-wise distributions can only be verified by augmenting sample number, the circumstantial evidences suggest the hierarchical structure between the individual and combination-wise distributions similar to that of the population and individual ones: The vowel combination-wise distributions seem to follow different lognormal distributions, under global constraint of the individual distribution.

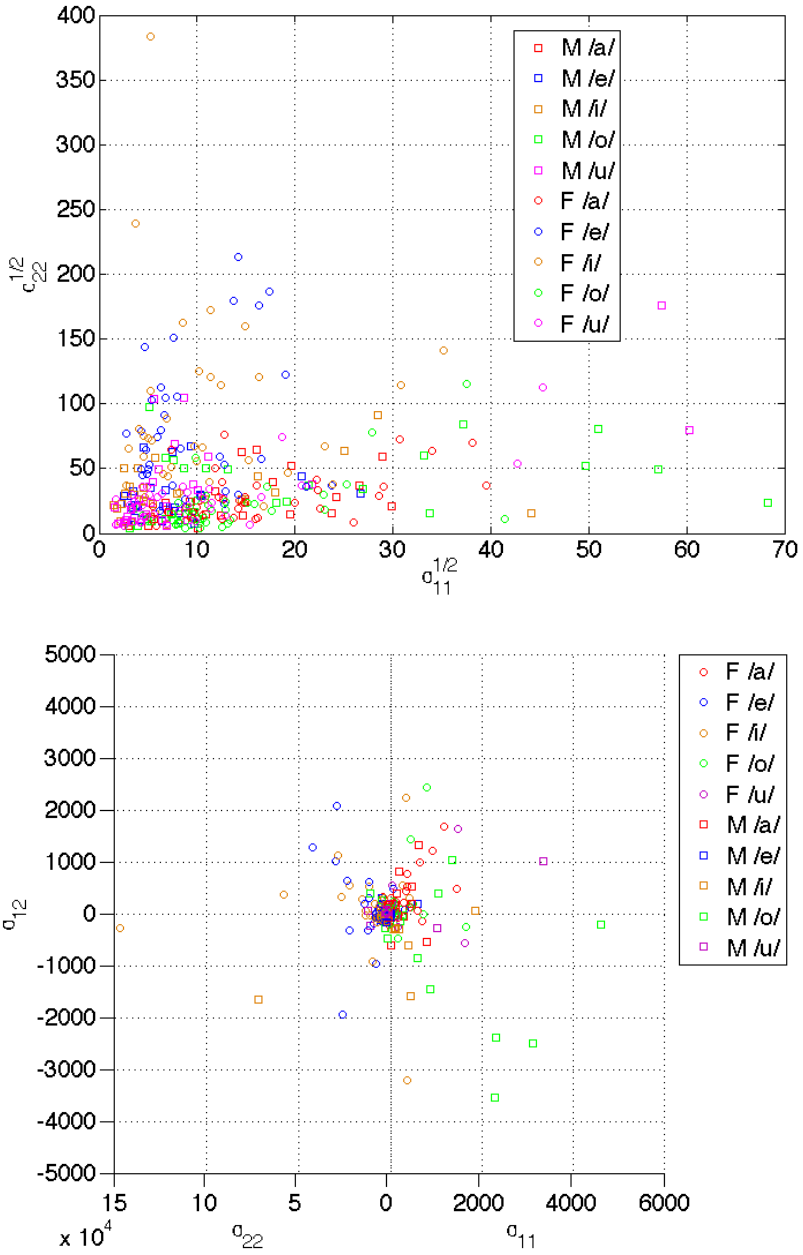


Figure 17.8: **Population distribution of five vowels' F1 F2 variances.** M and F denote male and female, respectively. The top figure is the distribution on $\sqrt{\sigma_{11}}$ - $\sqrt{\sigma_{22}}$ plane. The bottom figure is the side view of 3D plot. /a/, /o/ and /u/ are localized in relatively low F2 variance region compared to F1 variance, while /e/ and /i/ are in relatively low F1 variance region. The covariance axis does not seem to give distinctive feature among vowels.

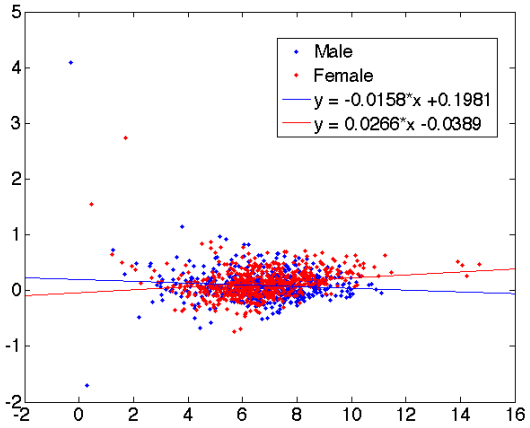


Figure 17.9: **Relation between mean value component** *vs.* **variance/mean component ratio α of KL divergence between vowels in 26 male (blue) and 29 female (red) population.** Horizontal axis: mean value component, Vertical axis: variance/mean component ratio α. The lines are the linear regressions.

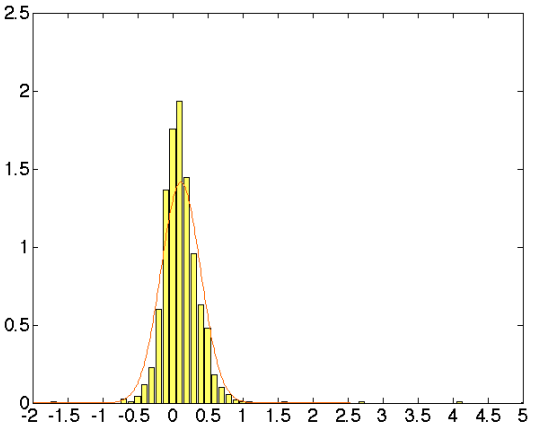


Figure 17.10: **Distribution of variance/mean component ratio α in 26 male and 29 female population.** Horizontal axis: value of α, Vertical axis: probability density.

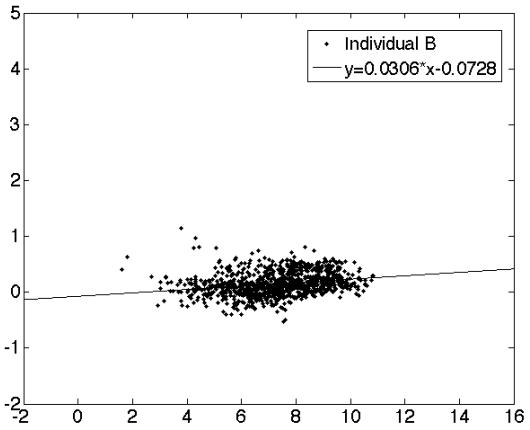


Figure 17.11: **Relation between mean value component** *vs.* **variance/mean component ratio α of KL divergence between vowels in 1 male person (provisionally called as individual B).** Horizontal axis: mean value component, Vertical axis: variance/mean component ratio α. The line is the linear regression.

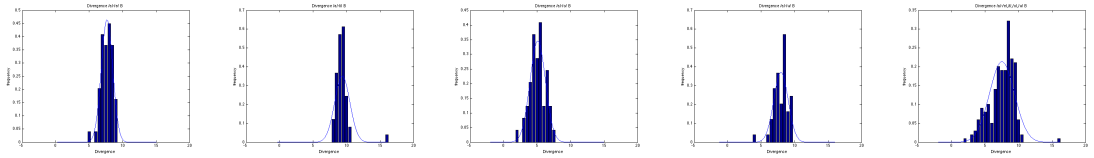


Figure 17.12: **Vowel combination-wise distribution of KL divergence of individual B.** From left to right: distribution of /a/-/e/, /a/-/i/, /a/-/o/, /a/-/u/, and /a/-/e//i//o//u/.

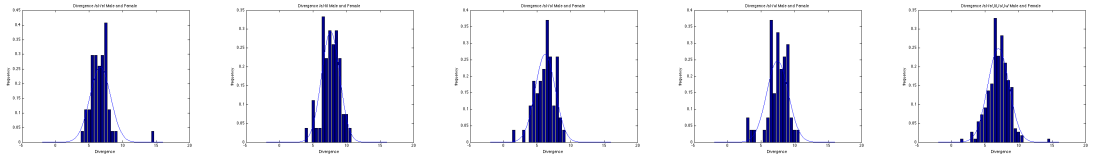


Figure 17.13: **Vowel combination-wise distribution of KL divergence of male and female.** From left to right: distribution of /a/-/e/, /a/-/i/, /a/-/o/, /a/-/u/, and /a/-/e//i//o//u/.

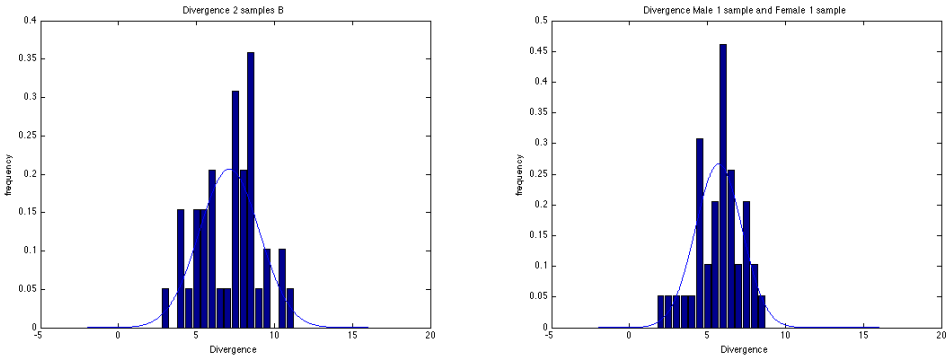


Figure 17.14: **Distribution of KL divergence between 5 vowels with reduced sample number.** Top: 2 samples of individual B (40 combinations of KL divergence). Bottom: 1 each sample from male and female (40 combinations of KL divergence).

17.3.4 The Origin of Lognormal Distribution in Vowel System: Weber-Fechner Law in η Coordinates

We investigate the origin of the observed lognormal distribution by analyzing the variation of vowels distribution in η coordinates.

It is known that independent multiplicative processes converge to a lognormal distribution [135] [136]. The distributions of 5 vowels in η coordinates do not show correlation between them. If the variance of each vowel in η coordinates is a multiplicative noise, the KL divergence between any combination of 5 vowels naturally converges to a lognormal distribution, because its definition is based on the linear combination of η coordinates (see eq. (17.12)).

Figs. 17.15 show the distributions of 5 vowels projected in $\eta_1 - \eta_2$ plane. To verify the existence of multiplicative noise, we calculate the following 2 kinds of variances: The first one is the simple variance of each vowel in η coordinates, namely the additive variance. The second one is also the variance of each vowel in η coordinates, but divided by their mean value, namely the multiplicative variance. If the variance of the vowels is multiplicative, the additive variance shows proportional increase with respect to the mean value in η coordinates, while the multiplicative variance remains constant.

Figs. 17.16 and 17.17 show the results which support the existence of multiplicative variation. Multiplicative tendency exists both in population and individual distribution in all dimensions of η coordinates. Therefore, the distribution on η coordinates implies that the origin of the observed lognormal distributions is grounded to its multiplicative variation. Note that the η_1 and η_2 coordinates corresponds to F1 and F2 value, respectively.

The multiplicative variation in frequency space has perceptual meaning known as Weber-Fechner law: For notes spaced equally apart to the human ear, the frequencies are related by a multiplicative factor. Humans hear pitch in a logarithmic or geometric ratio-based fashion. Musical scales are always based on geometric relationships for this reason. In case of vowel system, this relation would support the distinctiveness of each vowel. The observed lognormal distribution can be considered as the result of repulsive localization of each vowel according to the constant perceptual distance between them.

The multiplicative factor is also interesting when we assume a constant degree of accuracy in voice control. If the control precision of vocal tract is constant, the vocalization is naturally associated with multiplicative noise with respect to the produced frequency.

17.3.5 Relation to Ecological Linguistics: Invariants of Gibson Expanded

We started the analysis from the universalist point of view, seeking for the common structure of vowel system which would support our perception of harmonized resonance in human language. In visual perception, Gibson insisted geometrical invariants in optical flow as the foundation of perceptual significance [137]. The observed invariant relations clearly relate to the principle of Gibson's invariants, but outreach simple pictures such as formants localization [138] and invariant quantity under affine transformation [124]. The observed invariance has hierarchical structure between individual and population, different orders of statistics, and possibly

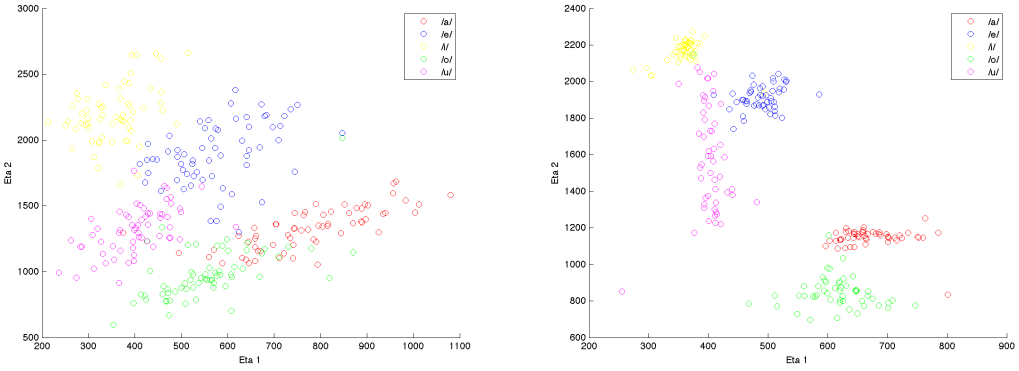


Figure 17.15: **Distribution of japanese 5 vowels in $\eta_1 - \eta_2$ plane.** Left: Male and female. Right: Individual B.

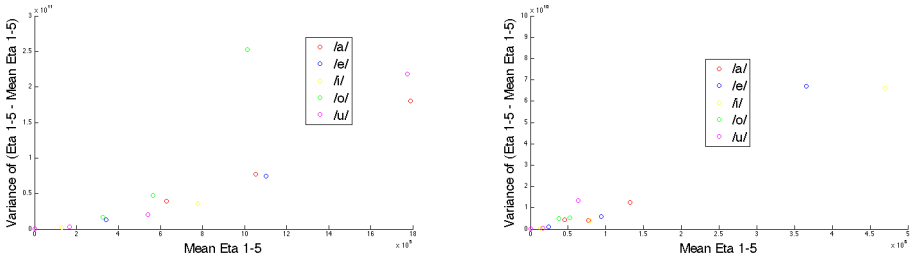


Figure 17.16: **Mean *vs* additive variance of japanese 5 vowels distributions in η coordinates.** Left: Male and female. Right: Individual B. The proportional increase of the additive variance implies multiplicative noise.

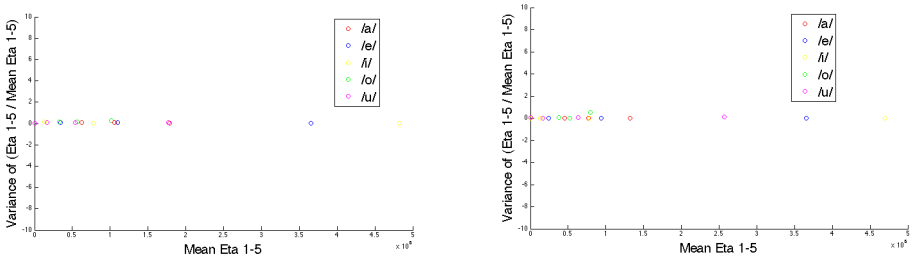


Figure 17.17: **Mean *vs* multiplicative variance of japanese 5 vowels distributions in η coordinates.** Left: Male and female. Right: Individual B. The invariance of the multiplicative variance implies multiplicative noise.

between the single pair and the whole combination of the vowels. One way to explain such complex structure was introduced in reference to the Weber-Fechner law, but the convergence of the population distribution and the relation between different orders of statistics remain untouched. These relations may become a phonetic expansion of Gibson's invariant, which reflect the complexity of phonetic perception deeply linked to our language faculty. The parallel study on the acquisition of vowel system in children would be necessary to clarify the development of these features [139][140].

It is of further interest how such invariance affect our perception of vowels, especially in case they are perturbed. Synthetic approach of vowel system controlling the discovered invariant relations would be fruitful to further attack these questions.

17.4 Conclusion

We investigated the distributions of KL divergence between 5 japanese vowels, and insisted that the population lognormal distribution and the variance/mean component ratio α are invariant features. The hierarchical relation between the population distribution, individual distribution, and vowel combination-wise distributions are also investigated. The origin of lognormal distribution is shown to be based on the multiplicative variation in formant frequency.

It is of further question whether such invariance can also be observed in other vowel systems, or by simply increasing the heterogeneity of linguistic profile in sample japanese population. Aside of statistical invariance, the geometrical composition of each vowel in the dual coordinates space remains to be analyzed. Further developmental and multilingual comparative study will be needed to relate the discovered invariance of vowel system to our cognitive mechanism.

Chapter 18

On Geometric Composition of Vowel System

Abstract

We further investigate statistical dependency between japanese five vowels with the use of information geometry. We question not only the simple correlational analysis but the dependency following the m-geodesic between the vowels which defined the projection of KL divergence in previous chapter.

We show that the five vowels are effectively in 2-dimensional surface in η coordinates, and the fluctuation of the introduced circumcenter on inter-vowel m-geodesic has uncorrelated noise with respect to its defining vowels.

The localization of the circumcenter in θ coordinates will be also characterized with the use of determinant $|S|$ in formants distribution.

The results support the origin of invariant lognormal KL divergence distribution found in previous chapter, and further imply the cognitive condition related to the quadratic dependency between the vowels in η coordinates.

Keywords: Information geometry, Vowel formant, Circumcenter, Determinant of formant distribution.

Methodology: Geometrical composition analysis of vowel system on dual-flat coordinates of information geometry → Definition of circumcenter between a pair of vowels with information geometry → Analysis of geometrical composition and fluctuation of circumcenters on dual-flat coordinates of information geometry

18.1 Introduction

In previous chapter, we investigated statistical property of KL divergence between 5 japanese vowels in both individual and population distribution. On the other hand, statistical dependency between vowels is also an important question considering the holistic order of vowel system. Do different vowels in an individual have correlated fluctuation in formant frequency during different trials ? Does individual variation contain correlated change inside of the population distribution ?

Vowel formants are also known to localize in specific region proper to each vowel [120]. The origin of the geometrical composition of vowel system is also a fundamental question. Why the vowels exist in such localization and not elsewhere ?

To tackle to these questions, we investigate the geometrical composition and the statistical dependency between the japanese 5 vowels in dual-flat coordinates defined in information geometry. For this purpose, we introduce the circumcenter on m-geodesic between two vowels using the dual-flat structure of the statistical manifold.

18.2 Geometrical Composition of Vowel System in the Dual Flat Manifold

We first investigate the geometrical composition of the japanese 5 vowels in η coordinates to clarify the parametric dependence between different dimensions. It is clear from the definition of η coordinates that each dimension has independent degree of freedom, with respect to the 5 parameters of 2-dimensional normal distribution $(\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})$ for the vowel formant distribution fitting in F1-F2 space:

$$\eta_1 = \mu_1, \tag{18.1}$$
$$\eta_2 = \mu_2, \tag{18.2}$$
$$\eta_3 = \sigma_{11} + \mu_1^2, \tag{18.3}$$
$$\eta_4 = \sigma_{22} + \mu_2^2, \tag{18.4}$$
$$\eta_5 = \sigma_{12} + \mu_1\mu_2. \tag{18.5}$$

Actual distribution of the vowels in η coordinates, however, shows certain approximative dependency between them. Figs. 18.1 and 18.2 plot the 5 vowels with individual and population distribution, respectively. Clearly η_3 and η_4 can be well approximated with quadratic function of η_1 and η_2, respectively. η_5 also has strong quadratic dependency to η_1 and η_2.

Indeed, from the definition of η coordinates, η_3, η_4, and η_5 are the quadratic function of η_1 and/or η_2 except the intercept term. The intercept terms are the elements of covariance matrix. The actual approximative relation means that the variance of the mean values in F1-F2 space is much larger than the variance of the formants. In other word, vowel distributions generally differ in mean formant frequency and not in their form. Therefore, there exist globally only 2 effective dimensions in the η coordinates. In short, the 5 japanese vowels and KL divergence between them can be represented approximately as 2-dimensional pentacle in 5-dimensional η coordinates (Fig. 18.3).

18.3 Dependency Test between Vowel Distributions

The statistics of KL divergence between vowels were studied in previous chapter. We did not however investigate the statistical dependency between vowels, which can generally vary under the constraint of the same lognormal distribution of KL divergence. The multiplicative noise of vowel formants was suggested as the generating mechanism of the lognormal distribution, but the independence between noise

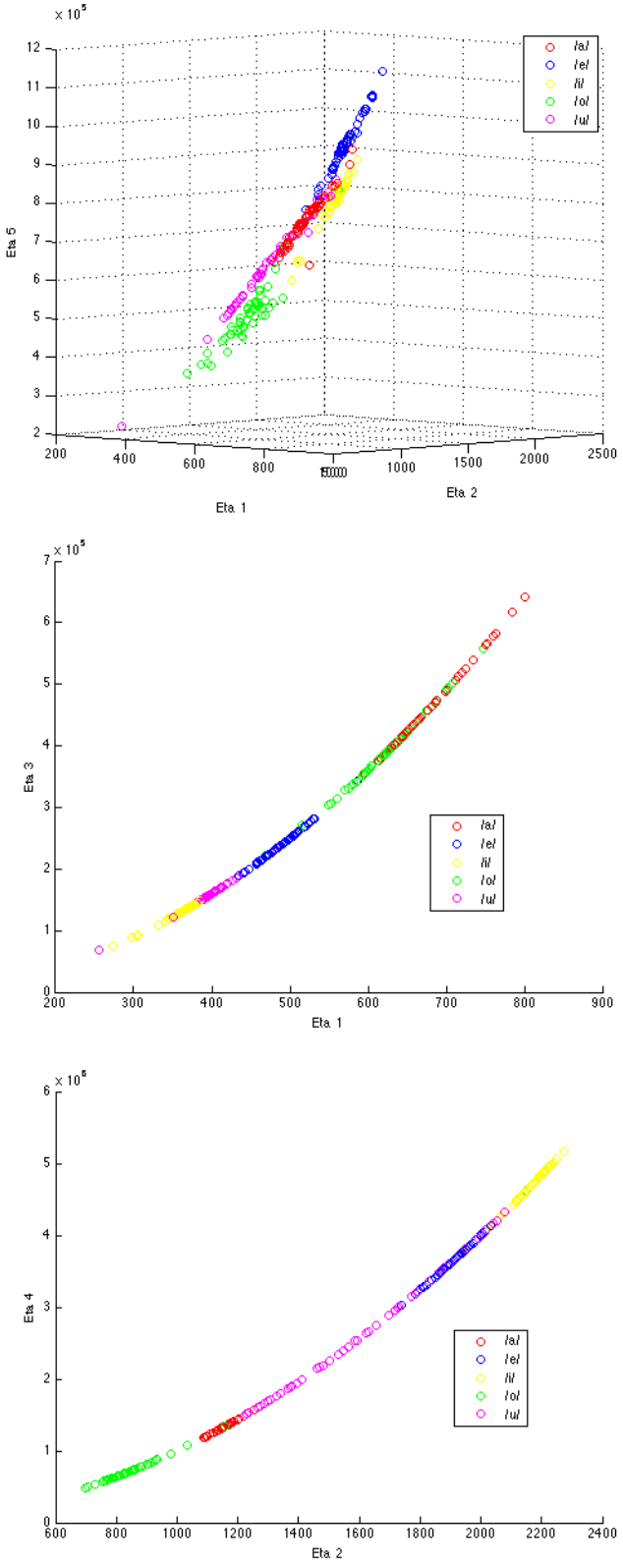


Figure 18.1: **Japanese five vowels distribution of individual B in η coordinates.** Top: $\eta_1 - \eta_2 - \eta_5$ plot. Middle: $\eta_1 - \eta_3$ plot. Bottom: $\eta_2 - \eta_4$ plot. η_3, η_4, and η_5 show quadratic dependency to η_1 and/or η_2.

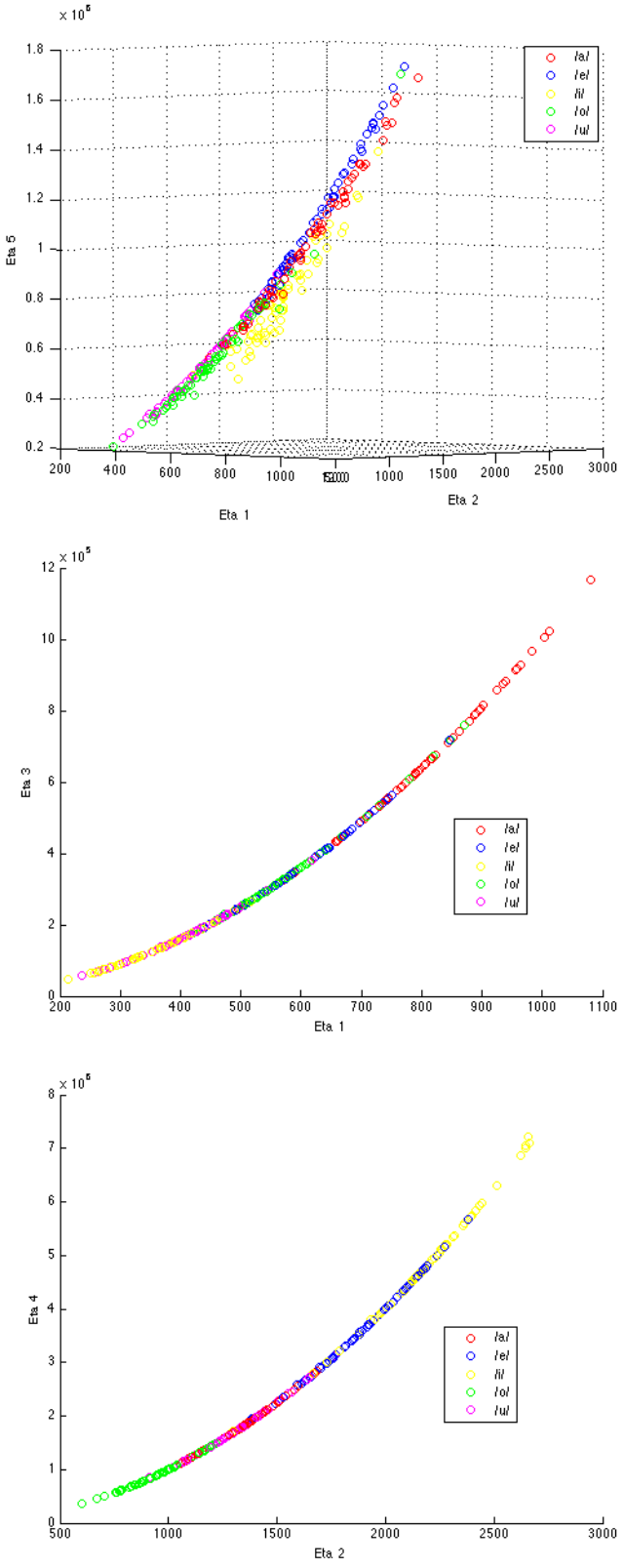


Figure 18.2: **Japanese five vowels distribution of 26 male and 29 female in η coordinates.** Top: $\eta_1 - \eta_2 - \eta_5$ plot. Middle: $\eta_1 - \eta_3$ plot. Bottom: $\eta_2 - \eta_4$ plot. η_3, η_4, and η_5 show quadratic dependency to η_1 and/or η_2.

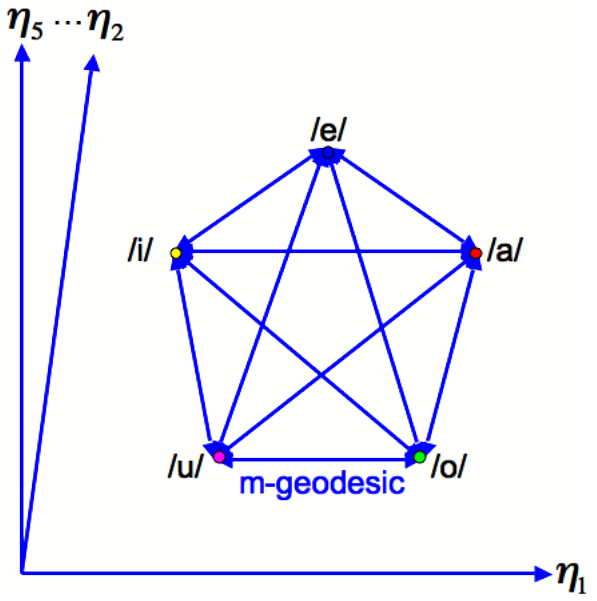


Figure 18.3: **Schematic representation of japanese five vowels distribution in η coordinates.** The vowels are aligned in effectively 2-dimensional surface according to the quadratic dependency between η coordinates. KL divergence follow m-flat projection and draw the edges of pentacle where 5 vowels form vertices. m-geodesic is expressed as linear function on η coordinates.

was not assured. To further clarify the geometrical relation in vowel system, we question the statistical dependence of fluctuation between vowels in this section.

We first investigate the simple correlation in η coordinates. Figs 18.4 and 18.5 show the identical η plot between different vowels. Little correlation is observed in an individual distribution, while population distribution shows positive correlation. The positive correlation of the five vowels with the spreading of their first principal components directions in a fan-like form support the possibility of the affine transformation approximation between individuals in η coordinates. In speech recognition, the distribution of vowel cepstrum vectors is studied with mixture Gaussian model. Invariant property of the vowel composition under affine transformation in Euclidian space is applied to the perception of speaker-variable speech [124]. Therefore, similar structural invariance may also hold in η coordinates in population distribution.

18.4 Circumcenter in the Dual Flat Manifold

We further investigate the dependency between vowels in higher order of principal components. We question the dependency contained in the 2nd order principal component of identical η plot. Under the presence of positive correlation in population distribution, such element corresponds to a negative correlation or a counterbalance between the fluctuation of different vowels. In case of decorrelation such as in individual distribution, the counterbalance between vowel fluctuation may still exist if they cancel out by the existence of hidden positive correlation. Even if both positive and negative correlation exist in generating mechanism of vowel system, the overall appearance can result in decorrelation. Indeed, an individual can easily shift formant frequencies according to his/her voice pitch to maintain the vowelness,

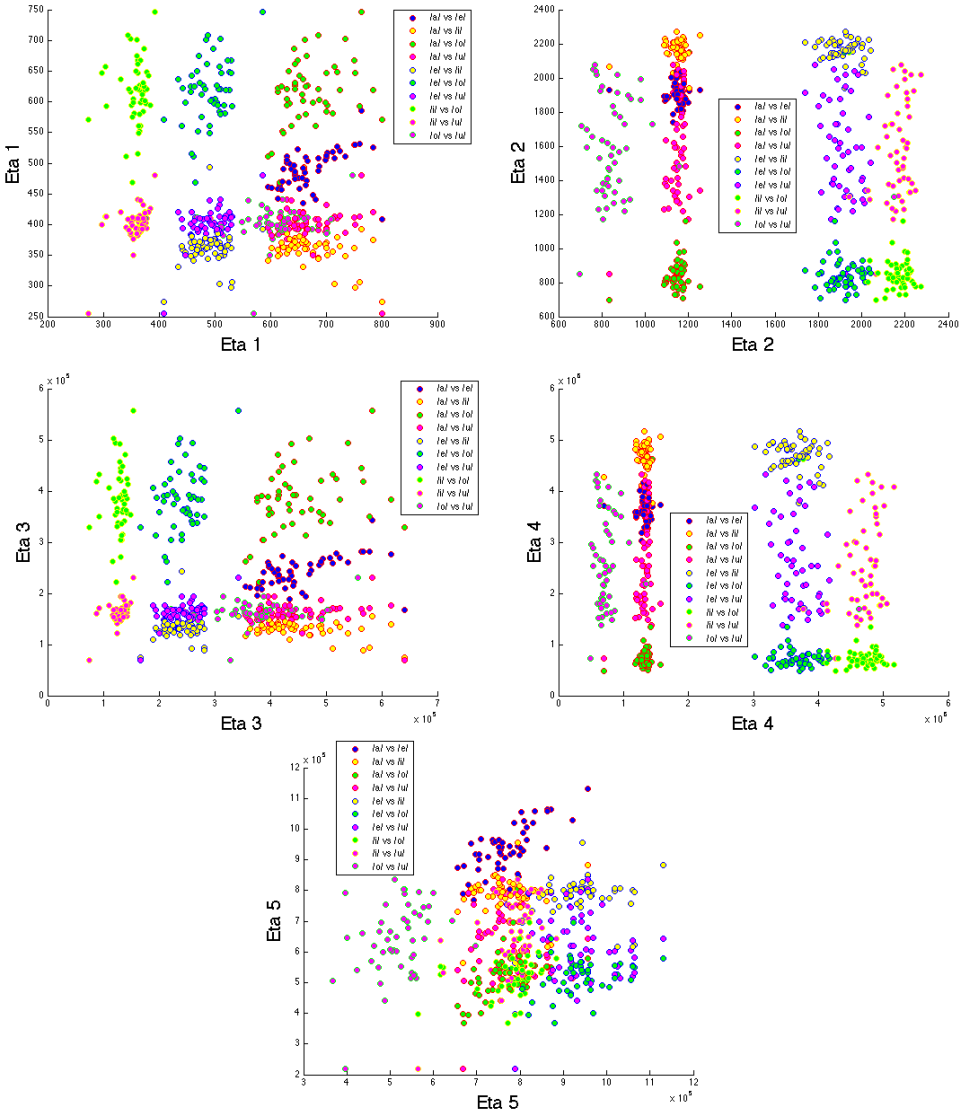


Figure 18.4: **Individual distribution of each pair of japanese five vowels plotted with the same element of η coordinates.** From top left to right bottom: Plot with η_1, η_2, η_3, η_4, and η_5. For example, the plot /a/ vs /e/ for η_1 coordinate is the plot using both the η_1 value of /a/ (horizontal axis) and /e/ (vertical axis). The data are from individual B.

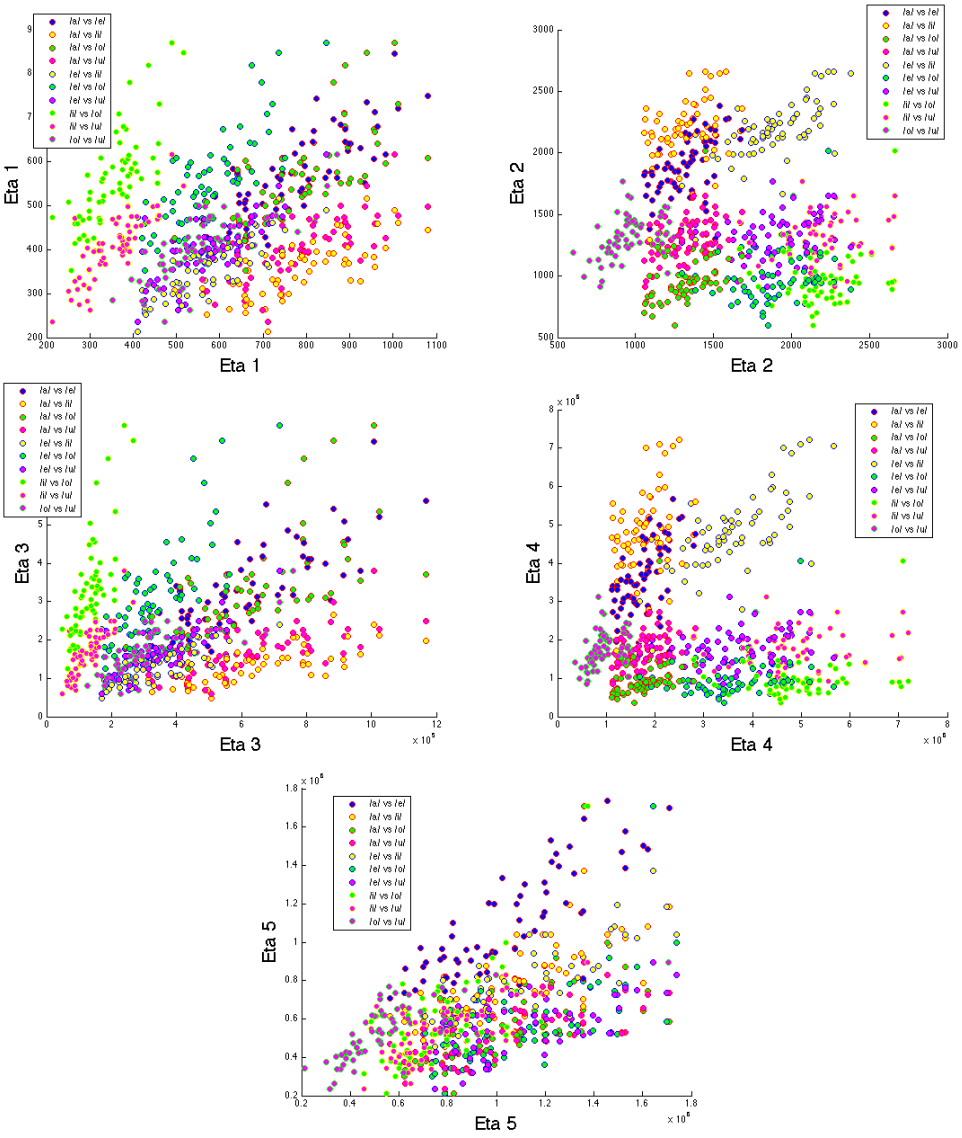


Figure 18.5: **Population distribution of each pair of japanese five vowels plotted with the same element of η coordinates.** From top left to right bottom: Plot with η_1, η_2, η_3, η_4, and η_5. For example, the plot /a/ vs /e/ for η_1 coordinate is the plot using both the η_1 value of /a/ (horizontal axis) and /e/ (vertical axis). The data combined both 26 male and 29 female samples.

which produces nothing but a positive correlation in individual distribution [141]. Since simple correlational analysis can not separate these components, we introduce the notion of *circumcenter* on dual-flat manifold to further test the existence of counterbalance between vowels.

In Euclidian space, the midpoint of two points is defined as their minimum equidistant point with respect to Euclidian norm. In the dual flat manifold of exponential family, we define the midpoint v_{ij} of two vowels v_i and v_j as the equal and minimum distance point with respect to m-divergence (KL divergence):

$$\underset{\eta^{\mathbf{v_{ij}}}}{\operatorname{argmin}}\{D(v_i||v_{ij})|D(v_i||v_{ij}) = D(v_j||v_{ij})\}. \tag{18.6}$$

Where$\eta^{\mathbf{v_{ij}}}$ is the η coordinates of v_{ij}, and $D(\cdot||\cdot)$ is the m-divergence (KL divergence) from the first to the second point in the argument. Such midpoint exists in the m-geodesic connecting v_i and v_j (Fig. 18.6).

The KL divergence minimum condition of v_{ij} can be proved with the use of Pythagorean relation by defining the e-flat geodesic $\theta^{\mathbf{e.l.}}$ as an equi- m-divergence line from v_i and $v_j$as follows:

$$D(v_i||\theta^{\mathbf{e.l.}}) = D(v_j||\theta^{\mathbf{e.l.}}), \tag{18.7}$$

where $v_{ij} \in \theta^{\mathbf{e.l.}}$. The line defined by $\theta^{\mathbf{e.l.}}$ is an orthogonal median line of the m-geodesic v_i-v_j with respect to Fisher metric and e-m- connection.

Then the following Pythagorean relations prove the KL divergence minimum condition of the midpoint v_{ij}:

$$D(v_i||v') = D(v_i||v_{ij}) + D(v_{ij}||v'), \tag{18.8}$$
$$D(v_j||v') = D(v_j||v_{ij}) + D(v_{ij}||v'), \tag{18.9}$$

where $v' \in \theta^{\mathbf{e.l.}}$. The minimum of $D(v_i||v') = D(v_j||v')$ corresponds to the condition that v' is in the m-geodesic between v_i and v_j, which coincides with v_{ij} (Fig.18.6).

The midpoint and orthogonal median line can be generalized for an arbitrary set of vowels $v_i, v_j, \cdots, v_k$ less than seven as the following $v_{ij\cdots k}$ and $\theta^{\mathbf{e.l.}}$:

$$\underset{\eta^{\mathbf{v_{ij\cdots k}}}}{\operatorname{argmin}}\{D(v_i||v_{ij\cdots k})|D(v_i||v_{ij\cdots k}) = D(v_j||v_{ij\cdots k}) = \cdots = D(v_k||v_{ij\cdots k})\}, \tag{18.10}$$

$$D(v_i||\theta^{\mathbf{e.l.}}) = D(v_j||\theta^{\mathbf{e.l.}}) = \cdots = D(v_k||\theta^{\mathbf{e.l.}}). \tag{18.11}$$

In case of the circumcenter of more than two vowels, the equidistant line is needed as an auxiliary line for numerical algorithm.

For geometrical reason, we call this $v_{ij\cdots k}$ as the circumcenter, and $\theta^{\mathbf{e.l.}}$ as the equidistant line of $v_i, v_j, \cdots, v_k$. The circumcenter of different number of vowel set represents different order of statistical dependency between vowels, which contains up to the equivalent order of principal components in identical η plot (Figs 18.4 and 18.5). Therefore, we call the circumcenter defined on n $(2 \leq n \leq 6)$ vowels as n-th order circumcenter. For example, the distribution of the 3rd order circumcenter

between /a/, /e/, and /i/ contains the information on the origin of 2nd and 3rd principal components in the identical η plot.

Examples of the results of the 2nd order circumcenter between two vowels are depicted in Figs. 18.7 and 18.8. The variance of the 2nd order circumcenter does not differ significantly with the average of the defining vowels. This means that there exist no significant pairwise counterbalance with respect to the informational distance(KL divergence) between different vowels fluctuation.

These results are negative for the existence of hidden negative correlation between vowels, though they contain some positive aspect concerning the origin of lognormal distribution. The decorrelation between vowels in both individual and population distributions assures the presence of uncorrelated dimension in multiplicative noise. Despite the positively correlated distribution in population sample, uncorrelated multiplicative process is still contained in both individual and population distribution. Furthermore, the m-geodesic connecting two vowels and their circumcenter coincides with the projection line of KL divergence between two vowels in previous chapter. The positive correlation was actually detected in population distribution, and can be easily introduced by varying the voice pitch in individual. The first principal components of these positive correlations generally differ from m-geodesic between vowels (Fig. 18.9). Therefore, the fluctuation of the circumcenter can be important to assure the statistical property of vowel system invariant of pitch variation. The decorrelation on the m-geodesic between different vowels strongly supports the origin of lognormal distribution of KL divergence as uncorrelated multiplicative noise.

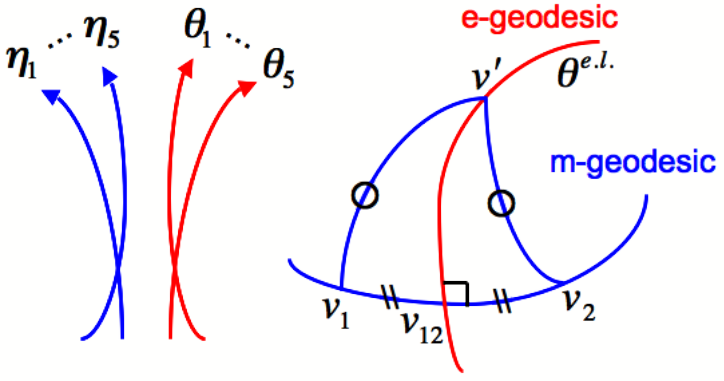


Figure 18.6: **Schematic representation of 2nd order circumcenter and equidistant line between two vowels.** The m-geodesic between two vowels v_i and v_j is orthogonal to the equidistant line $\theta^{\mathrm{e.l.}}$ which is e-geodesic. The circumcenter v_{ij} is situated at the intersection of these geodesics, and is the closest point from v_i and v_j on the equidistant line with respect to KL divergence.

18.5 Localization of Determinant $|S|$ in Vowels and Circumcenters

The distribution of the 2nd order circumcenter did not change significantly in variance with respect to its defining vowels. On the other hand, there exist significant

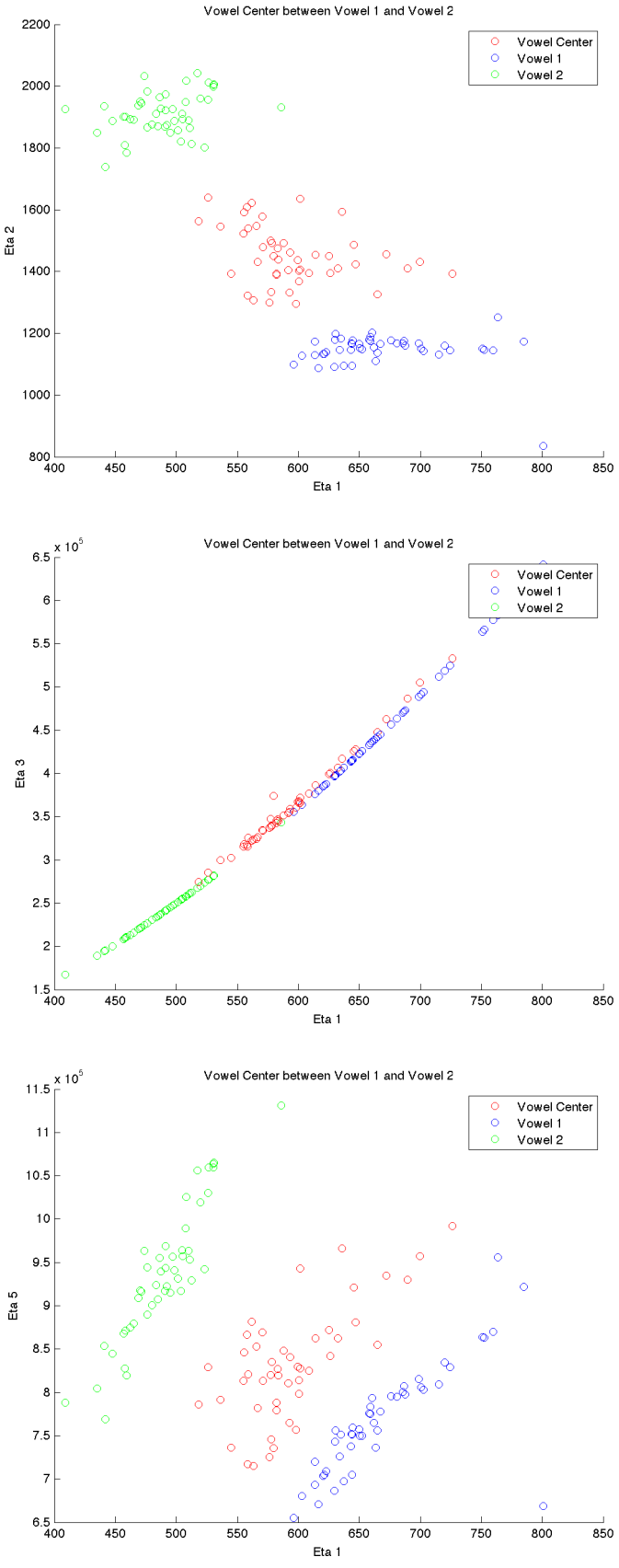


Figure 18.7: **2nd order circumcenter distribution between vowels /a/-/e/ of individual B in η coordinates.**

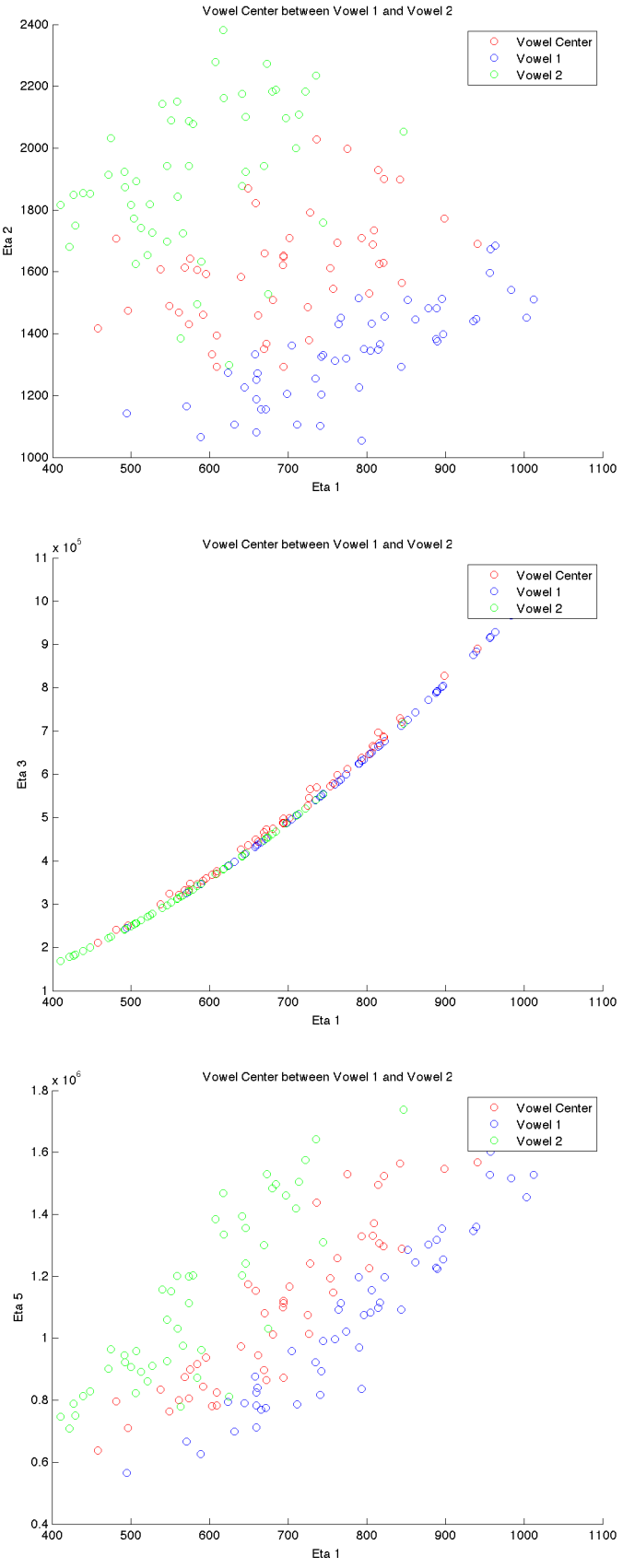


Figure 18.8: **2nd order circumcenter distribution between vowels /a/-/e/ of 26 male and 29 female in η coordinates.**

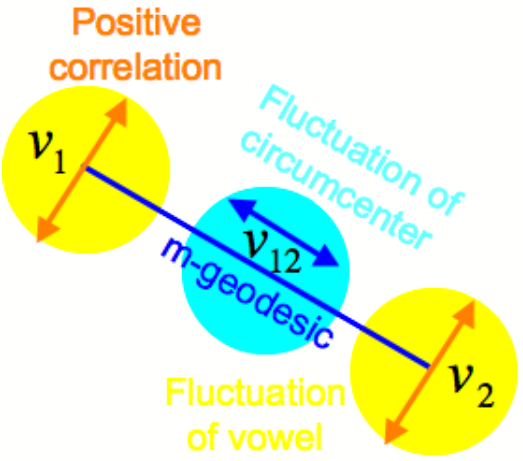


Figure 18.9: **Schematic diagram of the relation between positive correlation of two vowels and fluctuation of 2nd order circumcenter in η coordinates.** Vowel fluctuation contains positive correlation in population distribution, and can be introduce by voice pitch variation in individual distribution. The fluctuation of circumcenter reflects the components other than positive correlation, including the variation on m-geodesic where inter-vowel KL divergence was measured and followed lognormal invariant distribution. The decorrelation between the two vowels in the direction of the m-geodesic is observed as the non-decrease of the 2nd order circumcenter variance, which provides the uncorrelated multiplicative noise in KL divergence between vowels.

localization of circumcenter in θ coordinates. Figs. 18.10 show the examples of localization.

Such localization was dominantly under the effect of $|S|$ value, which is a common multiplicative factor of θ coordinates as follows:

$$\theta_1 = \frac{1}{|S|}(\sigma_{22}\mu_1 - \sigma_{12}\mu_2), \tag{18.12}$$

$$\theta_2 = \frac{1}{|S|}(\sigma_{11}\mu_2 - \sigma_{12}\mu_1), \tag{18.13}$$

$$\theta_3 = -\frac{1}{2|S|}\sigma_{22}, \tag{18.14}$$

$$\theta_4 = -\frac{1}{2|S|}\sigma_{11}, \tag{18.15}$$

$$\theta_5 = \frac{1}{|S|}\sigma_{12}. \tag{18.16}$$

Where $|S| = |\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{12}|$.

Figs. 18.11 and 18.12 show the distribution of $|S|$ value for each pair of vowels and their 2nd order circumcenter. The circumcenters are clearly distinguished from other vowels by $|S$ value in log scale. Figs. 18.13 show the color gradient expression of $|S|$ value projected in η_1-η_2 plane. The distribution of $|S$ value in Figs. 18.11 and 18.12 suggest the existence of a sea star shape barrier of $|S|$ value on the effective surface made of inter-vowel m-geodesics. Since there exist superposition between vowels and circumcenters in the projection, the color gradient interpolation are blurred at the frontier between vowels in Figs. 18.13.

Fig. 18.14 expresses schematically the sea star shape barrier of $|S|$ value. Such landscape of $|S|$ value can be explained by the geometrical composition of the five vowels in Figs. 18.1 and 18.2. The η coordinates of the vowels contain quadratic dependency between different axes, while the 2nd order circumcenter is defined on the m-geodesic, which is the first-order line connecting two vowels. This means the 2nd order circumcenter certainly deviate from the quadratic relation between vowels, as seen in Figs. 18.7(Middle) and 18.8(Middle). These deviations are the origin of the high $|S|$ value in circumcenter.

The high $|S|$ value is the result of unbalance between the variance parameters σ_{11}, σ_{22} and the covariance parameter σ_{12}. Indeed, the quadratic dependency between η coordinates of the five vowels has different scales in the variation of η_1 and η_2 values (Figs. 18.1 and 18.2 horizontal axes). The difference of these variations results in the variation of η_3, η_4 and η_5 values through quadratic dependency, which becomes the variance ratio between the parameters σ_{11}, σ_{22}, and σ_{12} in the 2nd order circumcenter: The variance ratio between σ_{11}: σ_{22}: σ_{12} is about $6 : 50 : 10$ in the individual distribution (Figs. 18.1 vertical axes), and $12 : 70 : 16$ in the population one (Figs. 18.2 vertical axes). Therefore, the linear deviation of the 2nd order circumcenter from the quadratic vowel surface is about 5 times larger in σ_{22} than σ_{12}, while σ_{11} remains close to σ_{12}. This causes the high $|S|$ value distribution in a sea star shape between vowels.

The 2nd order circumcenter exists between two vowels as their equidistant and nearest distribution with respect to the KL divergence. It is therefore the intermediate distribution of two vowels with respect to the form of distribution. The high $|S|$ value barrier between vowels may be a way to explain the quadratic localization of vowels in geometrical term: The intermediate vowels between the existing ones break the balance between the mean parameters and variance parameters of formant distribution, which contradicts to the invariance of the variance/mean component ratio α in previous chapter. The high $|S|$ value of the 2nd order circumcenter inversely suggests the role of the formants' variance/mean ratio in the recognition of vowels.

Following this hypothesis, the $|S|$ value may also be considered as the filtering function of vowels. This situation is similar to the color perception where it is possible to find filtering functions based on optical characteristics, which largely coincide with linguistic categorization of colors in many languages [142].

18.6 Conclusion

We investigated statistical dependency between five japanese vowels, as well as the property of the 2nd order circumcenter distribution in dual-flat coordinates. The results showed quadratic dependency between η coordinates, positive correlation between different vowels in population distribution, and uncorrelated fluctuation of 2nd order circumcenter which supports the origin of invariant KL divergence lognormal distribution as multiplicative noise. The observed high $|S|$ value in the 2nd order circumcenter distribution implies the role of variance-covariance parameters in vowel perception.

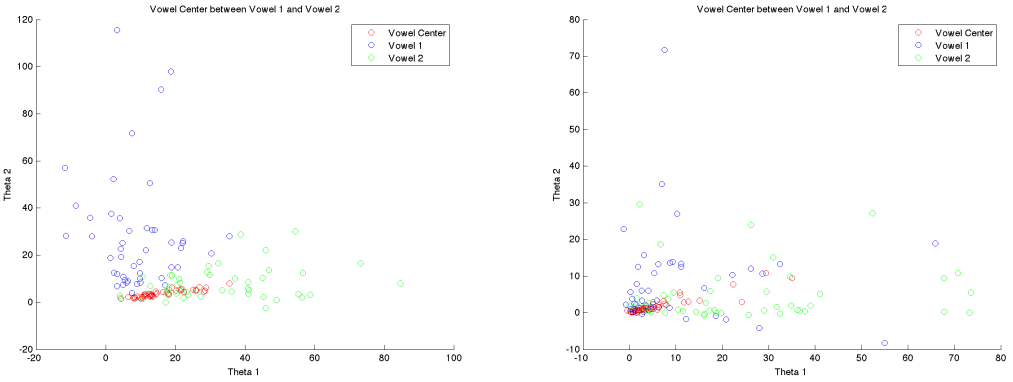


Figure 18.10: **Examples of localization of 2nd order circumcenter in θ coordinates.** Left: Plot of /a/(blue), /e/(green) and their 2nd order circumcenter(red) in θ_1-θ_2 plane of individual B. Right: Plot of /a/(blue), /e/(green) and their 2nd order circumcenter(red) in θ_1-θ_2 plane of 26 male and 29 female population.

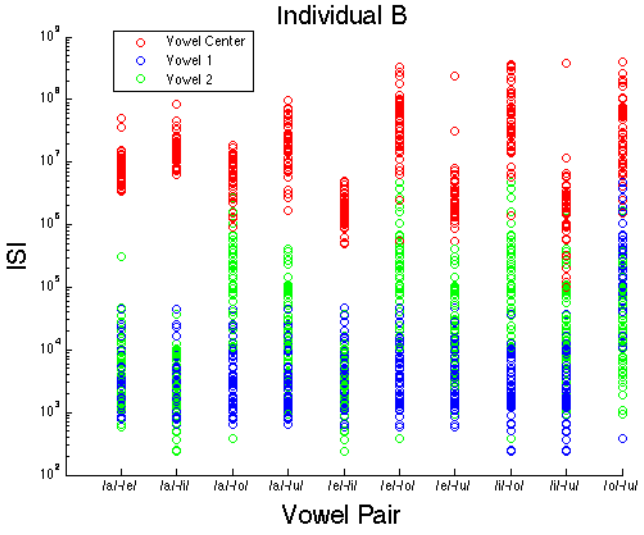


Figure 18.11: **$|S|$ value of five japanese vowels (blue and green) and their 2nd order circumcenter(red) of individual B.**

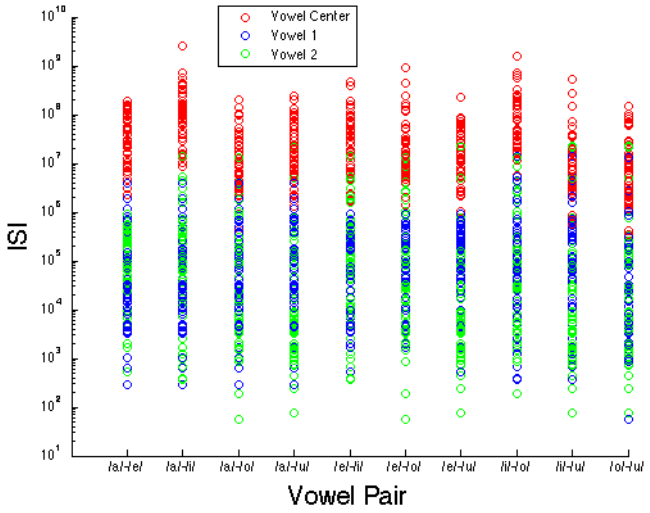


Figure 18.12: $|S|$ **value of five japanese vowels (blue and green) and their 2nd order circumcenters(red) of 26 male and 29 female.**

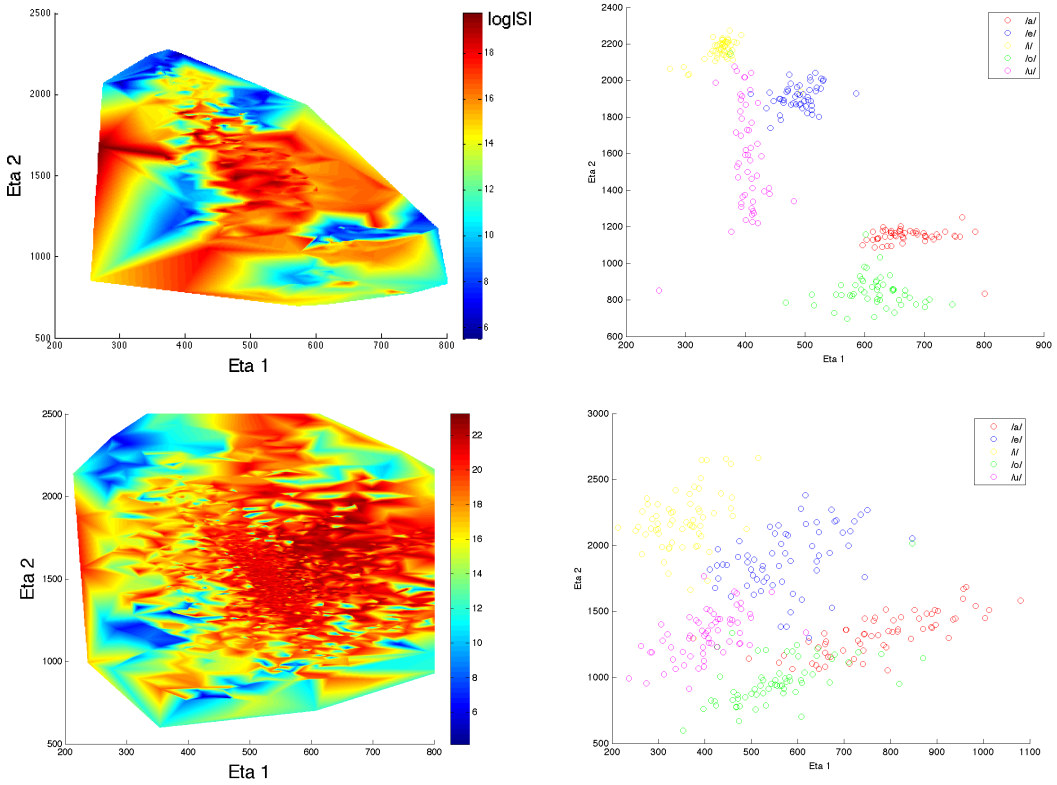


Figure 18.13: $|S|$ **value of five japanese vowels and their circumcenters projected as color gradient in η_1-η_2 plane.** Top 2: individual B. Bottom 2: 26 male and 29 female. The left figures show $|S|$ value as color gradient, and the right figures are the vowel distributions in η_1-η_2 plane. (The low $|S|$ regions in population distribution are blurred by the superposition of different vowels and their circumcenters.)

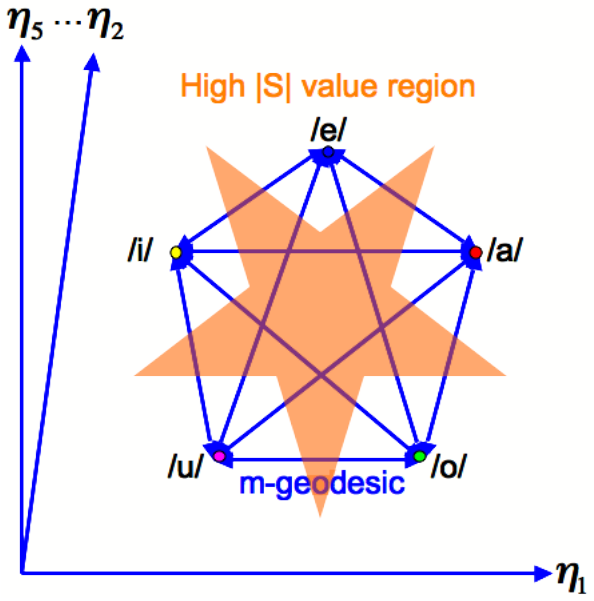


Figure 18.14: **Schematic representation of $|S|$ value landscape on the effective surface of inter-vowel m-geodesics.** The $|S|$ value is low around the five vowels, while becomes high around their 2nd order circumcenter.

Chapter 19

Methodology for Dialectic between Linguistic Theory and Mathematical Modeling and Its Application to Modeling Multilingual Environment with Contact Process

Abstract

We seek for a path to benefit a class of mathematical model to concretize and develop conceptual theory in ecological linguistics. We propose two steps of interaction both from linguistics side and modeling side: The first step is to appropriately define the model according to a linguistic theory by symbolizing the measurable primary variables and function. The second step is to perform analysis on the simulated model, and define the secondary variables to further expand the former linguistics theory.

An example of such dialectical modeling is demonstrated with the use of contact process, to model language acquisition in multilingual environment. The simulated model reproduced the spacio-temporal cluster dynamics of multilingual states with highly significant correlation among languages. The relative instability of bilingual states are also observed. The results support the working hypothesis of multilingual activity in Hippo Family Club.

Keywords: Contact process, Multilingual environment, Multilingualism, Natural acquisition of language, Ecology

Methodology: Formulation of dialectical strategy between linguistic theory and mathematical modeling compatible to modeling with dynamical system and analysis with information geometry → Modeling of multilingual environment with contact process as a stochastic dynamical system → Tentative correlation analysis of simulated contact process with information geometry → Theoretical development of a probability propagation analysis of the contact process model with dynamical system and

information geometrical perspectives

19.1 Preface: Context of work

This chapter is realized in a working group "Groupe de travail sur l'approche écologique en linguistique et en anthropologie (GdT- ELA)", at the Institute of Complex System Paris-Île de France (ISC-PIF), during the academic year 2007-2008.

The project aims to develop an interdisciplinary field situated principally in the interaction between linguistics, ecology, philosophy, anthropology, and mathematical modeling. The goal is not to complete an empirical work with already established academic framework, but to seek for a new methodology to break the barriers between separated fields, and to form novel transdisciplinary questions. This perspective inevitably shares some of important aspects with complex systems sciences, especially in the concept of emergent patterns in interacting population dynamics such as "écotone" proposed in ecology [143]. Similar concept in the organization of cognition and action also exists in linguistic anthropology, as a mutual elaboration of linguistic and gestural modalities to animate the discourse in individuals [144].

This work is realized with constant discussion and contribution of the members with different academic backgrounds, often contradicting in their proper terminology and methodology. The results put emphasis on the proposition of novel interdisciplinary methodology than the presentation of the findings. This work is also supported by the Lex Institute (the Institute for Language Experience, Experiment, and Exchange) in Japan, both to establish a working hypothesis in multilingual language acquisition and to collect the questionnaire data.

19.2 Introduction

The collaboration between physics and linguistics has been difficult in the way linguistic phenomena include human. The relativity of observation required to treat human-integrated systems destroys the fundamental principle of measurement in physics, which is based on isolating the observables from other interactions. The remarkable encounter between the celebrated linguist and physicist in 1958 ended up with the conclusion that declines the equivalence between linguistics and other "exact" sciences in term of the scientific precision [145].

The difficulty also lies in the indifference that separates the literacy forming each culture. Especially, there exist mutually exclusive dependencies in both ways around mathematics, completely exclude it or utilizes it as fundamental description of the nature [146].

Although, in linguistics in a large sense, there exist also interdisciplinary domains including physics of sound, neuroscience, mathematical modeling in various levels of syntax, etc [121][147][148]. Especially the newly emerging *complex systems science* aims to promote the interdisciplinary interactions to find the descriptive parameters specific to or universal in a variety of strongly and nonlinearly correlated systems. This has become possible only recently with the development of the computation technology with which we are able to simulate rapidly large-scale dynamical systems.

Before the innovation of the high-speed computation, mathematical description was limited to give linguists the observed facts and its mathematical orders for further linguistic consideration. Some analytically developed models also provided the abstract level of description that can be interpreted as essential structures of the phenomenon [149].

With the invention of simulation technology, we have become accessible to a large class of mathematical models that can incorporate conditions supporting linguistic plausibility. Here, the compensation principle of Niels Bohr may still find its way in the way of cooperation between mathematical modeling and linguistic theory.

In this article, we propose a dialectical way of interaction between mathematical modeling and linguistic theory, to overcome the operational limit of usual linguistics. We first try to implement the ecological approach in linguistics into mathematical expression. Next, we try to derive new variables from the simulated phenomenon, that can expand linguistic concepts to reinterpret the wider reality.

19.3 Minimum Requirement for Ecological Approach in Linguistics

The ecological approach assumes the systems where variables and their interactions are inseparable. The variables are the symbolized quantity which we can measure from real phenomena. Though, isolated observation of a single variable is not sufficient for ecological problem setting. We are interested in how the variables are affected by the interactions, and how the interactions change according to the variables. The mutual spiral of the feedback between the variables and the interactions is the irreducible unit to perceive the reality in this approach.

The variables and the interactions in issue generally range quite widely including nonlinear forms, time dependencies, etc. We have abstracted the minimum requirements for plausible mathematical models that can serve for ecological linguistics. The conditions consists of three principles:

1. The model consists of variables and functions as interactions among variables.
2. The variables and the interactions depend on each others.
3. Existence of time axis.

Possible candidate class of model can be expressed as stochastic dynamical system in the following general forms,

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t), \sigma), \tag{19.1}$$

in continuous time system, and

$$\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t-1), \sigma), \tag{19.2}$$

in discrete time system in case of order one. The $\mathbf{x}(t) = (x_1(t), \cdots, x_n(t))$ is the n-dimensional variable at time t, and the function F define the interactions. σ is the stochastic variable in case to model unavoidable noise or to adopt stochastic model such as contact process. Hence, this class of model includes the classical Lotka-Volterra equation in mathematical ecology [150] [151].

19.4 Methodology for Dialectic between Linguistic Theory and Mathematical Modeling

We define the above defined variables $\mathbf{x}$ as *primary variables*, which are the symbolization of observables in reality. The *primary variables* are not necessary the observables themselves, but should incorporate parameters that can be measured from real situation. Once the model is properly established, it becomes possible to perform a series of analysis including simulation on this model. The defined class of model is known to produce a variety of complex dynamics including chaos and percolation [152][153]. Both analytical and numerical analysis further reveal the nature of the established model, which belongs to the interest of physics. This analysis depending purely on logical operations leads us to consider on the simulated dynamics of the model, and gives us the idea to recode them with different symbols according to linguistic interests. We call such newly defined symbols on the model dynamics as *secondary variables*. For example, several statistics of simulated time series are candidates of the secondary variables. We express the secondary variables as the function of the primary variables $\mathbf{V}(\mathbf{x})$. The secondary variables are not necessary the observables in reality, but are derived from the combination of the linguistic conceptualization and mathematical analysis. This does not intend to create the reality in the model, but to help further hypothesis forming to expand the former linguistic theory, aiming to treat a wider and more profound range of reality. Such interaction is dialectical in the sense both mathematical modeling and linguistic consideration are supporting the novel concepts attributed on secondary variables, and the meaning is only defined in the cross-reference. The expanded theory resets the stage of discussion in purely linguistic framework, and starts in turn to seek for actual phenomena to evaluate the new hypothesis. The conceptual scheme of this strategy is depicted in Fig. 19.1.

19.5 Modeling Multilingual Environment with Contact Process

We take in this section a concrete example in sociolinguistics to realize the insisted interaction between linguistic theory and mathematical modeling. Note this is an on-going subject and we are limited in this article to give a progress report.

19.5.1 Contact Process with Interacting Variables for Modeling Language Acquisition in Multilingual Environment

Language acquisition is one of the important issues considering the human being in its interacting population. Most humans communicate via oral language, and the language faculty is one of the definitions that best differentiate human beings from other animals.

Though, it is widely known that language itself is not a genetically programmed characteristics. It is acquired from the environment in which other humans are communicating with the use of it. The inability to acquire human language of feral children is an incidental example that revealed the language environment is necessary for the first language acquisition [154]. Languages do not generate spontaneously in

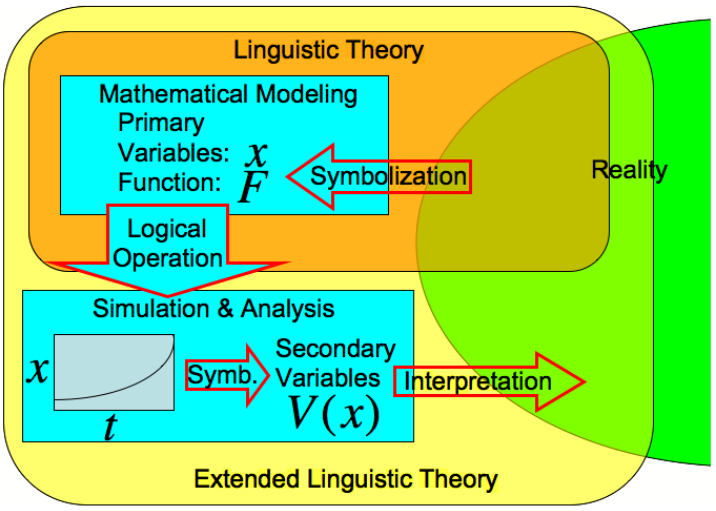


Figure 19.1: **Conceptual scheme of the dialectic between linguistic theory and mathematical modeling.** Based on a linguistic theory, we symbolize a part of linguistic reality as primary variables and functions to form a mathematical model which satisfies the minimum requirement. Then, the logical operation on this model enable us to perform simulation and its analysis, which have the possibility to overcome conceptual and operational limit of former linguistic theory. The result of the simulation and the analysis bring us the inspiration to define secondary variables, which can be used to reinterpret the reality with an extended theoretical framework.

a short time period from isolated individual, but rather maintained and transmitted between individuals, evolving its structure. In this sense, the acquisition of language in human population has a similar property with the transmission of epidemic disease [155].

Since the acquisition of language in each individual is inseparable with the interactions with other individuals, the subject typically requires the ecological approach to treat the phenomenon in its synthetic entity. When considering the language acquisition, we need to reconsider what the language environment is. Although there exist politically labeled official languages in each nation, actual dynamics of languages propagation cannot be necessary encapsulated in political system, and is reported to be based on multilingual state where plural languages are in constant competition [156].

Such situation can be abstracted with the use of the contact process model, which we adopted to model language acquisition in multilingual environment. The contact process is also used as a model of epidemic disease propagation, communication model on the web, trend model in sociology, etc [157].

We define the contact process with interacting n-dimensional binary variables $\eta(j) \in \{0,1\}^n$ on one-dimensional circle lattice with N nodes ($1 \leq j \leq N$). Here, the state $\eta_i(j) = 1$ represents that the node j is able to communicate in i-th language, while $\eta_i(j) = 0$ is the lack of competence. The node can represent an individual or a group of homogeneous language profile. If $\eta_1(j) = \eta_2(j) = 1$ and $\eta_3(j) = \cdots = \eta_n(j) = 0$ then the person j is bilingual of the first and second languages in the list. Network topology can further incorporate realistic contact structure, but is limited to the circle lattice in this article for basic analysis. The model can be formulated

using the generator Ω as follows [153]:

$$\Omega f(\eta) = \sum_{j \in \{1,\cdots,N\}} \mathbf{c}(j,\eta)[f(\eta^j) - f(\eta)]. \tag{19.3}$$

Where the variables η^j represent $\eta^j(k) = \eta(k)$ for $(1 \leq j \neq k \leq N)$, and $\eta^j(j) = 1 - \eta(j)$ for $(1 \leq j = k \leq N)$. $f(\cdot)$ is any two bounded increasing function on the real line. The transmission rate c_i of the i-th language for the node j and the variables η is given by

$$\mathbf{c} = (c_1, \cdots, c_i, \cdots, c_n), \tag{19.4}$$

$$c_i(j,\eta) = (1 - \eta_i(j)) \times \lambda \sum_{|k-j|=1 \ (mod\ N)} \frac{\eta_i(k)}{2} \times \frac{Poly\,(R)}{R} + \eta_i(j), \tag{19.5}$$

$$R = \frac{\sum_{|k-j|=1 \ (mod\ N)} \sum_{1 \leq i' \leq n} \eta_{i'}(k)}{2}. \tag{19.6}$$

The weighting function $Poly(R)$, its regularization factor R and parameter λ $(0 \leq \lambda \leq 1)$ define the transmission probability of languages to the j-th node in the next time step. To incorporate the interactions between languages in multilingual environment described in the following sections, we define $Poly(\cdot)$ as polynomial function for the accessibility to data fitting. Note that the elimination rate is fixed to 1 in contact process, therefore the time step does not necessary corresponds to the actual generation of human population. This assumption does not change qualitative feature of the model.

In this model, the variables η correspond to the *primary variables* $\mathbf{x}$ in equation (19.2), defined with the parameters $\frac{Poly(R)}{R}$ measurable from the questionnaire (See sections 19.5.3, 19.5.4, 19.5.5).

19.5.2 Multilingual Effect on Natural Language Acquisition in Experimental Activity at Hippo Family Club

The multilingual activity of Hippo Family Club started in Japan to overcome the limit of analytical linguistics in practical acquisition of languages ranging in all age groups. Language educations based on the combination of reduced units such as vocabulary, grammar, pronunciation, etc., tend to tear apart the integrity of the growing language in individual in the name of "foreign" language, which attenuate the learning motivation far before reaching one's actual capacity [158]. This can be considered as a typical example in complex systems that the induction from segmented levels of observation does not allow us to deduce the synthetic rules of the entire phenomenon. Natural acquisition of plural languages in multilingual countries, on the other hand, are usually associated with daily practical use, where the achievement of communication is the primal task. The inductive units of languages emerge through unconscious trial and error process, which is the same natural path as the newborn baby's first language acquisition.

In Hippo Family Club, there exist no teachers, no classes, no tests. Members participate to the so-called "Family activity" where they play, sing and dance together in the immersion of 19 languages, by imitating the registered sound [159].

This activity is inspired by the founder (Yo Sakakibara) from the ordinary scene in Luxembourg, where children in the park play in several languages [160]. It is further developed through experiences of accepting non-japanese people from over 128 countries in members' houses in the form of homestay. Besides the Family activity, members also have occasions to participate in short-time homestay programs in 21 countries, not to learn the languages but simply to become friends with people who speak these languages. One of the main objects in this program is to have the experience of speaking spontaneously without logically considering whether one is making mistakes.

Through these activities in Hippo Family Club, we have come to discover the fact that people who speak more than one languages tend to acquire more rapidly new languages. This suggests the existence of multilingual effect on language faculty, that the acquired languages reinforce the acquisition of new languages.

Several members of Hippo Family Club also insist that intermittent exposure to three or more languages of different competences help us to relativize them, clear away the mental barrier to face them, and subjectively facilitate the acquisition.

These collective experiences lead us to establish the two working hypothesis:

1. The multilingual environment may facilitate simultaneous plural languages acquisition.
2. Multilingual person may have greater capacity and efficiency to acquire new languages.

In this article, we consider the first hypothesis for modeling language acquisition in multilingual environment.

19.5.3 Investigation on Multilingual Environment in Singapore

We investigated the multilingual environment in Singapore by means of the questionnaire concerning the language environment and acquired languages of the respondent. The form of questionnaire is shown in Appendix Fig. 19.12.

Singapore is a multilingual country where there exist four official languages with equal status, Mandarin, Malay, Tamil, and English. Language education has been strongly promoted by the People's Action Party, which has been the ruling political party since 1959. They insist the "bilingual policy" to manage ethnic diversity, and expect people to acquire at least two languages including English. At the same time, equality of three major ethnic groups, the Malays, the Chinese, and the Indians, is strongly controlled with the principle of "multi-racialism", so that each population enhance their distinctive cultures, languages, heritages and values [161] [162] [163]. The situation gives us an idealistic field of investigation where people are exposed to four distinctive languages with writing systems in daily life. Besides the official languages, there exist local variants in each language. For instance, people originated from the south part of china tend to speak in Cantonese, which is a dialect of chinese languages, but phonetically quite different from Mandarin. We also reflected such variations in the questionnaire.

We obtained the data from 30 volunteers living in Singapore. As every person cooperated to the questionnaire spoke more than one languages, we combined the results of additional 8 Japanese for monolingual data. We will investigate the result with the proposed linguistics-modeling interaction in the following sections.

Since the sample number is not sufficient to establish an empirical model with accurate statistics, we are limited to give some hypothetical analysis based on some typical cases of simulation results.

19.5.4 Competition Model: Linear-Response Acquisition Hypothesis

We define here the exact form of the transmission rate defined in equation (19.5). Let us assume the total amount of information rate I_T that exists in multilingual environment with l_{env} languages. The I_T is a measure of language stimulus that can serve for the acquisition, represented in one-dimensional quantity. This is for example the amount of time one spend to listen, speak, read, and write in a day. Then the mean amount of information exchanged for each language is $\frac{I_T}{l_{env}}$. If we judge the acquisition of a language by putting certain threshold in its competence, we can assume the existence of the mean critical information I_C to acquire a language. This represents the amount of experience necessary to acquire a new language. Note that the unit of I_T and I_C is not important in this approach. It cannot even be measured properly in real situation. What is important is the ratio between I_T and I_C that determines the mean acquisition rate out of l_{env} languages. This is an invariant structure of the model as long as languages competitively divide the information holding additive axiom, regardless of its way of measurement.

If we assume that each of l_{env} languages equally divide the total information I_T, there exist critical value n_C of l_{env} where $\frac{I_T}{l_{env}=n_C}$ becomes less than I_C. This situation means that the individual is not able to acquire any language due to the lack of information. At this point, it is not realistic because such incompetence of language acquisition does not occur in real situation. The distribution of I_T is not necessary equal among coexisting languages, and anyone can acquire at least one language more or less frequently used around him or her. Therefore, we derive the fluctuation around the mean divided information $\frac{I_T}{l_{env}}$ with normal distribution $N(\mu, \sigma)$, where $\mu = \frac{I_T}{l_{env}}$ is the mean value and σ is the variance. Since the change of σ is qualitatively replaceable to the corresponding scale transformation of I_T and I_C that maintain the I_T/I_C ratio invariant, we fix the variance as $\sigma = 1$.

With such fluctuation, the mean acquisition rate l_{acq} in the multilingual environment with l_{env} languages can be estimated as follows.

$$l_{acq} = l_{env} \times \int_{I_C}^{I_T} \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(z - \frac{I_T}{l_{env}})^2}{2\sigma^2}\right) dz. \tag{19.7}$$

The calculation results with realistic parameters is plotted in Fig. 19.3. The following linear relation is observed in realistic situation $l_{env} \geq 1$.

$$l_{acq} \propto l_{env}. \tag{19.8}$$

We call this relation as linear-response acquisition hypothesis, based on the purely competitive interaction between languages. The corresponding contact process model is with the linear interaction $Poly(l_{env}) = p_1 l_{env} + p_2$ $(p_1, p_2 \in \mathbf{R})$ and the regularization factor $R = l_{env}$ as the parameters of *primary variables* in equation (19.5).

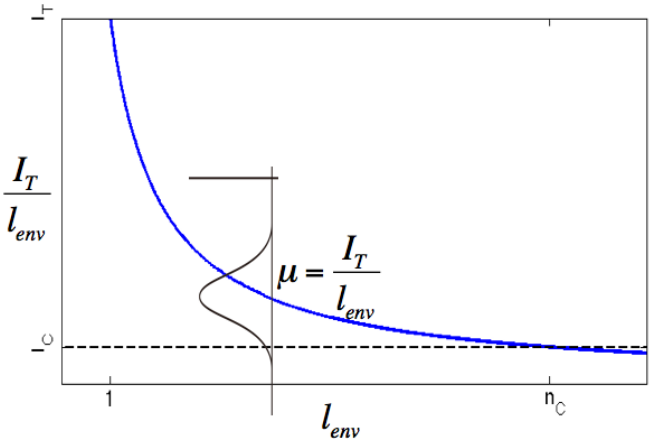


Figure 19.2: **Divided amount of information for each language in competition.** Horizontal axis: Number of languages l_{env} in the environment, Vertical axis: Mean information $\frac{I_T}{l_{env}}$ divided to each language. Additional fluctuation of information distribution was defined with the normal distirbution $N(\mu, \sigma)$ where $\mu = \frac{I_T}{l_{env}}$.

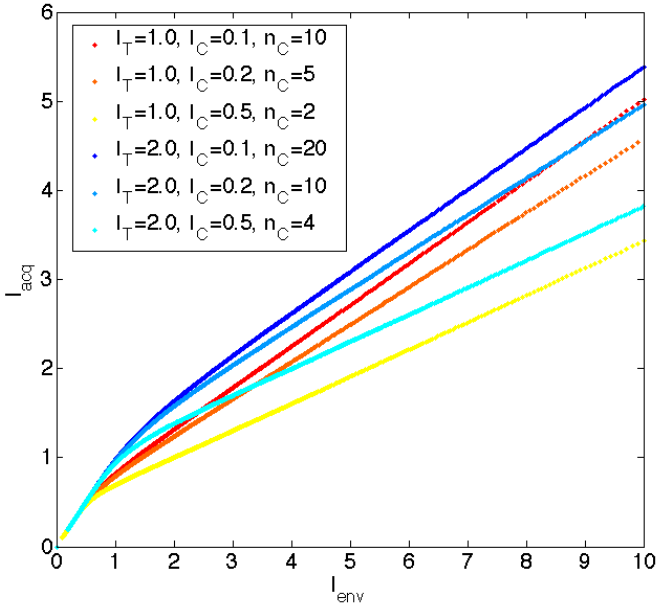


Figure 19.3: **Estimated acquisition rate of languages in competition model.** Horizontal axis: Number of languages l_{env} in the environment, Vertical axis: Expecting number l_{acq} of acquired languages. l_{acq} is proportional to l_{env} in a wide range of realistic parameters.

19.5.5 Positive/Negative Interaction Model: Non-Linear Effect of Multilingual Environment on Language Acquisition

We test the formerly developed linear-response hypothesis with the use of actual data, and consider a more plausible model as an expansion.

From the questionnaire, we calculated the number of languages l_{env} that existed in the respondent's ordinary language environment. We also define the acquired number of languages l_{acq} to investigate the acquisition rate from the environment. The precise definition of the language environment and the acquired languages for data extraction are described in Appendix.

Acquisition rate of actual data is plotted with linear and polynomial fitting in Fig. 19.4. Actual data generally conserved the linear relation, supporting the competition model. On the other hand, augmenting the order of polynomial, there exist slightly positive mean effect in $3 \leq l_{env} \leq 5$ region, and negative effect in $l_{env} = 2,\ 6$ regions. Such deviation from linear-response model can be considered as the positive and negative non-linear interaction effects on the acquisition between competing languages. We call this acquisition curve as positive/negative interaction model or simply as interaction model. We chose the polynomial of order 5 for the fitting, since it is the minimum order that assures the minimum squared error fitting for 6 values of l_{env}.

The positive/negative interaction model can be formulated as contact process using the following definition of the parameters of *primary variables* in equation (19.5).

$$Poly(l_{env}) = p_1 l_{env}^5 + p_2 l_{env}^4 + p_3 l_{env}^3 + p_4 l_{env}^2 + p_5 l_{env} + p_6, \tag{19.9}$$

$$R = l_{env} = \frac{\sum_{|k-j|=1 \ (mod\ N)} \sum_{1 \leq i' \leq n} \eta_{i'}(k)}{2}, \tag{19.10}$$

where $p_1, \cdots, p_6 \in \mathbf{R}$ are the coefficients of the fitting polynomial.

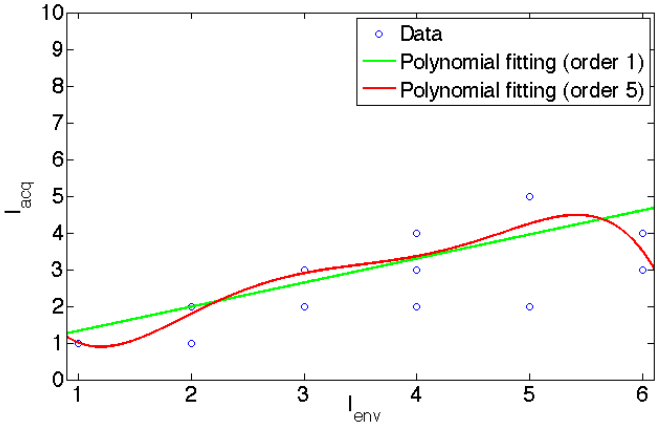


Figure 19.4: **Acquisition rate of languages in multilingual environment.** Horizontal axis: l_{env}. Vertical axis: l_{acq}. The blue circles are the results of the questionnaire in Singapore with monolingual data of Japan. The green and red lines are the fitting with first- and fifth- order polynomials, respectively.

19.5.6 Simulation Result

We investigate the dynamics of the established models by simulating the contact process. To focus on the percolation state where languages prevail in the population, we fix the model parameter $\lambda = 0.3$ in equation (19.5).

The dynamics of the competition model and the positive/negative interaction model are depicted in Fig. 19.5 and Fig. 19.6, respectively. In each case, the evolution of spatio-temporal clusters representing multilingual states (for example, nodes with $l_{acq} \geq 3$) are observed. The clusters tend to be more distinctive in the positive/negative interaction model. This implies the effect of non-linear interactions to aggregate the multilingual state.

Since the relative stability of multilingual state and instability of bilingual state are reported through Hippo activity, we also simulated the interaction model with the bilingual initial condition in Fig. 19.7. The bilingual states are shown to be extinguished after falling into monolingual states. This dynamics qualitatively fit to the conflict between two languages in bilingual states and the associating inhibition of the acquisition [158].

To analyze the statistical feature of the modeled dynamics, the histograms of l_{acq} of the models are shown in Figs. 19.8 and 19.9. Both models express equivalent appearance rate in $l_{acq} \leq 4$ states, except the relatively high bilingual and trilingual states in the interaction model. The uniformity of distribution in the histograms imply the existence of correlations in high orders of interaction.

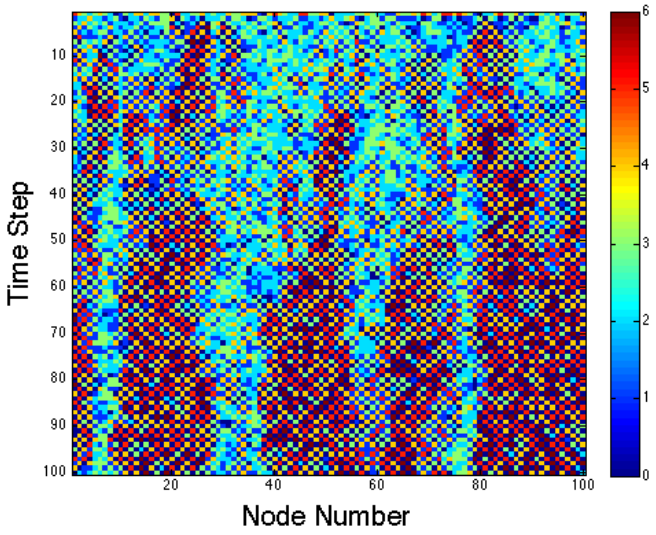


Figure 19.5: **Dynamics of the competition model.** Horizontal axis: Node number $1 \leq j \leq N$. Vertical axis: Time step. The color indicates the value of l_{acq}. $N = 100$. Initial condition was taken randomly.

19.5.7 Analysis of Order-wise Correlations

To detect the degree of interactions in simulated process, we calculated the order-wise correlations among languages with χ^2-value based on [77]. We simply analyzed the correlations among the first four variables (x_1, x_2, x_3, x_4) summarized for all

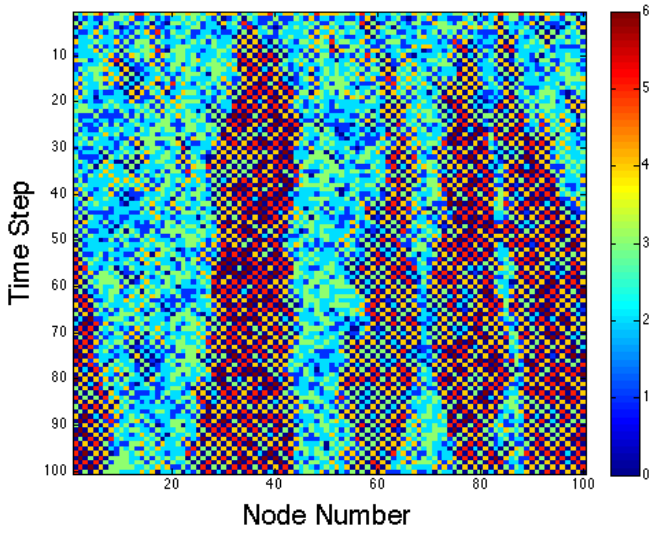


Figure 19.6: **Dynamics of the positive/negative interaction model.** Horizontal axis: Node number $1 \leq j \leq N$. Vertical axis: Time step. The color indicates the value of l_{acq}. $N = 100$. Initial condition was taken randomly.

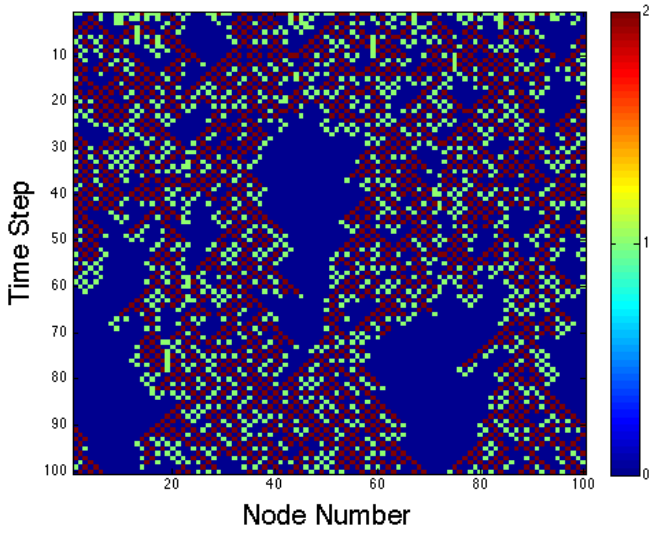


Figure 19.7: **Dynamics of the positive/negative interaction model with bilingual initial condition.** Horizontal axis: Node number $1 \leq j \leq N$. Vertical axis: Time step. The color indicates the value of l_{acq}. $N = 100$. Initial condition was taken randomly for η_1 and η_2.

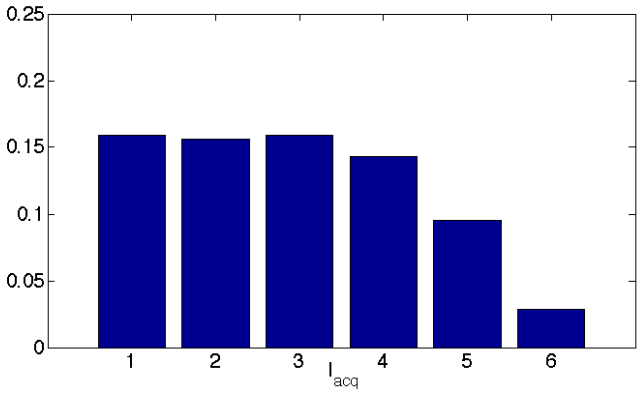


Figure 19.8: **Histogram of l_{acq} in competition model.** Horizontal axis: l_{acq}, Vertical axis: Occurrence number divided by the node number $N = 100$ and total time step 100.

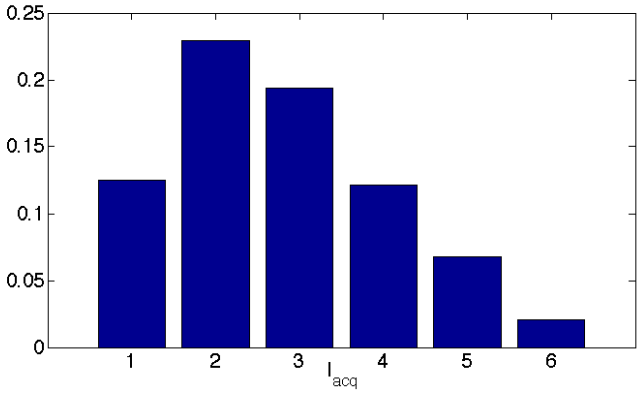


Figure 19.9: **Histogram of l_{acq} in positive/negative competition model.** Horizontal axis: l_{acq}, Vertical axis: Occurrence number divided by the node number $N = 100$ and total time step 100.

nodes, ignoring x_5 and x_6. This statistics is an example of the *secondary variable* introduced in the methodology section.

The results are shown in Figs.19.10 and 19.11. In all orders of correlations, there exist significance with respect to the 5 percent significance level of χ^2 test. Such state can be considered as the product of transmission through contact. Remarkably, the fourth-order correlation in the interaction model is more significant than some of the third- and second-order correlations. The spontaneous transition to multilingual activity in Hippo Family Club experience may relate these statistics, where simultaneous acquisition of plural languages was significantly promoted with the presence of four languages [158]. If we consider the direct analogy of the simulated results, it implies that such spacial accumulation of languages is one of the emergent property arising from the contact relation. Indeed, Hippo activity emphasizes to play with language sounds, together with other people, which realizes the contact relation taking place in natural acquisition [160].

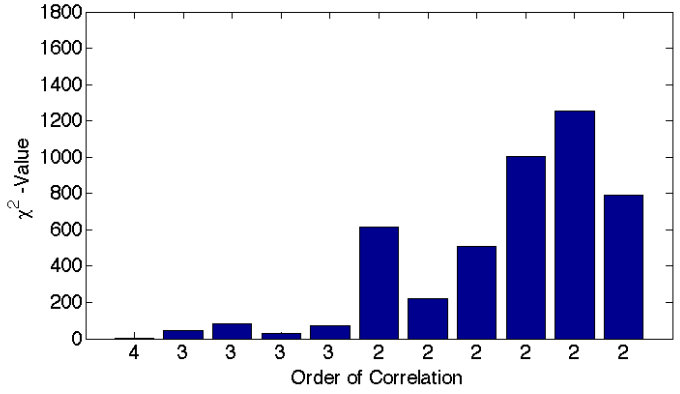


Figure 19.10: **χ^2-values of order-wise correlations in competition model.** Horizontal axis: Each combination of correlation represented simply with the order. Vertical axis: χ^2-value (positive/negative sign imply the positive/negative correlation, respectively). The 5 percent significant level of χ^2 test is 3.841.

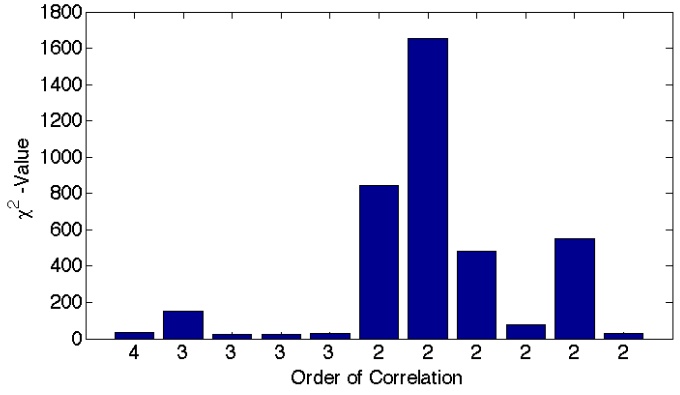


Figure 19.11: **χ^2-values of order-wise correlations in positive/negative interaction model.** Horizontal axis: Each combination of correlation represented simply with the order. Vertical axis: χ^2-value (positive/negative sign imply the positive/negative correlation, respectively). The 5 percent significant level of χ^2 test is 3.841.

19.6 Further Consideration

19.6.1 On working Hypothesis of Multilingual Acquisition

The models can further be expanded to incorporate the second working hypothesis: non-linear effect of multilingual person on language acquisition. For this purpose, it suffices to expand the model, introducing the axis of previously acquired language number of each node which positively modifies the transmission rate **c**. We need however much more empirical data to fit this model.

Since contact relation is estimated to be essential for the genesis of multilingual state, network topology should be further considered including the implementation of a real social network structure. The visiting dynamics of Hippo members to Family activities should be investigated. If the model can propose optimal contact frequency between individual to maximize the propagation of multilingual state, it may be possible to apply it in the way of organizing Family and foreign exchange activities at a collective level.

19.6.2 Theoretical Consideration

The break of symmetry in each order of the correlations may also relate the mechanism of transmission through contact. The nature of the established models is also an important issue, such as the classification of the universality class.

The contact process is mostly studied with variables without non-linear interaction. The overall transmission rate is simply calculated by linearly summing the single transmission rate according to the state of adjacent nodes. However, recent study in epidemiology reports the existence of non-linear interaction between variables which affect the transmission rate [164]. Our model gives a standard framework to treat such interaction with contact process. In other words, the positive/negative interaction model incorporates non-linear interaction between variables in classical SIR model. We add some theoretical remarks on the stability analysis of the steady state in probability propagation.

Let us consider the probability propagation of the n-node contact process with 2 sets of n-dimensional variables, namely $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$, which interact non-linearly with each other. The transition probability function matrix $\mathbf{T}$ is formally defined as follows:

$$\mathbf{T} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{T}_\mathbf{x} \\ \mathbf{T}_\mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix}, \tag{19.11}$$

$$\begin{bmatrix} \mathbf{T}_\mathbf{x} \\ \mathbf{T}_\mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{y}_{t+1} \end{bmatrix}. \tag{19.12}$$

Where t is the discrete time.

Then the k-periodic steady state $\mathbf{x}_1, \cdots, \mathbf{x}_k$ and $\mathbf{y}_1, \cdots, \mathbf{y}_k$ of the probability propagation $\mathbf{T}$ is defined as follows for $1 \leq t \leq k$:

$$\begin{bmatrix} \mathbf{T}_\mathbf{x} \\ \mathbf{T}_\mathbf{y} \end{bmatrix}^k \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix}. \tag{19.13}$$

Such steady state is under constraint of network structures between nodes independently defined for $\mathbf{x}$ and $\mathbf{y}$. For example, two interacting epidemics can have different transmission modalities such as vector-borne infection of malaria and sexual transmission of HIV.

The stability of the steady state is compatible to linear stability analysis, and the Lyapunov exponents can be derived from the product of Jacobian $\mathbf{T}'$. If we define the perturbation of $\mathbf{x}$ and $\mathbf{y}$ as $\delta\mathbf{x}$ and $\delta\mathbf{y}$, respectively, we have the following first-order approximation:

$$\mathbf{T}^t \begin{bmatrix} \mathbf{x}_0 + \delta\mathbf{x} \\ \mathbf{y}_0 + \delta\mathbf{y} \end{bmatrix} \sim \mathbf{T}^t \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} + \mathbf{T}' \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{y}_{t-1} \end{bmatrix} \cdots \mathbf{T}' \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} \begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{y} \end{bmatrix}. \tag{19.14}$$

Then the following product J of Jacobian T' gives the Lyapunov spectrum with a standard numerical method such as Gram-Schmidt orthogonalization.

$$J = \mathbf{T}' \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{y}_{t-1} \end{bmatrix} \cdots \mathbf{T}' \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix}. \tag{19.15}$$

The analysis of probability propagation can be further extended to the information geometrical framework. The dynamics of probability propagation should be considered as a dynamical system in the Riemannian space with Fisher information matrix as the metric. The linear stability can be measured with respect to the infinitesimal distance on this manifold which is approximately twice the KL divergence between adjacent points. The space also includes singular points of the zero parameters of a discrete distribution which are expressed as positive/negative infinity in the natural parameters of exponential family.

19.7 Conclusion

We have developed a framework for a synergetic interaction between linguistic theory and mathematical modeling. Ecological approach for modeling language acquisition in multilingual environment with the use of contact process derived some examples of qualitative correspondence between typical model statistics and the working hypothesis from multilingual acquisition experiences. Further investigation with sufficiently large number of samples is needed to augment the plausibility of the model.

19.8 Appendix

The actual format of the questionnaire used in this study is shown in Fig. 19.12. In modeling section, the acquisition of a language was judged whether the respondent achieved at least one level B competence in one of the Listening/Speaking/Reading/Writing ability. The number of environmental factors l_{env} were calculated from the number of languages filled in the questions 2,4,5,6,7.

Questionnaire about Multilingual Environment and Multilingualism 言語環境･言語使用状況調査

The aim of this questionnaire is to find out the social factors that support the multilingual environment. We would like to investigate how humans can acquire naturally several languages, and from this data, we aim to develop a conceptual model that supports multilingualism. Any information filled in this form will only be used for the research.

名前 Name :

性別 Sex : F 女 / M 男

年齢 Age :

出生地(国･都市) Birthplace(Country and city) :

記入年月日 Date :

To answer the questions, please select the corresponding number (1 to 45) from the the list of the next page. If you wish to answer the language that is not on the list, please write the name of language instead of the number.

On the languages that you are able to use:

Q. 1 Please check all languages that you are able to use, regardless of the ability. Then please evaluate by yourself the ability of Listening/Speaking/Reading/Writing.
(A: Fluent, using for the work or at classroom / B: Sufficient for use in daily life / C: Able to communicate only in simple conversation, or reading, writing simple text / D: Poor, almost impossible to communicate)

Languages / 言語	Listening / 聴く				Speaking / 話す				Reading / 読む				Writing / 書く			
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
No.	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D

On the language environment in your life:

Q. 2 Please list up all languages spoken in the family you were raised.

No.	No.	No.	No.	No.	Comment :

Q. 3 Besides Q.2, if there are any languages spoken in your neighborhood (but not in your family) before entering a school, please list them up.

No.	No.	No.	No.	No.	Comment :

Q. 4 What were the languages used for education in your primary school ?

No.	No.	No.	No.	No.	Comment :

Q. 5 What were the languages used for education in your junior high and high school ?

No.	No.	No.	No.	No.	Comment :

Q. 6 What were the languages used for education in your university ?

No.	No.	No.	No.	No.	Comment :

Q. 7 If you are working in the society, what are the languages used in your working place ?

No.	No.	No.	No.	No.	Comment :

On the languages you are using now:

Q. 8 What are the languages used in TV, film, radio, etc., that you frequently see/listen ?

No.	No.	No.	No.	No.	Comment :

Q. 9 What are the languages used in newspaper, magazine, book, webpage, etc., that you frequently read ?

No.	No.	No.	No.	No.	Comment :

Q. 10	What languages do you use when you make a wish ?					
	No.	No.	No.	No.	No.	Comment :
Q. 11	What languages do you use in case you write a diary ?					
	No.	No.	No.	No.	No.	Comment :
Q. 12	What do you think is your mother tongue (first language) ?					
	No.	No.	No.	No.	No.	Comment :

List of languages 言語の選択肢					
1 英語 English	2 中国語（普通話・華語） Mandarin Chinese	3 福建（閩南）語 Chinese (Hokkien)	4 広東語 Chinese (Cantonese)	5 潮州語 Chinese (Teochew)	
6 客家語 Chinese (Hakka)	7 海南語 Chinese (Hainanese)	8 上海語 Chinese (Shabainese)	9 福州語 Chinese (Foochew)	10 その他の中国語（具体的に ） Chinese (Other Dialect: Please specify. ____)	
11 マレー語 Malay (Bahasa Melayu)	12 インドネシア語 Indonesian	13 ジャワ語 Javanese	14 スンダ語 Sundanese	15 バリ語 Balinese	
16 その他のインドネシア諸語（具体的に ） Other Indonesian Language (Please specify. ____)	17 タミル語 Tamil	18 テルグ語 Telugu	19 マラヤラム語 Malayalam	20 カンナダ語 Kannada	
21 ヒンディ語 Hindi	22 ウルドゥ語 Urdu	23 パンジャビ語 Punjabi	24 グジャラティ語 Gujarati	25 ベンガル語 Bengali	
26 その他のインド諸語（具体的に ） Other Indian Language (Please specify. ____)	27 タガログ語 Tagalog	28 イロカノ語 Ilocano	29 ビサヤ・セブアノ語 Visayas / Cebuano	30 その他のフィリピン諸語（具体的に ） Other Filipino (Please specify. ____)	
31 タイ語 Thai	32 ラオス語 Lao	33 クメール（カンボジア）語 Khmer	34 ベトナム語 Vietnamese	35 ミャンマー（ビルマ）語 Burmese	
36 日本語 Japanese	37 韓国・朝鮮語 Korean	38 アラビア語 Arabic	39 スペイン語 Spanish	40 ポルトガル語 Portuguese	
41 フランス語 French	42 ドイツ語 German	43 オランダ語 Dutch	44 ロシア語 Russian	45 その他（具体的に ） Others (Please specify. ____)	

Figure 19.12: **Template of the questionnaire about multilingual environment and multilingualism.**

Part VII

General Discussion

Chapter 20

General Discussion

Abstract

We review the results of this thesis and examine the contribution to complex systems sciences with respect to the proposed complementary strategy between dynamical system and information geometry in Part 1 (General Introduction). Comparison between the emergent properties in neural and social organization will be particularly issued as a concrete example. Further possibility along this strategy and novelty of the established models are also discussed. The results imply the effectiveness and possibility of the proposed strategy toward a universal characterization of complex systems in relationalistic viewpoint.

Keywords: Constructive methodology, interaction-analytical methodology, functor, meta-functor, organization of interactions

20.1 Introduction for General Discussion

In Parts 2 to 6, we investigated concrete subjects of complex systems using analytical tools derived from both dynamical system and information geometry. The analysis was performed in the global view of the complementary strategy between the constructive and interaction-analytical methodology. In Parts 2 to 6, we mainly discussed the contribution to each discipline. In this Part, we consider inter-subject comparison based on our modeling results to seek for universal properties of emergence beyond particularity of components as universal structures of meta-functor network, as introduced in Part 1.

20.2 Complementary Approach to Complex Systems: Dialectic Between Constructive and Interaction-Analytical Modeling

We review and integrate the complementary strategy between the constructive and interaction-analytical methodologies, based on the main results of modeling in Parts 2 to 6. Fig. 20.1 summarizes the results with respect to the proposed strategy.

The analysis of constructive modeling with the use of functor network is demonstrated in Part 2 and 3. The chaotic neural network (CNN) is a constructive model

of brain as a complex system. The firing patterns of the network, including the hierarchical restriction based on the structure of invariant subspaces are the function of neuron outputs, therefore correspond to functors. The analysis of system decompositionability after autonomous learning is the analysis of functor network with information geometry. Investigation in Part 2 revealed the property of chaotic itinerancy as a catalyst of learning, which implied certain role of dialectical information processing in CNN. This contribution serves to the model identification of autonomous learning in actual brain.

The dynamics of chaotic roving robot in Part 3 are also functors, since they are also based on the dynamics of CNN. Analysis on the dynamics or the robots were performed using plural statistical measures, including theoretical development of multi-scale evaluation on collective infotaxis using information geometry. The design of robot is not a modeling of actual system, but their collective search capacity is the emergent property based on the interaction between robots, which are themselves chaotic elements. The statistical analysis on the dynamics therefore serves to the identification of the collective infotaxis as complex systems.

Analysis of social network data in Part 4, on the other hand, directly measures the dynamics of complex systems without constructive model. The order-wise correlations, edge information, system decompositionability and the proposed complexity measures are the functors of data variables. The extraction of these functors are a part of the measurement process of complex systems and establish functor networks for further analysis including meta-functors.

We also proposed a way to link this interaction-analytical methodology of social network to a constructive methodology, with the use of the return map analysis from dynamical system theory. This is the case where a functor network model serves directly as the phase space of dynamical system in the formalization of constructive methodology. This realizes the constructive modeling based on the functor network model of actual data. Here, the interaction-analytical methodology serves as an analytical interpreter of complex systems, to better extract its interactions for a constructive modeling.

Such association of the constructive methodology from the interaction-analytical methodology is also proposed in Part 5. In EM-clustering, the primary variables of an embryo are approximated with Gaussian mixture model, whose Gaussian components form functors of data. The temporal development of embryo expressed in the parameter space of EM-clustering is nothing but a dynamical system defined on functor network model.

On the other hand, the primary variables for the detection of dynamical tissue is based on the vector field, which is a dynamical system representation. The definition of dynamical tissue is based on the statistical relation between these vectors, with the use of information geometrical measures. This is the inverse case where we calculate the functors from the phase space of constructive model.

The invariants of vowel system detected in Part 6 are an example of deterministic structure in functor network. Such invariance of functors is a candidate features to define the universality in complex systems, and is also accessible to further constructive modeling. We will further investigate the relation between the invariance and universality of functor network in the following sections.

The modeling and analysis of multilingual environment with the use of contact process with non-linear interaction in Part 6 is simply a constructive modeling using

stochastic dynamical system and its tentative analysis with information geometry. Though, the conceptual framework unifying the methodologies of linguistics and mathematical modeling aims to establish further dialectical refinement of the model. The analysis of functor network on the constructive model is expected to bring novel insight beyond the intuition of existing linguistics, which would lead to discover hidden reality of the phenomena. The accumulation effect of multilingual state via contact relation is an example of such novelty. Reconsidering the non-trivial fact derived from the analysis of functor network, another constructive model should be proposed including the extended reality of the target system.

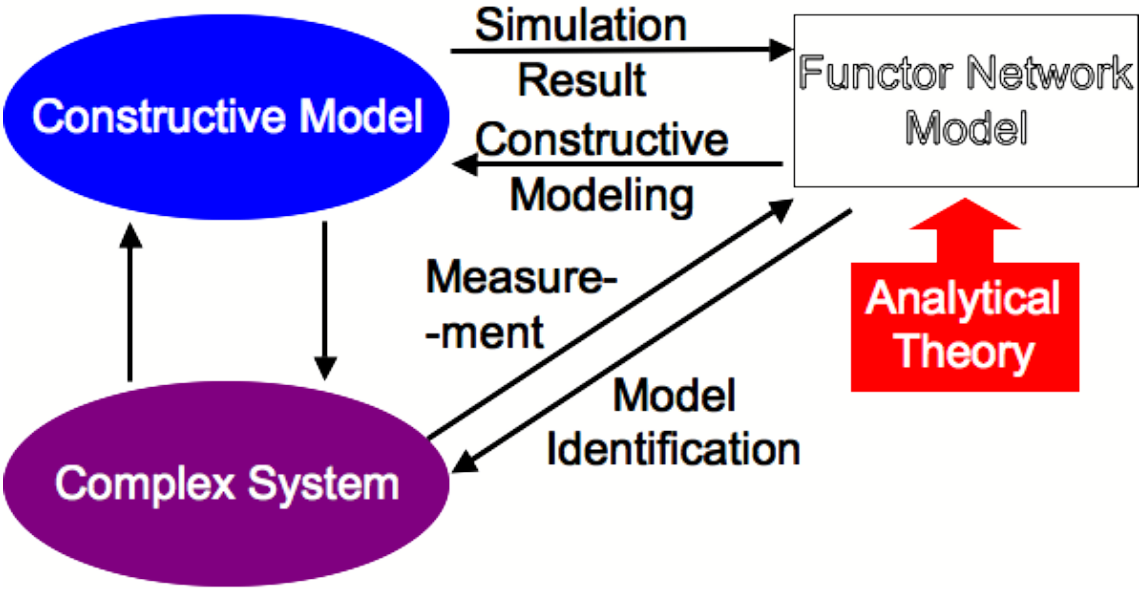


Figure 20.1: **Integrated constructive and interaction-analytical methodology.**

20.3 Universal Structure of Functor Networks in Different Levels of Emergence

In this section, we evaluate the contribution of the complementary strategy on the understanding of the emergence in terms of the organization of interactions. Our strategy had in global scope the characterization of different levels of emergence with compatible mathematical frameworks, so that to discover the universality that does not depend on the particularity of the components. The effectiveness of this strategy can be examined by comparing the results of Part 2 and Part 4, which connect the neuronal to neural network level and individual to social level of cognition, respectively. The micro-macro hierarchy of neuron-brain-society is an important axis for the understanding of intelligence as complex systems, whose emergent property is always based on the interaction between different levels of dynamics.

20.3.1 Multiple Complexities Analysis as Meta-Functor Network Analysis

The comparison between the results of Part 2 and 4 is schematized in Fig. 20.2.

In Part 2, we investigated the effect of autonomous learning on memory structure during the chaotic itinerancy state of CNN. Information geometric characterization of this process was performed by the measurement of inter- subsystem interactions, which are mathematically identical to the system decomposition in Part 4. Inter-subsystem interactions in CNN are functors since they are functions of the output variables. The statistical interactions between each subsystem becomes larger and more homogeneous as the autonomous learning proceed in non-trivial manner. Such dynamics of attractor recombination is a simplified candidate model of the autonomy of the brain producing novel information with the use of neuronal plasticity. In terms of statistical interactions, this process can be regarded as a complexification associated with the emergence of novel attractors.

In Part 4, direct analysis of political weblog data was performed with the use of multiple complexity measures. Here, the order-wise interactions, multi-information, system decompositionability, and edge information are functors. We evaluated the temporal dynamics of the french presidential in terms of the functor network analysis. The emergence of public opinion was characterized with multiple perspectives according to the property and relations of these functors.

On the other hand, other complexities based on the system decomposition, such as cuboid-bias complexities and modular complexities, are classified as meta-functors, since they are functions of system decompositionability being a functor. The analysis of meta-functor network brought a finer view on the emergence of collective semantic trend in the blogosphere, relating the dynamics of a single subsystem to the total coherence of the system.

If we consider the comparison between the emergence in CNN and social network treated in Part 2 and 4, simple comparison of functor networks is mostly limited to state the augmentation of statistical interactions. The statistical complexification is a long repeated insistence on the process of emergence, but we cannot find further utility of the functors besides this ordinary regime. The dynamics of some characteristic subsystems seems too much grounded to the particularity of the system, and is difficult to find the universal platform of comparison.

The meta-functor networks of the proposed complexities in Part 4, on the other hand, has a possibility to compare these different levels since they do not explicitly depend on particular variables nor interactions, but on the way interactions are organized. The meaning of the dynamics of functors are also clarified and put in universal platform by contextualizing them with meta-functors. The theoretical framework enables the comparison between functors, based on the concepts associated with the definition of meta-functors. The relation between meta-functors, or the analysis of meta-functor network, further create multiple criteria of comparison that are self-contained and time-independent since they refer to each other. We can close the description between them without referring to other global dynamics depending explicitly on the time axis. The increase and decrease of a complexity measure can be analyzed in comparison to other measures, as the way of organization of interactions comparing different hierarchy of the system. The actual events depending on the time axis can therefore be evaluated inversely from the functor and meta-functor network. As the level l of meta^l-functor augments, possible relations between meta^l-functors increase in combinatorics order. The investigation of the universality between different levels of emergence therefore requires the theoretical invention and selection of meta^l-functors in view of the comparison between different

levels.

For a concrete comparison of CNN and social network in meta-functor level, the regularized cuboid-bias complexity C_c^R is expected to play a key role in meta-functor network analysis. The augmentation of statistical interactions during autonomous learning of CNN is associated with the increase of homogeneity of system decompositions, that corresponds to the maximization of C_c^R. For the analysis of social network, the meta-functors corresponding to novel complexities C_c^R,C_m^R,C_m, and C_m all showed different dynamics. Since the overall degree of statistical interaction such as the multi-information I can not be an universal measure to compare different phenomena, a self-contained measure of relation between subsystems and total system is necessary for the comparison. For that purpose, the novel regularized complexities such as C_c^R and C_m^R can be regarded as universal measures. Other complexities such as C_c and C_m are dependent to the multi-information I which can differ according to the particularity of the system. The relation between the proposed meta-functors and other functors based on the C_c^R and C_m^R would be a concrete strategy in our theoretical framework to compare different models such as CNN and social network.

In Part 2, we only analyzed the system decompositionability of CNN during chaotic itinerancy before and after autonomous Hebb learning. To compare the emergent properties between the CNN and social network, we calculated the cuboid-bias and modular complexities as meta-functors of CNN. The results are shown in Figs. 20.3 and 20.4.

The both learning dynamics of Fig. 20.3 are the emergent process of novel attractor E⇌F with different learning coefficient ϵ. The top figure shows almost monotonic increase of all complexities until it reaches to the maximum value of C_c^R and C_m^R. The value of the multi-information I and other related complexities are also bounded and maintain the maximum value after 60 learning steps. The system decompositionability after the 0 and 350 learning steps are analyzed in Part 2.

The bottom figure, on the other hand, shows longer transient dynamics of complexities change until the novel attractor emerges around after 300 learning steps. Since the overall system is completely deterministic, the dynamics until 300 learning steps is necessary as the accumulative learning history, and in terms of internal measurement framework is a part of emergent process seeking for a pseudo-solution with intrinsic mechanism. The dynamics of the complexities show compensatory oscillation after 140 learning steps, between the cuboid-bias complexities C_c and C_c^R, as well as between the modular complexities C_m and C_m^R. Especially around the learning steps 160, 260, and 400, the relative values of I, C_c and C_m show local maxima, while those of C_c^R and C_m^R are local minima. This means that the augmentation of the multi-information I is owned by some specific subsystem.

Indeed, as we investigate the shortest decomposition path defining the modular complexities of this dynamics in Figs. 20.4, different specific subsystems show high internal coherence at the learning steps 160, 260, and 400. At the learning step 160 and 400, the interaction between the 1st and 4th nodes drastically increase and own much of the information increase. While at the learning step 260, the interaction between the 2nd and 3rd nodes is the emergent core of interaction.

Such dynamics is qualitatively similar to that of emergent process of collective opinion in social network during the 1st and 2nd round of French presidential 2007 detected in Part 4. If we compare the emergent process of CNN in Part 2 and social network in Part 4, in terms of the relative dynamics between meta-functors,

or the meta-functor network analysis, such qualitatively similar structure can serve to construct a typology of emergent process. As we have insisted, the abstraction up to the level of meta-functor realized to set a common language to compare between neural and social emergence in terms of the organization of interactions.

With the complementary strategy between the constructive and interaction-analytical methodologies, we have established an information geometrical framework that enables to compare between different levels of emergence with multiple analytical perspectives. The novel complexities as functors and meta-functors provided finer characterization of emergence by referring to each other on the complementarity basis between them.

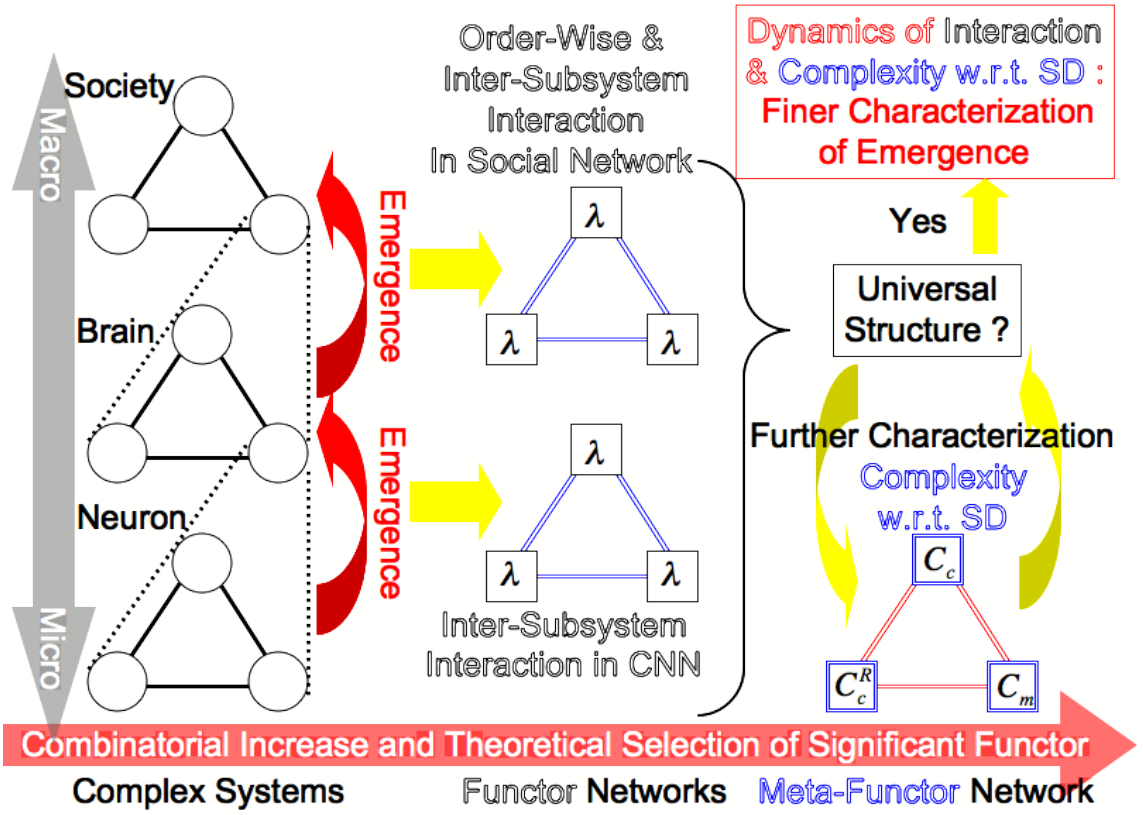


Figure 20.2: **Discovered universality in functor networks structures in different levels of emergence.**

20.3.2 Example of Common Structure as Meta-Functor in Theoretical Neuroscience and Systems Biology

The proposed framework to compare commonality between different levels of organization can be utilized to integrate the knowledges of different research areas with complex systems perspectives. The reinterpretation using the concepts of variables, functors, and meta-functors sets a common platform to discover the universal structure of emergent property in different systems. This will serve for the integration of scientific knowledge, overcoming the adverse effect of highly segmentalized professionalism, and cultivate holistic insight on what we have achieved as the understanding of living nature in relationalistic viewpoint.

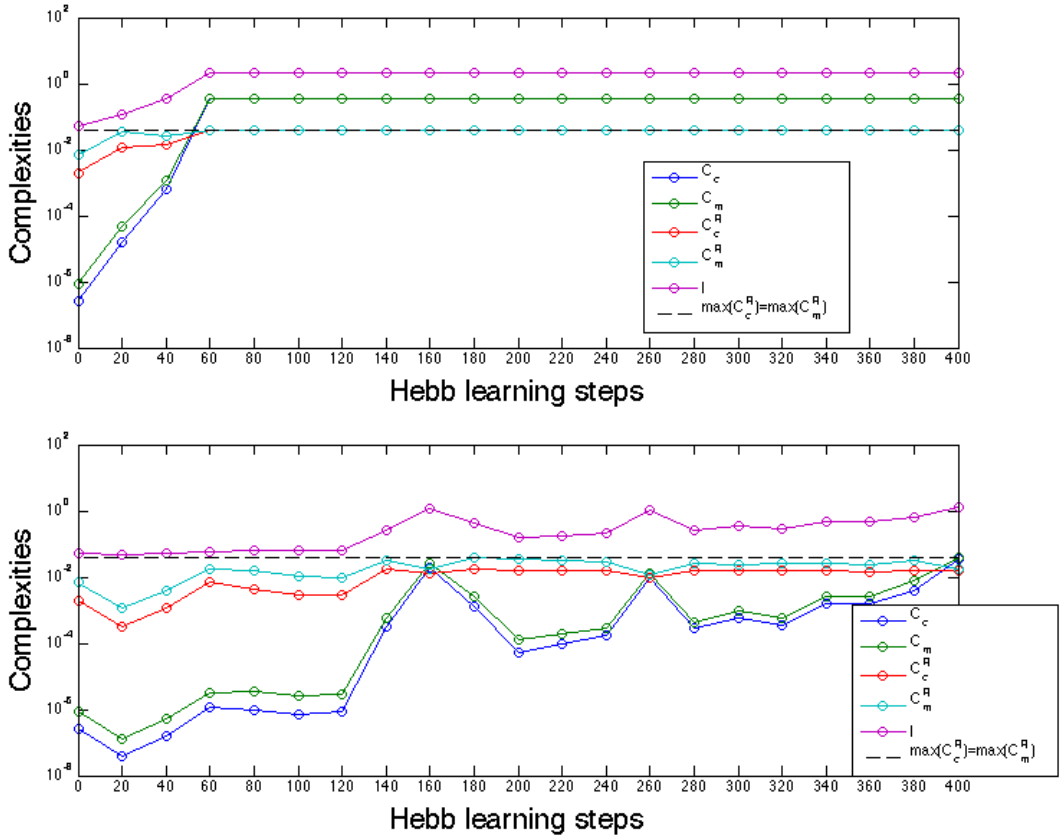


Figure 20.3: **Dynamics of cuboid-bias complexities during chaotic itinerancy after autonomous Hebb learning of CNN.** Top: $k_r = 0.4$, $\alpha = 5.0$, $T = 10000$, $\epsilon = 0.01$. Bottom: $k_r = 0.4$, $\alpha = 5.0$, $T = 10000$, $\epsilon = 0.001$.

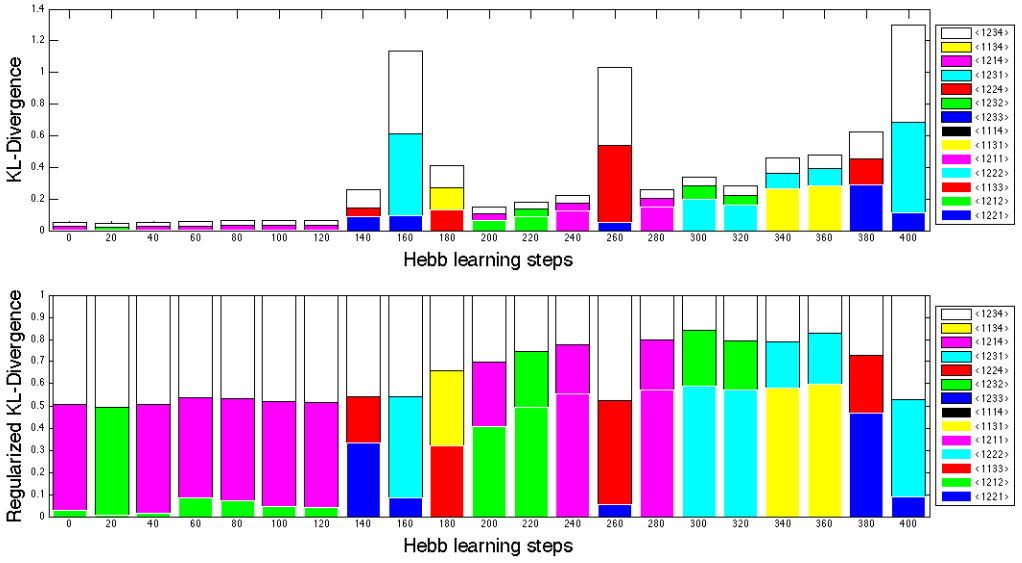


Figure 20.4: **Shortest decomposition paths and KL divergences of modular complexities during chaotic itinerancy after autonomous Hebb learning of CNN.** Top: Modular complexity C_m. Bottom: Regularized modular complexity C_m^R. Vertical axes represent KL divergence and regularized KL divergence, respectively. $k_r = 0.4$, $\alpha = 5.0$, $T = 10000$, $\epsilon = 0.001$ were used that corresponds to the Bottom of Fig. 20.3.

Here, we cite one of the interesting common structures of systems expressed as meta-functors in neural and cellular cognition. A multilayer perceptron is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output. It is more powerful than the standard linear perceptron in that it can distinguish data that is not linearly separable [165]. The chaotic neural network model introduced in Part 2 is also a part of it. If we interpret the multilayer perceptron with the use of meta-functor network model, the variables of neuron model correspond to the variables, interactions between neurons to functors, and the network structure to meta-functor. The three or more layer structure supporting the non-linear separation ability can be first defined in meta-functor level.

Such network structure was also found in various signal transduction pathways by systems biology, such as epidermal growth factor receptor (EGFR), G-protein coupled receptor (GPCR), and toll-like receptor (TLR) signaling networks [166] [167]. The so-called bow-tie structure of these networks are topologically equivalent to the multilayer perceptron, therefore expressed as identical feature in meta-functor level.

In both cases, the existence of classifier hyperspace supported by the mid-layer elements enables the network to realize complex nonlinear input-output relations. Although the nature of components are quite different between neurons and proteins, as well as its interactions, the meta-functor level expression codes the same structure and the same function. It may imply more than a simple mathematical analogy, but a possible universal structure supporting various levels of cognition, whether it is in cellular/molecular or neural ones. The use of meta-functor concept helps us to clarify the relation between the knowledges of different research areas in terms of common function. Such correspondence between meta-functor levels implies further extension of the concept of functor from function to category, which is recently investigated in quantum physics.

20.3.3 Invariance/Determinisity in the Distribution of KL Divergence

Similar comparison of emergent property also applies in other results that are not necessary integrated in micro-macro hierarchy. The invariance of inter-vowel KL divergence distribution in Part 6 is another example of meta-functor, since it describes the relation between the KL divergences that are functors of formant frequency variables. Such deterministic characteristic features were first observed at the level of meta-functor. This is different from other invariant quantities in statistical physics such as scale-free features in self-organized critical phenomena, since these belong to the functors. The invariants that first appear at the meta-functor level are the novel discovery for the characterization of emergence in vowel system.

The invariance of meta-functors based on KL-divergence distribution is also strategically proposed in Part 5, for the characterization of a normal process of embryogenesis in the EM-clustering space. Individual variations are the relation between individuals, that are formalized with KL-divergence as functors taking each individual process as variable. The quantification of individual variations and fuzzy definition of normal developmental process can be formalized as the functions of these functors, that are nothing but the quantitative classification of meta-functors. Categorization of normal/pathological processes is possible with respect to the invariance of meta-functors.

20.3.4 Toward Relationalistic Typology of Emergence

As a conclusion of this section, the comparison of functors and meta-functors dynamics between different systems set up a transversal perspective to universally characterize the emergence and derive its typology. Fig. 20.5 shows the three axes to relate emergent properties of different complex systems: The first axis is the spacial localization of interactions, such as the degree and heterogeneity of edge information on network presentation. The second axis is the variability of temporal evolution of these leading interactions, such as in case of CNN highly coherent subsystems changes temporally its spacial localization during autonomous learning (Figs. 20.3 and 20.4), while studied social dynamics showed less variability and always maintained the hardest core of interaction. The third axis is the statistical order of interactions first revealed and independently decomposed with the use of information geometry. How high-order statistics are involved in emergent process is an important key to characterize the nature of interaction in collective level. With the use of these three axes, one can situate all kinds of emergent property of different complex systems in terms of the organization of interactions, as long as they are accessible to the proposed dialectical methodology.

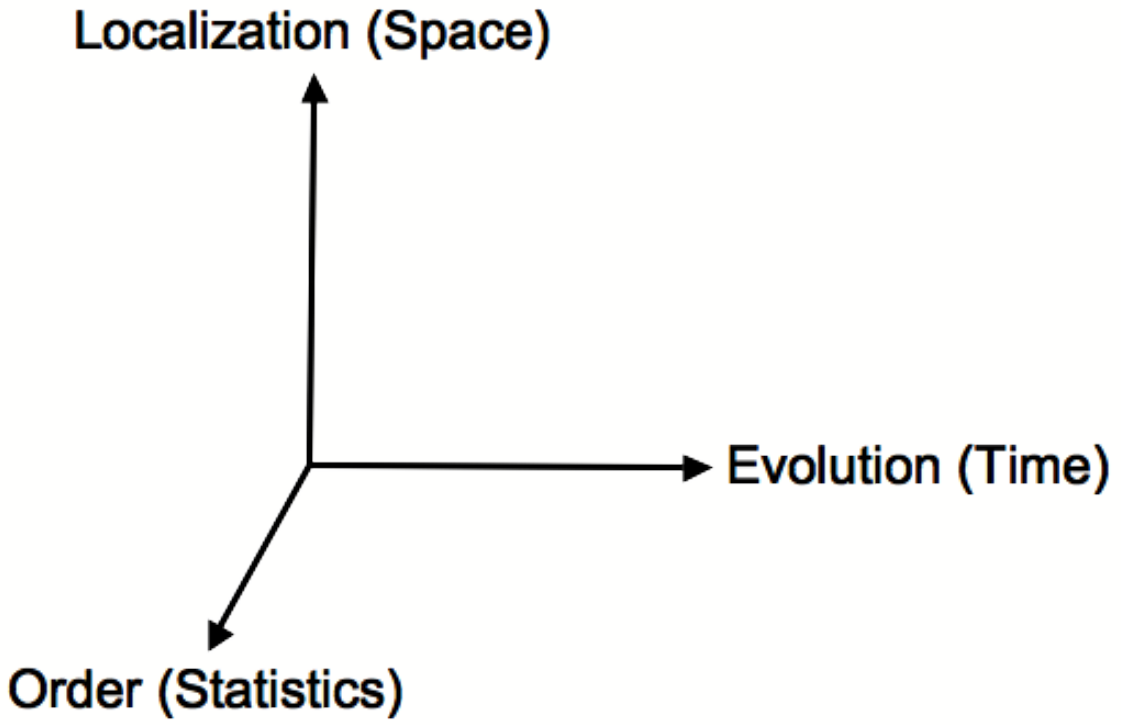


Figure 20.5: **3 axes for relationalistic typology of emergence.** Spacial localization, temporal evolution variability, and statistical order of interactions enable to characterize emergent properties of complex systems and to transversally compare in relationalistic level as the dynamical organization of interactions.

20.4 Integration of Micro Model

One of the critics against our approach would be the relative lack of importance on conventional micro models of each subject, which corresponds to the variation of variable networks in our framework. In this thesis, we have centralized our effort to generalize from particular to meta-level interactions to obtain universal property of emergence in complex systems. Once the universality is characterized, we need the opposite direction of modeling to contextualize the phenomenological models in each subject. The integration of micro models based on various variable networks in each discipline into our proposed universal framework is necessary to solve particular problems in each scale.

Such effort to integrate complex systems science perspective and other system theoretical approach is not yet necessary pursued even in rapidly developing field such as theoretical biology. Research field seems to be divided into whether to solve concrete problems with concrete models such as in systems biology [168], or to establish an universal class of model beyond particularity of individual system with strong intuition from physics which would encompass essential property of living organism such as in complex systems life science [10].

20.5 Dynamical Tissue and Internal Measurement

The modeling and detection of dynamical tissue in Part 5 was an attempt to incorporate internal measurement theoretical view into empirical phenomenological model. As discussed in Part 2, morphogenetic process during embryogenesis is a typical example of internal measurement, where the stability of the whole coordination is achieved by the components only having local information. Since the bottom-up mechanism from cellular to tissue level can be properly explained with measurement-oriented than state-oriented theory, the modeling also requires to incorporate internal observers as a representation of cells.

Indeed, the definition of dynamical tissue is based on the local relative coordinates between neighboring cells, and do not depend on global information. This is the very information that contribute to the mechanical relation between cells as internal observers. The local relative coordinates are the functors of cell movement vector fields, and their correlation are meta-functors. The dynamical tissue is therefore based on the thresholding of meta-functors. A comprehensive theoretical model based on such internal observers will only be simulated in high-performance computer, to reproduce *in silico* experimental platform for drug test.

To validate the internal measurement theoretical formalization, we always need the empirical measurement that belongs to the state-oriented paradigm to obtain collective dynamics data. Internal measurement theory gives an appropriate definition of the coordinates to describe effective interactions for the cells, while experimental measurement gives the database to decide the parameters of internal observers. We believe that the dialectic between the external measurement experiment and the internal measurement theoretical modeling is a fruitful way to decipher complex emergent property in morphogenesis. The limit of empirical modeling in the measurement-oriented theory can be compensated by the advancement of measurement technology. Comprehensive simulator integrating both theoretical precision of

local interaction and empirical measurement data will be a promised way toward the augmented virtuality of embryogenesis.

20.6 Relation with Open Systems Science

The concept of complex systems sciences is widely spread and taken with deep consideration in various research fields. Though, in most cases, researchers tend to simulate complex systems based on primary variables derived from reductionist approach, and investigate the effect of interaction between them. Such approach is strongly limited to the primal symbolization of the variables, which in complex systems of real world, neither guaranteed to be measurable in precision, nor even definable as appropriate components of the system since the effective unit of components may change in time line.

There has recently been a proposition to further expand the concept of complex systems science and overcome those difficulties facing real world problems, with the name of "*open systems science*" [169]. In contrast to closed systems that are accessible with external observation, expert view with deep but narrow knowledge, static modeling, and possible to provide strong/complete solutions, the open systems are characterized by the necessity of internal observation, holistic view, consideration of temporal development, and are only possible to "manage" the problems in the time line with best effort. In open systems, not only the understanding of the system by analysis and synthesis, but the "*management*" becomes one of the most important scientific issues. Complex systems science actually deal with both closed and open systems, but is basically anchored to the first definition of the system components and does not accept the modification of their definition in the time line, which is not sufficient to treat open systems in a long run. The concept of "*management*" is neither sufficiently considered in application to technology, including the long-term social effect of scientific invention.

In open systems science, the methodology of reductionism is strongly conditioned by the 4th criterion of René Descartes, the process of *verification*, by which we should thoroughly evaluate whether we did not omit anything to sufficiently reconstruct the whole property of the system [170]. After a tentative establishment of a model, if there is still something wrong or missing, we should reconsider the definition of the components itself both in quantitative and qualitative manner to attain the sufficient quality of modeling. The ever-changing tentative definitions of system components are called "*micro theories*" in open systems science, which can be usually extracted from a sufficiently huge and empirical database with the use of computer. With such incessant trial and error approach always profoundly doubting the plausibility of components definition, we may manage to adapt to the dynamical change of open systems, find path to cope with problems in real world, and realize sustainable society.

With this respect, the three axes for the typology of emergence should also consider the way of *observation* behind the axis of statistical order (Fig. 20.5). Since statistics practically depend on the past data to obtain sufficiently large sample number, in the real-time management of open systems from inside, it is not always useful nor available. One should consider the methodological limit of conventional statistics for transient phenomena, and intelligently incorporate internal observer's

view always conditioned by the observation.

The concepts and mathematical models developed in this thesis are also profoundly related to open systems science. The autonomous learning in Part 2 can be considered as the coupling between the variables and parameters in dynamical systems, which modify the model definition in the time line. This bring an important methodological framework in open systems science in which the parameters of a subsystem are basically variables of other ones.

The novel complexity measures, system decomposition, edge information, and order-wise interactions developed in Part 4 can quantify the inter-subsystem interactions, therefore express the temporally changing dominant source of information in the system. This is compatible to the derivation of "micro theories" by flexibly adjusting the effective definition of the system components, in this case the division of strongly correlated subsystems.

The unification of the internal measurement model and empirical external measurement of cell movement in Part 5 suggests a common ground that would unify novel measurement-oriented theory, *in toto* experimental measurement and massive calculation *in silico*.

Interfacing linguistics theory and mathematical modeling in Part 6 provides a novel way of expanding science and knowledge in an opposite but constructive way to the specialization of academic fields. To obtain deeper and wider understanding of multilingual environment is also an important subject in open systems science, toward the mutual understanding of human being beyond language barrier in the coming century.

20.7 Conclusion

We reviewed the results of each Part, from a global comparative viewpoint along the complementary strategy between the constructive and interaction-analytical methodologies proposed in Part 1. The translation of the concrete analysis in each Part into the definition of functors and meta-functors are performed. Meta-functor network analysis on different levels of emergence between the chaotic neural network and social network derived a typology of emergence in terms of the organization of interactions. Further open-ended discussions are listed to apply the proposed complimentary methodology in a wide range of complex systems sciences and beyond.

Chapter 21

Méthodologie dialectique entre système dynamique et géométrie informationnelle : Vers une typologie transversale des systèmes complexes

Résumé

Ce chapitre résume la contribution méthodologique et métathéorique de cette thèse. Nous étudions d' abord l' ensemble des interactions possibles du système qui deviennent les paramètres dominants par rapport à l' augmentation de sa taille. Ensuite, nous développons la méthodologie dialectique entre la théorie de système dynamique et géométrie informationnelle afin de analyser l' effet des interactions comme l' origine des phénomènes émergentes. Application de cette méthode aux plusieurs systèmes à différentes échelles implique la possibilité de construire la typologie transversale de l' émergence, basé sur la nature statistique et la dynamique de l' organisation des interactions qui ne dépendent pas explicitement de la propriété des composants.

21.1 La définition des systèmes complexes et les paramètres dominants

Selon l' enquête du communauté scientifique autour de la science des systèmes complexes au japon, les systèmes complexes sont définis comme le suivant [3]:

" les systèmes qui se constituent des composants variés, multiples, non-linéaires, parmi lesquels existent les interactions non-linéaires qui est la source de la propriété globale émergente, et réciproquement, les composants mêmes sont sous l' influence de l' état globale. " (Traduction par M. Funabashi)

La définition rigoureuse mathématique des systèmes complexes est difficile, cars sa propriété émergente est conditionnée par l' échelle d' observation. Ce qui paraît trivial par la vue de l' extérieur n' est pas explicite par l' observation intérieure, et vice

versa. Rien que le procès de relaxation vers l' équilibre globale en thérmodynamique n' est pas explicite si lÔ on se limite l' observation aux interactions locales entre les particules. Il est cependant évident que certain effet non-linéaire des interactions locales jouent un rôle essentiel des phénomènes émergents. S' il existe la caractérisation transversale de l' émergence, la formalisation nécessite une approche relationnelle ancrée au niveau des interactions, qui est la propriété distincte de la particularité des composants hors de l' état d' organisation. Notre approche commence donc par examiner le nombre des paramètres nécessaires pour définir un système de plusieurs composants. Considérons la formalisation stochastique d' un système (qui peut être déterministe) avec en général n variables, et étudions l' effet et nombre des interactions. Dans l' étude des systèmes complexes, la vitesse de propagation des interactions est en général essentiellement limitée par rapport à sa taille, qui limite le pouvoir de contrôle global ou centré de chaque composant. Cette prémisse est différente de celle des systèmes physiques où on peut ignorer la plupart de cas le temps de transmission des interactions tels que la lumière, la gravité, ou ignorer analytiquement la phase de transmission même, ce qui n' est pas le cas des interactions locale des systèmes complexes comme l' effet physico-chimique entre les cellules de l' embryogenèse. Si on limite la vitesse de propagation des interactions par un constant, l' effet que peut donner un composant au système entier se diminue à la puissance -1 par rapport à la taille n de système (Fig. 21.1 haut). Contrairement, le nombre des paramètres des interactions nécessaires pour définir complètement la distribution jointe de système augmente en combinatoire, et devient tout de suite les paramètres dominants du système (Fig. 21.2 bas). Les interactions se définissent non seulement entre chaque pair des composants, mais entre tous les sous-ensembles $2 \leq k \leq n$ comme les éléments indépendants de point de vue mathématique. Déjà à la taille $n = 7$, plus de 99% des paramètres représentent l' interaction entre les distributions marginales de chaque composant. Cette richesse des interactions possibles soutient la possibilité sous-jacente des phénomènes émergents. Les paramètres des interactions de haut niveau ne sont pas encore étudiés intensivement dans le domaine de système complexe, bien qu' il est important pour caractériser les phénomènes au-delà de simple corrélation binaire.

21.2 Méthodologie pour l' étude des interactions de la propriété émergente du système

Considérons maintenant la méthodologie pour aborder cette richesse des interactions. La formalisation conventionnelle de système complexe utilisant la représentation de réseau se fait de manière centrée aux composants comme variables, et ajoute les liens entre ces nœuds selon le degré des interactions (Fig. 21.2). Or, il est difficile de incorporer les interactions de haut niveau dans cette représentation. Nous considérons donc la représentation duale de ce réseau en mettant les interactions comme nœuds et les variables comme liens. Cette transformation nous permet de représenter tous les ordres d' interactions qui s' échappaient auparavant, et sur lequel on peut considérer et formaliser la relation entre les interactions. Nous allons appeler ces interactions symbolisées par les nœuds de réseau dual comme " foncteurs " d' après le nom d' objet de langage C++, et définissions la relation entre les foncteurs avec des fonctions mathématiques nommées " méta-foncteurs ". Cette échelle d'

abstraction se fera jusqu' il atteint suffisamment le haut niveau selon l' intérêt d' investigation sur l' organisation des interactions.

21.3 Stratégie concrète : Dialectique entre système dynamique et géométrie informationnelle

Nous intégrons maintenant la méthodologie sur l' analyse des interactions avec la modélisation concrète des systèmes complexes. Pour cela, nous introduisons une théorie qui traite l' ensemble des distributions de probabilité et considère la relation géométrique entre eux : La géométrie informationnelle [119]. La géométrie informationnelle représente la distribution jointe d' un système comme un point sur la variété statistique avec le métrique Riemannien basé sur la nature statistique de la théorie d' estimation. L' analyse géométrique sur cette variété nous permet d' évaluer la contribution de chaque ordre des interactions entre chaque sous-système: Les paramètres dominants des interactions peuvent être décomposé à chaque ordre de statistique et à chaque distinction de sous-système de manière orthogonale, comme les éléments mathématiquement indépendants (Partie IV). Cette formalisation avec la géométrie informationnelle est d' abord accessible à l' extraction des interactions substantives à partir des données réelles des systèmes complexes, qui déduira la modélisation système dynamique des phénomènes au niveau des interactions détectées (Fig. 21.3). Cette approche est surtout prometteuse quand le système est extrêmement complexe comme la société, et qu' on arrive pas à avoir une approche bottom-up à partir de modélisation directe de son mécanisme au niveau des composants. Nous pouvons aussi bénéficier de la géométrie informationnelle pour l' analyse des modèles système dynamique/stochastique simulés, au niveau de l' analyse des interactions émergées afin de valider le modèle par rapport au phénomène réel et interpréter sa propriété émergente. Plus particulièrement en cas de système dynamique, l' apparition du chaos impose une formalisation en terme de distribution globale comme la mesure invariante, qui est aussi compatible à la géométrie informationnelle. Nous appelons l' analyse des interactions par la géométrie informationnelle comme la méthodologie "interaction-analytique ", en contraste de la modélisation " constructive " des systèmes complexes par le système dynamique/stochastique. Cette méthodologie tente donc d' avoir l' interaction théorique entre les modèles conventionnels des systèmes complexes comme système dynamique/stochastique et l' analyse par géométrie informationnelle de manière dialectique, en sens que cela augmente le niveau de représentation vers l' étude de l' organisation des interactions.

21.4 Résumé métathéorique des résultats obtenus : Vers une typologie transversale de l' émergence

En appliquant la méthodologie dialectique entre système dynamique et géométrie informationnelle, nous avons examiné la dynamique, le contraint déterministe et l' invariance des plusieurs systèmes différents qui se varient entre système neuronale, système sociale, robotique collective, embryogenèse de poisson zèbre, système des voyelles, et système sociolinguistique (Partie II-VI). Parmi les résultats obtenus, nous démontrons quelques extraits d' exemples essentiels pour la typologie transversale d'

émergence. L' analyse des interactions par la géométrie informationnelle est effectué en principe sur le modèle de distribution log-normale, qui représente l' ensemble de distribution discrète et appartient à la famille exponentielle de la distribution de probabilité. Quel que soit le système modélisé, en symbolisant proprement les variables discrètes afin de représenter de manière compatible à la distribution log-normale comme réseau de Booléen, il est possible de comparer les différents systèmes avec l' échelle variée sur la formalisation mathématiquement identique. Les foncteurs représentent ici le degré des interactions selon l' ordre de statistique ou les variables concernées, à partir duquel nous avons défini les mesures de complexité comme méta-foncteurs par rapport à la difficulté de décomposition du système aux sous-systèmes. Ces nouvelles mesures représentent la plausibilité/difficulté d' appliquer le réductionnisme en décomposant le système aux parties indépendantes, et se forment les axes de l' espace de caractérisation non-linéaire sur la variété statistique comme celui de support vector machine (Partie IV).

21.5 Ex.1 Réseau neuronale : Emergence des nouveaux attracteurs dans la modélisation de l' apprentissage autonome avec réseau neuronal chaotique

L' enjeu de plasticité neuronale est un issue fondamentale pour savoir l' origine de la propriété émergente de système neuronale, l' intelligence. Il est récemment découvert que la dynamique chaotique de l' activité neuronale à l' échelle de réseau est une source d' introduire la nouveauté dans la structure de mémoire d' odorat [17]. Nous avons étudié l' effet de l' interaction entre la plasticité neuronale et la dynamique chaotique qui crée spontanément les nouveaux attracteurs de mémoire à partir de la structure déjà acquis. Un modèle de réseau neuronal chaotique (CNN) est établi de manière synthétique dans un cadre d' apprentissage autonome sans superviseur. L' analyse des mesures de complexité qui reflètent la dépendance statistique entre sous-systèmes montre que pendant l' apprentissage autonome qui consiste de la modification synaptique des neurones selon la propriété locale de la dynamique chaotique, la complexité totale de système augmente graduellement, qui est dirigé de manière transitoire par plusieurs sous-systèmes changeant temporellement sa distribution spatiale (Fig. 21.4). Il est donc important pour la caractérisation de l' émergence spontanée des nouveaux attracteurs d' introduire la dimension spatio-temporelle afin de évaluer cette variation des mesures de complexité comme méta-foncteurs, qui représentent le changement de la manière d' organisation des interactions entre les neurones.

21.6 Ex.2 Réseau social : Analyse des tendances des weblogs politiques sur le présidentiel 2007 en France

Passons ensuite à une autre échelle et prenons un exemple de système social. L' étude sur la cooccurrence des mots dans les 120 weblogs politiques le plus cité pendant le dernier présidentiel 2007 en France nous montre l' utilité de corrélation de haut niveau pour la caractérisation de la dynamique collective. La cooccurrence dans le même weblog des mots " J.M. Lepen " le candidat extrême droite, " N. Sarkozy

" le président élu, et " S. Royale " la rivale la plus compétente de N. Sarkozy, se trouve plus signifiante que les simples corrélations binaires entre chaque 2 personnes (Fig. 21.5). Ceci implique le fonctionnement catalytique de J.M. Lepen lors d' argument public dans le blogosphère, En contraste, un autre candidat F. Bayrou, malgré sa position centriste entre N. Sarkozy et S. Royale, n' était pas forcément pris en discussion compréhensive au niveau de la corrélation de 3e ordre. Si on ignore la distinction entre les différents ordres d' interactions ou ignore entièrement les interactions de hauts ordres, on risque de perdre ces informations qualitatives pour la caractérisation de dynamique collective. Cette dimension est peu considérée dans l' étude de réseau complexe en général, faute de manque de présentation approprié de réseau. La définition mathématique de graphe considère uniquement les liens entre deux nœuds et ignore entièrement le moyen de présenter les autres ordres des interactions, bien qu' ils sont nécessaire à définir la distribution jointe de réseau. La même analyse que CNN avec les mesures de complexité est effectuée sur la dynamique du présidentiel, qui mesure la difficulté de décomposition entre les noms des quatre candidats (Fig. 21.6). Nous observons l' émergence de sous-système Sarkozy-Royal qui reste comme noyaux dure de corrélation pendant toute la durée de présidentiel, et qui dirige l' augmentation imminente de la complexité totale lors de 2e tour (pendant 112e-127 e jour). Ce sous-système le plus corrélé ne change pas la distribution spatiale en contraste du cas de CNN.

21.7 Discussion : La caractérisation transversale de la propriété émergente des différents systèmes

Considérons enfin la comparaison transversale des propriétés émergentes qui n' appartient pas à la même échelle. Les exemples introduits se composent des composants extrêmement différents et se situent à l' échelle complètement distincte (Fig. 21.7): le système neuronal se compose des neurones, qui donne naissance à la propriété émergente au niveau de la dynamique de réseau neuronale. Le système social, qui se compose des humains et les autres infrastructures, fait émerger la dynamique sociale aux plusieurs niveaux collectifs y compris le contenu sémantique. Chaque échelle d' émergence dépend à la base la propriété des composants, qui selon contexte et la modalité de couplage fait fonctionner les interactions. Les foncteurs mêmes dépendent donc encore explicitement à la propriété des composants, et il serait difficile d' avoir la comparaison significative sur la nature des interactions. Pour aborder à la manière d' organisation des interactions, il faut introduire le niveau de méta-foncteur qui traite la relation des interactions comme la structure de réseau entre les foncteurs. Ce niveau dépend donc essentiellement sur les interactions, et il est possible de comparer entre plusieurs systèmes mettant en critère la manière d' organiser les interactions. Si on trouve certain trait commun de la dynamique à ce niveau, la caractéristique est le candidat de la propriété transversale, idéalement universelle de l' émergence. En outre, la variance de dynamique à ce niveau implique la nécessité de l' axe pour la classification des phénomènes qui constitue l' espace de typologie de l' émergence avec la compatibilité transversale.

21.8 Conclusion : 3 axes de l' analyse des interactions vers une typologie transversale des systèmes complexes

Nous considérons que la caractérisation transversale et " universelle " de l' émergence ignore trop la particularité de l' échelle et/ou des composants, qui ne s' échappe pas de critique sur le concept universaliste de système complexe. Au contraire, notre but est de clarifier la variance et la communauté des phénomènes émergents de manière transversale, et s' il existe, la particularité de l' échelle et des composants. Nous renonçons donc la théorie universelle de l' émergence, mais plutôt insistons à obtenir la typologie pour conditionner et caractériser les phénomènes émergents au niveau des interactions du système. Selon les résultats de l' application de la méthodologie dialectique entre système dynamique et géométrie informationnelle aux plusieurs systèmes différents, il est possible de proposer les trois axes relationnels suivants pour établir la typologie transversale (Fig. 21.8): Premièrement, la localisation spatiale des interactions est importante. La distribution ou l' hétérogénéité des fortes interactions dans le système implique le centre de contrôle ou le fournisseur d' information. Seconde, la variabilité temporelle de l' hétérogénéité spatiale représente le côté dynamique et transitoire du phénomène. i.e. Par rapport à la dynamique transitoire entre plusieurs sous-systèmes dirigeant l' augmentation de complexité dans CNN, la dynamique de réseau sociale restait relativement stable au niveau de l' évolution temporelle de noyaux dure de sous-système. Troisièmement, l' ordre statistique des interactions caractérise la manière d' organisation collective du phénomène, qui est le côté peu développé dans l' étude de système complexe. La cohérence de sous-système soutenue par des simples corrélations de seconde ordre et celle de plus haut niveau sont essentiellement différentes comme la dynamique des interactions. L' étude au niveau de méta-foncteur est encore moins étudiée. Ces éléments sont essentiels pour trouver les interprétations de l' émergence de manière transversale au niveau de l' organisation des interactions qui ne dépend pas explicitement de la propriété des composants.

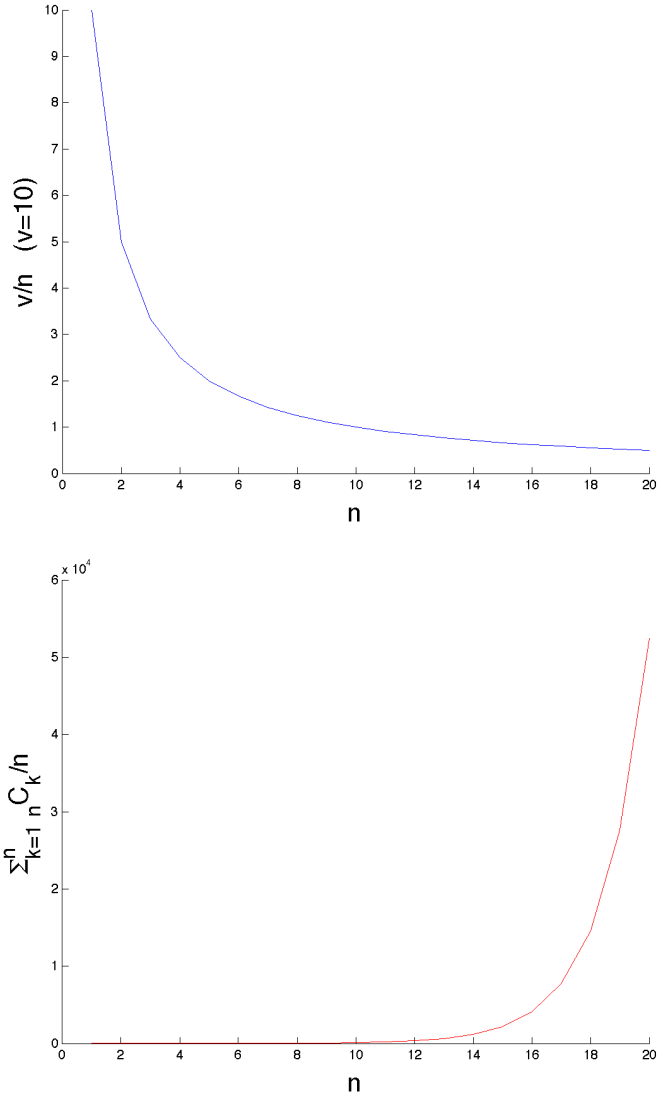


Figure 21.1: **Degré de dominance des composants et les interactions dans système complexe.** Haut : La taille ou le nombre des variables de système n vs la vitesse constante de propagation des interactions renormalisé par la taille de système v/n (ligne blue). Bas : La taille ou le nombre des variables de système n vs le nombre des combinaisons possibles des variables qui définissent les interactions indépendantes (ligne rouge).

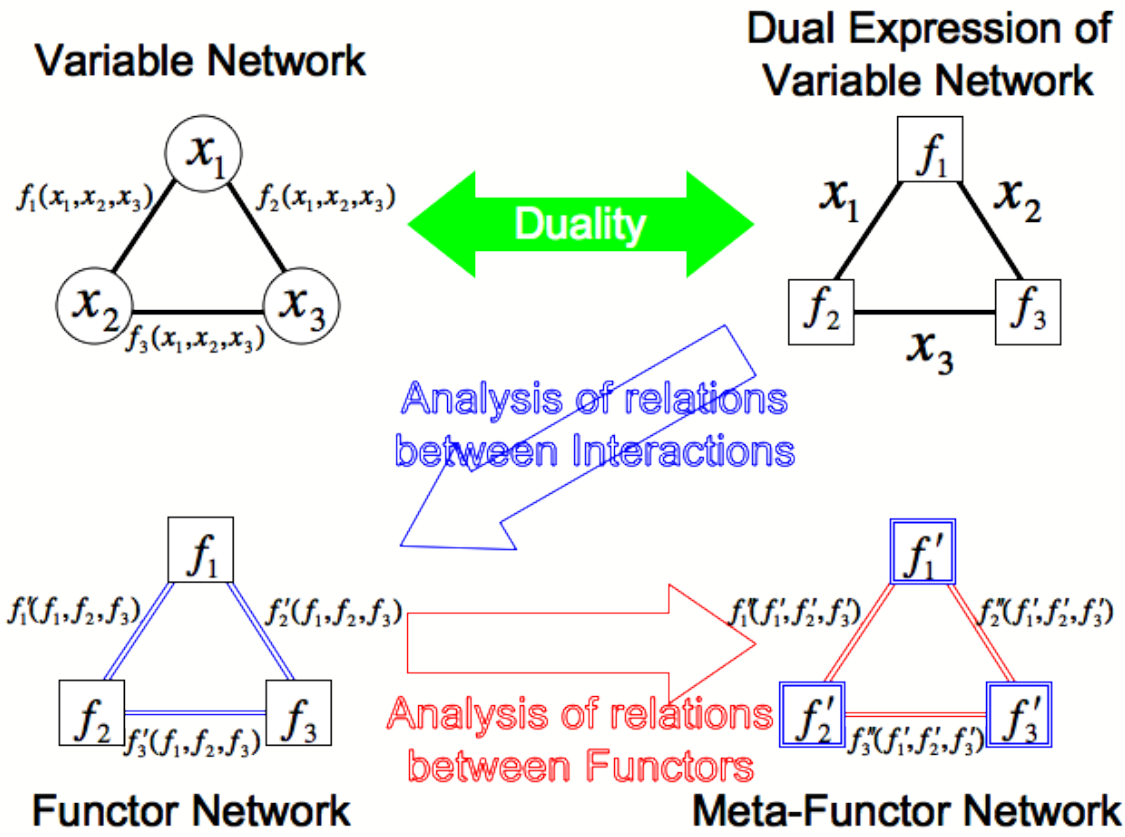


Figure 21.2: **Dualité entre variables, interactions, et reconstruction du réseau des (méta-)foncteur.** Les symboles x représentent les variables et les f sont des fonctions qui définissent les interactions. Nous appelons f comme " foncteur " et en considérant la relation entre les foncteurs, on obtient un niveau augmenté d' abstraction avec les " méta-foncteurs " f' qui sont les fonctions des foncteurs.

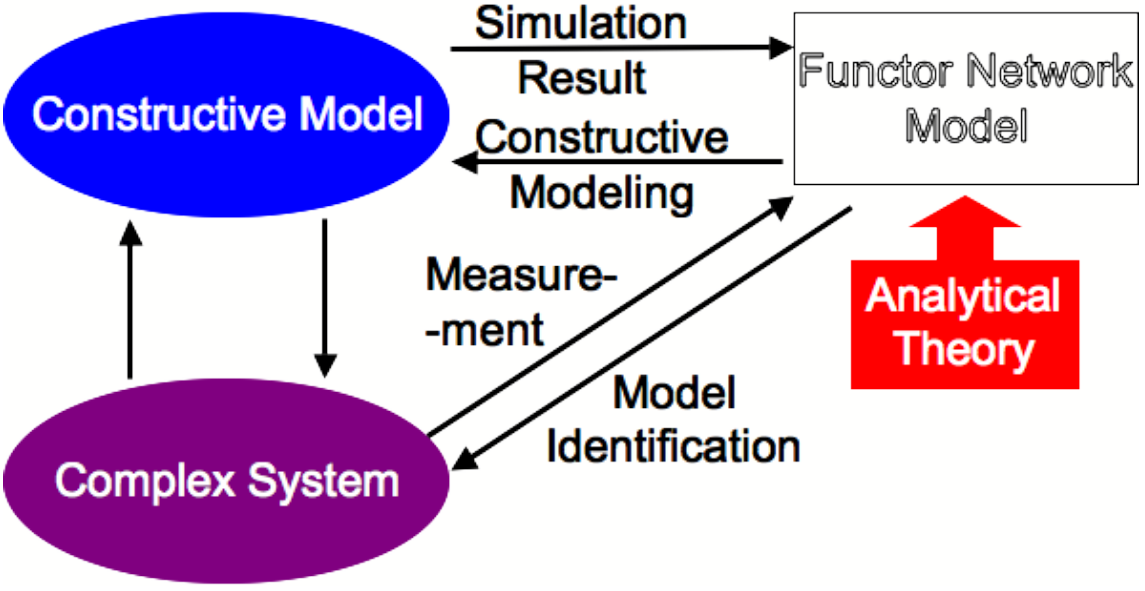


Figure 21.3: **Intégration des méthodologies constructive et interaction-analytique.**

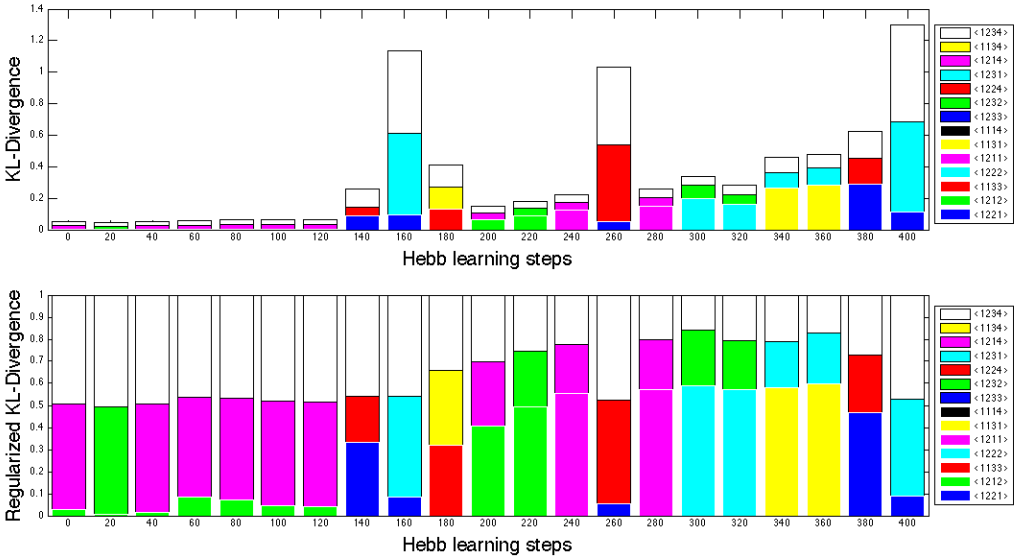


Figure 21.4: **Dynamique des mesures de complexité pendant l' apprentissage autonome de CNN.** Axe horizontal : Pas de temps d' apprentissage autonome. Axe vertical : Mesures de complexité représentant la difficulté de décomposition statistique de chaque sous-système.

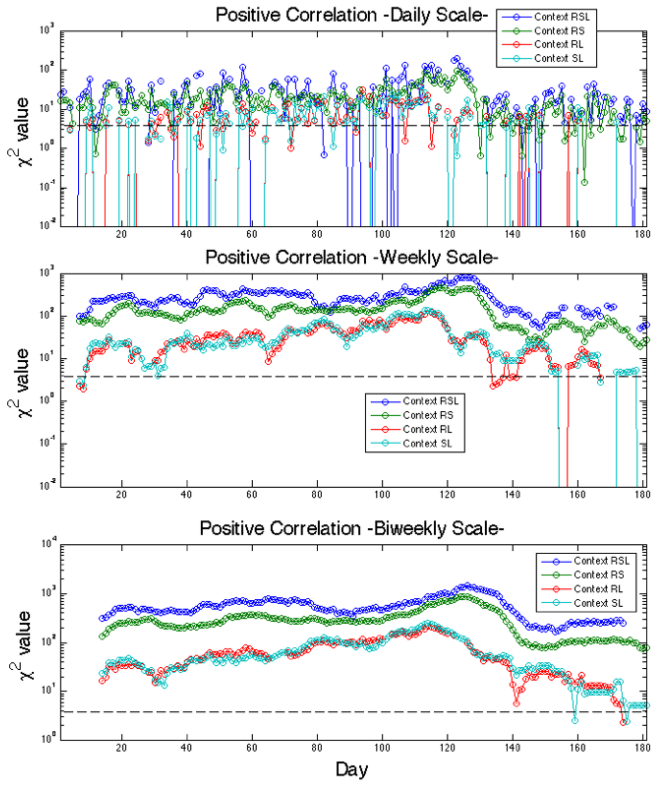


Figure 21.5: **Dynamique de corrélation entre S.Royal, N.Sarkozy, et J,M,Lepen dans les weblogs politiques pendant le présidentiel 2007.** Axe horizontal : jour compté du début du présidentiel. Axe vertical : Degré de signifiance des interactions de 2e et 3e ordre de statistique.

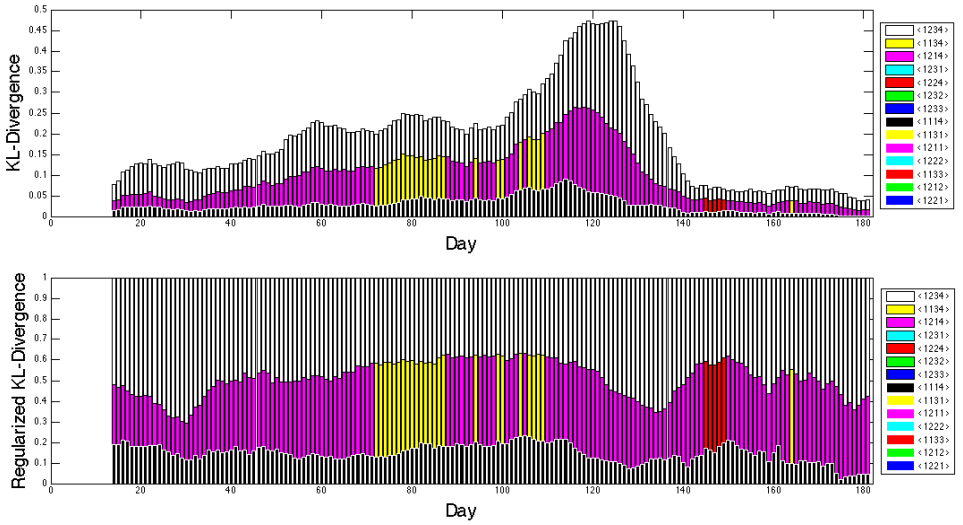


Figure 21.6: **Dynamique des mesures de complexité entre 4 candidats pendant le présidentiel 2007.** Axe horizontal : Jour compté du début du présidentiel. Axe vertical : Mesures de complexité représentant la difficulté de décomposition statistique de chaque sous-système (sous-ensemble des 4 candidats).

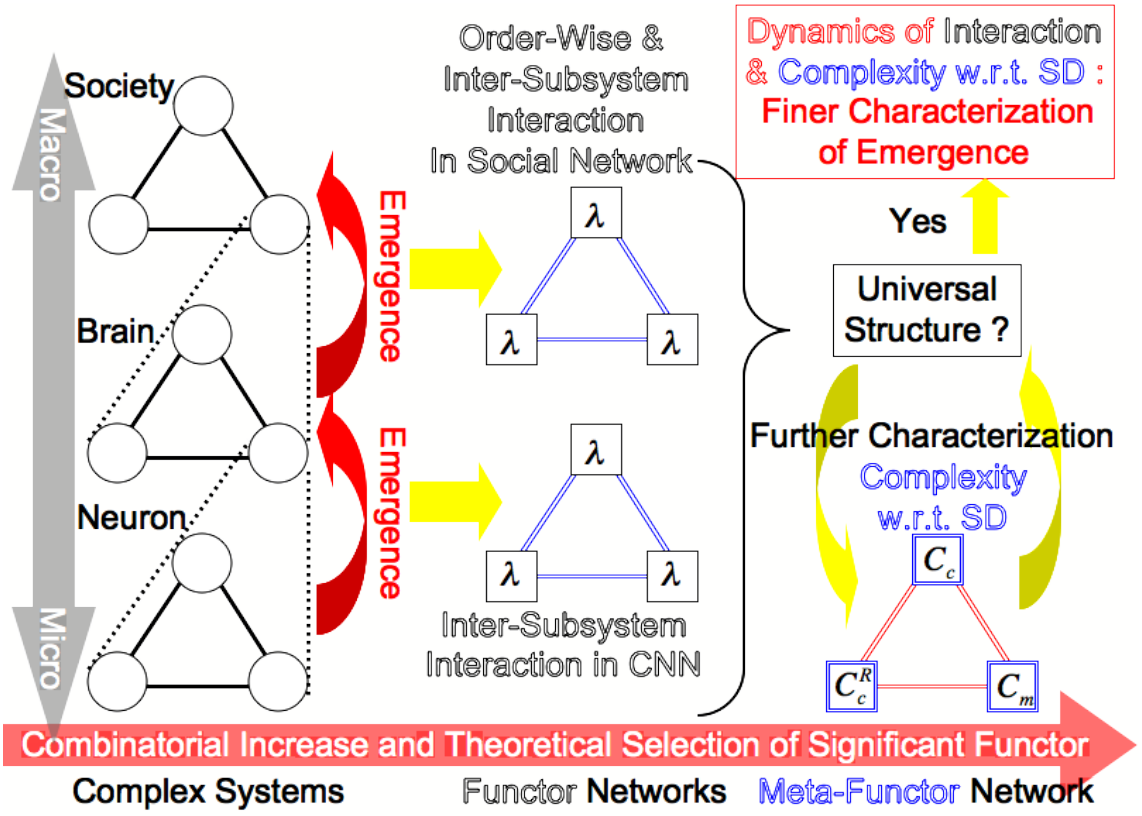


Figure 21.7: **Caractérisation transversale de réseau de foncteur par méta-foncteur à différentes échelles de l' émergence.**

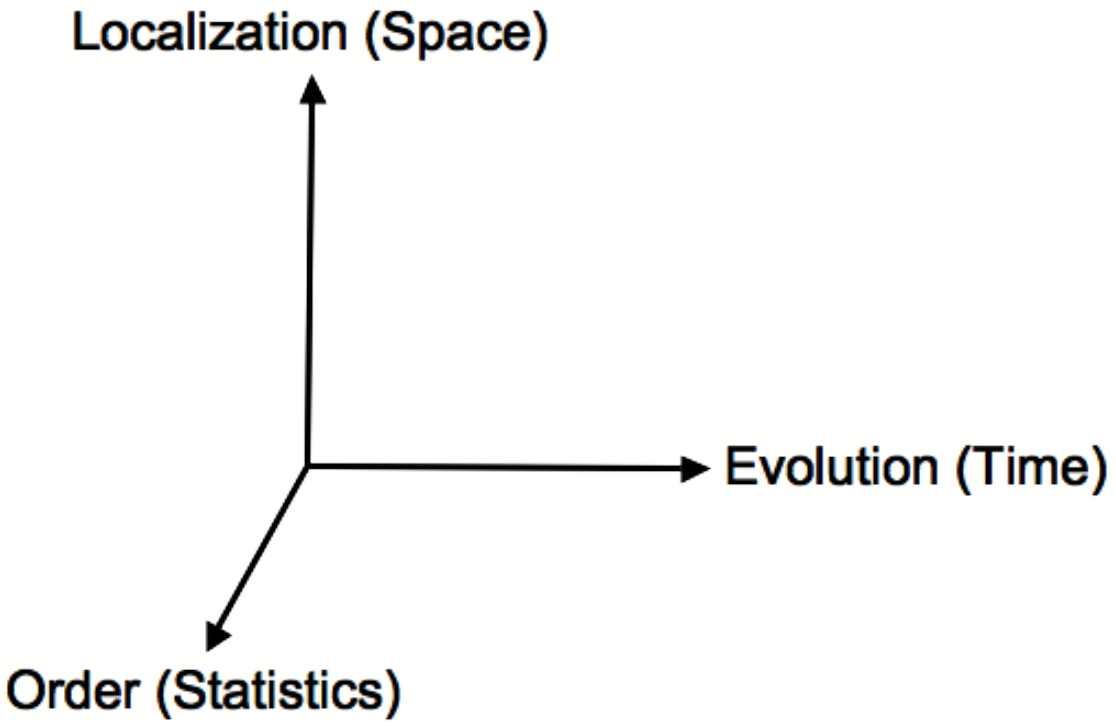


Figure 21.8: **3 axes pour la typologie transversale de l' émergence.**

References

[1] M. Komuro. *Dynamical System from the Basis (in Japanese).* Saiensu-sya, Tokyo, Japan, 2005.

[2] S. Akaho. Information Geometry and Machine Learning (in japanese). http://staff.aist.go.jp/s.akaho/infogeo/infogeo.html.

[3] K. Aihara et al. *Introduction to Chaology (Kaosu Gaku Nyumon, in japanese).* Teaching Material of the Open University of Japan, Japan, 2001.

[4] T. Iba and Y. Fukuhara. *Introduction to Complex Systems (Fukuzatukei Nyuumon, in japanese).* NTT Publication, Japan, 1998.

[5] J. Vuillemin. *La Philosophie de l'Algèbre.* Presses Universitaires de France, 1993.

[6] W. Weaver. Science and complexity. *American Scientist*, 36:536–544, 1948.

[7] P.E.Rapp, T.I.Schmah, and A.I.Mees. Models of knowing and the investigation of dynamical systems. *Physica D*, 132:133–149, 1999.

[8] P. Checkland. *Systems Thinking, Systems Practice.* John Wiley & Sons Ltd, 1981.

[9] K. Aihara et al. *Chaos (Kaosu, in japanese).* Information & Computing=49. Saiensu-sya, Japan, 1990.

[10] K. Kaneko and I. Tsuda. *Complex Systems: Chaos and Beyond.* Springer, 2000.

[11] K. Matsuno. *What is Internal Measurement (Naibukansoku toha Nanika, in japanese).* Seidosya, Japan, 1981.

[12] P.-Y. Gunji, K. Matsuno, and O. E. Rossler. *What is Internal Measurement (Internal Measurement (Naibu Kansoku, in japanese).* Seidosya, Japan, 1997.

[13] P. Goyal. From information geometry to quantum theory. *New Journal of Physics*, 12(023012), 2010.

[14] D. C. Dennett and D. R. Hofstadter. *The Mind's I: Fantasies and Reflections on Self and Soul.* Penguin Books Ltd, 1982.

[15] H. Nagashima and Y. Baba. *Introduction of Chaos (Kaosu Nyuumon in japanese).* Baifuukan, Tokyo, 1992.

[16] S. Amari and H. Nagaoka. *Method of Information Geometry*, volume 191 of *Translations of Mathematical Monograph.* Oxford University Press, 2000.

[17] C. A. Skarda and W.J. Freeman. How brain make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10:161–195, 1987.

[18] I. Tsuda. Dynamic link of memory - chaotic memory map in nonequilibrium neural network. *Neural Networks*, 5:313–326, 1992.

[19] H. Fujii, K. Aihara, and I. Tsuda. Functional relevance of 'excitatory' gaba actions in cortical interneurons: A dynamical systems approach. *Journal of Integrative Neuroscience*, 3(2):183–205, 2004.

[20] H. Fujii and I. Tsuda. Neocortical gap junction-coupled interneuron systems may induce chaotic behavior itinerant among quasi-attractors exhibiting transient synchrony. *Neurocomputing*, 58-60:151–157, 2004.

[21] I. Tsuda and H. Fujii. A complex systems approach to an interpretation of dynamic brain activity i: Chaotic itinerancy can provide a mathematical basis for information processing in cortical transitory and nonstationary dynamics. In *Computational Neuroscience: Cortical Dynamics*, volume 3146 of *Lecture Notes in Computer Science*, pages 109–128. Springer Berlin / Heidelberg, 2004.

[22] K. Kaneko, I. Tsuda, S. Tadokoro, T. Yasuoka, and Y. Yamaguti. Chaotic itinerancy as a mechanism of irregular changes between synchronization and desynchronization in a neural network. *J. of Integrative Neuroscience*, 3:159–182, 2004.

[23] K. Kaneko and I. Tsuda. Chaotic itinerancy. *Chaos*, 13(3):937–946, 2003.

[24] K. Aihara, T. Takabe, and M. Toyoda. Chaotic neural networks. *Physics Letters A*, 144(6-7):333–340, 1990.

[25] M. Adachi and K. Aihara. Associative dynamics in a chaotic neural network. *Neural Networks*, 10(1):83–98, 1997.

[26] S. Kang, K. Kitano, and T. Fukai. Structure of spontaneous UP and DOWN transitions self-organizing in a cortical network model. *PLoS Comp. Biology*, 4(3):e1000022, 2008.

[27] J. Ito and K. Kaneko. Spontaneous structure formation in a network of chaotic units with variable connection strengths. *Physical Review Letters*, 88(2):028701, 2002.

[28] M. Komuro and K. Aihara. Hierarchical structure among invariant subspaces of chaotic neural networks. *Japan Journal of Industrial and Applied Mathematics*, 18(2):335–357, 2001.

[29] E. Ott and J.C. Sommere. Blowout bifurcations: The occurrence of riddled basins and on-off intermittency. *Physics Letters A*, 188:39–47, 1994.

[30] E. Ott, J.C. Sommere, T.M. Antosen Jr, and S. Venkataramani. Blowout bifurcations: Symmetry breaking of spatially symmetric chaotic states. *Lecture Notes in Physics*, 450:182–195, 1995.

[31] R. S. MacKay. Nonlinearity in complexity science. *Nonlinearity*, 21:T273–T281, 2008.

[32] R. S. MacKay. *Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems*, chapter Indecomposable Coupled Map Lattices with Non-unique Phase, pages 65–94. Lecture Notes in Physics. Springer, Berlin / Heidelberg, 2005.

[33] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, 1949.

[34] H. Markram, J. Lubke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 10:213–215, 1997.

[35] I. Tsuda. Chaotic itinerancy as a dynamical basis of hermeneutics in brain and mind. *World Futures*, 32:167–184, 1991.

[36] I. Tsuda. Chaotic itinerancy is a key to mental diversity. *Behavioral and Brain Sciences*, 27(4):585–602, 2004.

[37] G. Tononi, O. Sporns, and M. Edelman. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U.S.A.*, 91:5033, 1994.

[38] N. Ay, E. Olbrich, N. Bertschinger, and J. Jost. A unifying framework for complexity measures of finite systems. Report 06-08-028, Santa Fe Institute, 2006.

[39] S. Morra, C. Gobbo, Z. Marini, and R. Sheese. *Cognitive Development: Neo-Piagetian Perspectives*, pages 367–368. Lawrence Erlbaum, 2007.

[40] K. W. Fischer and T. R. Bidell. *Theoretical Models of Human Development, Handbook of Child Psychology*, volume 1, chapter Dynamic Development of Action, Thought, and Emotion, pages 313–399. Wiley, New York, 6th edition, 2006.

[41] B. P. Kennedy K. W. Fischer. *Change and Development: Issues of Theory, Method, and Application*, chapter Tools for Analyzing the Many Shapes of Development: The Case of Self-in-Relationships in Korea, pages 117–152. The Jean Piaget Symposium Series. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[42] M.D. Lewis. *Emotion, Development, and Self-Organization*, chapter Emotional Self-Organization at Three Time Scales, pages 37–69. Cambridge University Press, 2000.

[43] M.D. Lewis. *Appraisal Processes in Emotion: Theory, Methods, Research*, chapter Personal Pathways in the Development of Appraisal: A Complex Systems/Stage Theory Perspective, pages 205–220. Series in Affective Science. Oxford University Press, New York, US, 2001.

[44] J. Pascual-Leone. A mathematical model for the transition rule in Piaget's development stages. *Acta Psychologica*, 32:302–345, 1970.

[45] A. Raftopoulos A. Demetriou. The shape and direction of development: Teleologically but erratically lifted up or timely harmonious ? *Journal of Cognition and Development*, 5(1):89–95, February 2004.

[46] G.W.F. Hegel. *Phänomenologie des Geistes*. Suhrkamp, 1970.

[47] K.F.Riegel. Dialectical operations: the final period of cognitive development. *Human Development*, 16:346–370, 1973.

[48] W. C. Hoffman. *Mind in Time: The Dynamics of Thought, Reality, and Consciousness*, chapter The Formal Structure of Dialectical Psychology. Advances in Systems Theory, Complexity, and the Human Sciences. Hampton Press, 2003.

[49] W. C. Hoffman. Symmetry in psychology. *Visual Mathematics*, 5(4), 2003.

[50] H. M. Sheffer. A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society*, 14:481–488, 1913.

[51] P. Norvig S. J. Russell. *Artificial Intelligence: Modern Approach*. Prentice Hall, 3rd edition, December 11, 2009.

[52] P.-Y. Gunji. *Internal Measurement (Naibukansoku, in japanese)*, chapter Tekiounou to Naibukansoku. Seido-sha, Japan, 1997.

[53] M. Kawato. *Nou no Keisan Riron (in japanese)*. Sangyo Tosyo, Tokyo, 1996.

[54] K. Doya. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, 10:732–739, 2000.

[55] K. Samejima, Y. Ueda, K. Doya, and M. Kimura. Representation of action-specific reward values in the striatum. *Science*, 310:1337–1340, 2005.

[56] Doya K. Metalearning and neuromodulation. *Neural Networks*, 15:495–506, 2002.

[57] S. Tanaka, K. Doya, G. Okada, K. Ueda, Y. Okamoto, and S. Yamawaki. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, 7:887–893, 2004.

[58] K. Doya and E. Uchibe. The cyber rodent project: Exploration of adaptive mechanisms for self-preservation and self-reproduction. *Adaptive Behavior*, 13:149–160, 2005.

[59] Y. Kuniyoshi, Y. Ohmura, Y. Terada, A. Nagakubo, S. Eitoku, and T. Yamamoto. Embodied basis of invariant features in execution and perception of whole body dynamic actions – knacks and focuses of roll-and-rise motion. *Robotics and Autonomous Systems*, 48(4):189–201, 2004.

[60] Y. Kuniyoshi and S. Suzuki. Dynamic emergence and adaptation of behavior through embodiment as coupled chaotic field. In *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, pages 2042–2049, 2004.

[61] Y. Kuniyoshi, Y. Ohmura, Y. Terada, A. Nagakubo, S. Eitoku, and T. Yamamoto. From humanoid embodiment to theory of mind. *In Iida F, Pfeifer R, Steels L, Kuniyoshi Y (Eds.), Embodied Artificial Intelligence, Lecture Notes in Artificial Intelligence*, 3139, 2004.

[62] J.-J. Aucouturier, Y. Ogai, and T. Ikegami. Making a robot dance to music using chaotic itinerancy in a network of fitzhugh-nagumo neurons. In *Proceedings of the 14th International Conference on Neural Information Processing (ICONIP 14 th)*, Kitakyushu, Japan, NOV 2007.

[63] Y. Li, T. Tanaka, and S. Nara. Control of roving robot using chaotic dynamics in a quasi-layered recurrent neural network for sensing and driving. In *SICE Annual Conference 2007*, pages 976–979, Takamatsu, Japan, 2007.

[64] J. Kuroiwa, N. Masutani, S. Nara, and K. Aihara. Sensitive responses of a chaotic wandering state to memory fragment input in a chaotic neural network model. *International Journal of Bifurcation and Chaos*, 14(4):1413–1421, 2004.

[65] M. Vergassola, E. Villermaux, and B. I. Shraiman. 'infotaxis' as a strategy for searching without gradients. *Nature*, 445:406–409, 2007.

[66] M. Funabashi and K.Aihara. Modeling birdsong learning with a chaotic elman network. *Artificial Life and Robotics*, 11(2):162–166, 2007.

[67] J. Namikawa. Chaotic itinerancy and power-law residence time distribution in stochastic dynamical system. *PRE*, 72(026204), 2005.

[68] O. Bénichou, M. Coppey, M. Moreau, P-H. Suet, and R. Voituriez. Chaotic itinerancy and power-law residence time distribution in stochastic dynamical system. *PRL*, 94(198101), 2005.

[69] W.J. O'Brien, H.I. Browman, and B.I. Evans. Search strategies of foraging animals. *American Scientist*, 78:152–160, 1990.

[70] L.D. Kraumer and R.L. McLaughlin. The behavioral ecology of intermittent locomotion. *American Zoologist*, 41(2):137–153, 2001.

[71] R. Jeanson, S. Blanco, R. Fournier, J.-L. Deneubourg, V. Fourcassié, and G. Theraulaz. A model of animal movements in a bounded space. *J. Theor. Biol.*, 225:443–451, 2003.

[72] O. Bénichou, M. Coppey, M. Moreau, P-H. Suet, and R. Voituriez. Optimal search strategies for hidden targets. *Phys. Rev. Lett.*, 94(19):198101, 2005.

[73] S. Wasserman and K. Faust. *Social network analysis*. Cambridge University Press, Cambridge, 1994.

[74] E. Garfield. Historiographic mapping of knowledge domains literature. *Journal of Information Science*, 30(2):119–145, 2004.

[75] S. Boccalettia, V. Latorab, Y. Morenod, M. Chavezf, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, February 2006.

[76] S. H. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001.

[77] H. Nakahara and S. Amari. Information-geometric measure for neural spikes. *Neural Computation*, 14(10):2269–2316, 2002.

[78] J. Holzfuss and U. Parlitz. *Lecture Notes in Mathematics*, volume 1486, chapter Lyapunov Exponents from Time Series, pages 263–270. Springer, Berlin / Heidelberg, 1991.

[79] M. Funabashi, J.-P. Cointet, and D. Chavalarias. *Complex Network*, volume 207 of *Studies in Computational Intelligence*, chapter Order-Wise Correlation Dynamics in Text Data, pages 161–172. Springer, 2009.

[80] S. Amari. Information geometry on hierarchy of probability distributions. *IEEE Transaction on Information Theory*, 47(5):1701–1711, July 2001.

[81] The political weblog data were collected by Jean-Phillipe Cointet.

[82] J.-P. Cointet and C. Roth. Socio-semantic dynamics in a blog network. In *IEEE proceedings of 2009 International Conference on Computational Science and Engineering*, 2009.

[83] R. Badii and A. Politi. *Complexity: Hierarchical Structures and Scaling in Physics*. Cambridge University Press, 2008.

[84] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and its Applications*. Texts in Computer Science. Springer, 2nd edition, 1997.

[85] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Whiley, July 2006.

[86] C.H. Bennett. On the nature and origin of complexity in discrete, homogeneous, locally-interacting systems. *Foundations of Physics*, 16:585–592, 1986.

[87] P. Grassberger. Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics*, 25(9):907–938, 1986.

[88] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: The entropy convergence hierarchy. *Chaos*, 15:25–54, 2003. and refs therein.

[89] J. P. Crutchfield. Inferring statistical complexity. *Phys. Rev. Lett.*, 63(2):105–108, 1989.

[90] D. P. Feldman and J. P. Crutchfield. Measures of statistical complexity: Why? *Physics Letters A*, 238(4-5):244–252, 1998.

[91] E. Olbrich, N. Bertschinger, N. Ay, and J. Jost. How should complexity scale with system size? *The European Physical Journal B*, 63(3):407–415, 2008.

[92] D. P. Feldman and J. P. Crutchfield. Measures of statistical complexity: Why? *Physics Letters A*, 238(4-5):244–252, 1998.

[93] R. Lopez-Ruiz, H. Mancini, and X. Calbet. A statistical measure of complexity. *Physics Letters A*, 209:321–326, 1995.

[94] G. Fraser, editor. *The New Physics for the Twenty-First Century*. Cambridge University Press, 2006.

[95] Project url: http://www.embryomics.eu/.

[96] Project url: http://www.bioemergences.eu/.

[97] K. Moriyoshi, L.J. Richards, C. Akazawa, D.D. O'Leary, and S. Nakanishi. Labeling neural cells using adenoviral gene transfer of membrane-targeted GFP. *Neuron*, 16:255–260, 1996.

[98] N.C. Shaner, P.A. Steinbach, and R.Y. Tsien. A guide to choosing fluorescent proteins. *Nat Methods*, 2:905–909, 2005.

[99] Z. Krivá, K. Mikula, N. Peyriéras, B. Rizzi, and A. Sarti. 3d early embryogenesis image filtering by nonlinear partial differential equations. *Medical Image Analysis*, 2008.

[100] A. Sarti, R. Malladi, and J.A. Sethian. Subjective surfaces: A method for completing missing boundaries. In *Proceedings National Academy Science of USA*, volume 97, pages 6258–06263, 2000.

[101] A. Sarti and G. Citti. Subjective surfaces and Riemannian mean curvature flow graphs. *Acta Math. Univ. Comenian.*, 70:85–103, 2000.

[102] S. Corsaro, K. Mikula, A. Sarti, and F. Sgallari. Semi-implicit co-volume method in 3D image segmentation. *SIAM Journal on Scientific Computing*, 28:2248–2265, 2006.

[103] C. Zanella et al. Segmentation of cells from 3-D confocal images of live zebrafish embryo. In *Conf Proc IEEE Eng Med Biol Soc*, pages 6028–6031, 2007.

[104] K. Mikula, N. Peyrieras, M. Remesíková, and A. Sarti. 3D embryogenesis image segmentation by the generalized subjective surface method using the finite volume technique. In *Proceedings of FVCA5 – 5th International Symposium on Finite Volumes for Complex Applications*, Paris, 2008. Hermes Publ.

[105] K. Mikula, A. Sarti, and F. Sgallari. *Handbook of Medical Image Analysis: Segmentation and Registration Models*, chapter Co-volume Level Set Method in Subjective Surface Based Medical Image Segmentation, pages 583–626. Springer, New York, 2005.

[106] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

[107] C. Melani et al. Cells tracking in a live zebrafish embryo. In *Conf Proc IEEE Eng Med Biol Soc*, pages 1631–1634, 2007.

[108] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, September 1998.

[109] W. Heisenberg. *Der Teil und das Ganze*. Piper, Germany, 1969.

[110] J. Xuan, T. Adali, and X. Liu. Information geometry of maximum partial likelihood estimation for channel equalization. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 6(7-10):3533–3536, May 1996.

[111] V. Garcia and F. Nielsen. Simplification and hierarchical representations of mixtures of exponential families. *Signal Processing*, In Press, 2010.

[112] S. Amari. Information geometry of the EM and em algorithms for neural networks. *Neural Networks*, 8(9):1379–1408, 1995.

[113] F. Nielsen and S. Boltz. The Burbea-Rao and Bhattacharyya centroids. *CoRR*, 2010.

[114] J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory*, 37:145–151, 1991.

[115] J. Burbea and C. R. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28(3):489–495, 1982.

[116] J. Burbea and C. R. Rao. On the convexity of higher order Jensen differences based on entropy functions. *IEEE Transactions on Information Theory*, 28(6):961–962, 1982.

[117] S. Akaho. Dimension reduction method for mixture parameters based on information geometry. IEIC Technical Report 657, Institute of Electronics, Information and Communication Engineers, 2006. Vol.105; No.657 (NC2005 104-124); Pages 57-62.

[118] Personal communication with Miguel A. Luengo-Oroz. The definition of morphogenetic entropy was established by him.

[119] S. Amari and H. Nagaoka. *Method of Information Geometry*, volume 191 of *Translations of Mathematical Monograph*, page 61. Oxford University Press, 2000.

[120] S. Furui. *Digital Speech Processing, Synthesis, and Recognition.* Marcel Dekker Inc, 1989.

[121] T. Chiba and M. Kajiyama. *The Vowel: Its Nature and Structure.* Phonetic Society of Japan, 1958.

[122] D.F. Klatt and L.C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. of Acous. Soc. Am.*, 87:820–857, 1990.

[123] W. Labov. La transimission des changements linguistiques. *Langages*, 26(108):16–33, 1992.

[124] N. Minematsu, T. Nishimura, K. Nishinari, and K.Sakuraba. Theorem of the invariant structure and its derivation of speech gestalt. In *SRIV-2006*, pages 47–52, 2006.

[125] S.Takano and S.Nakamura. Multilingual enviroment and natural aquisition of language. *Statistical Physics*, CP519:785–796, 2000.

[126] B. de Boer. *The Origins of Vowel Systems.* Oxford University Press, 2001.

[127] M.K. Sonmez. Information geometry of topology preserving adaptation. In *Proceedings of IEEE International Conference onAcoustics, Speech, and Signal Processing, ICASSP*, volume 6, pages 3743 – 3746, 2000.

[128] A. Gunawardana. *The Information Geometry of EM Variants for Speech and Image Processing*. PhD thesis, The Johns Hopkins University, 2001.

[129] http://www.fon.hum.uva.nl/praat/.

[130] A.P. Simpson. Gender-specific articulatory-acoustic relations in vowel sequences. *Journal of Phonetics*, 30(3):417–435, 2002.

[131] A.P. Simpson. Dynamic consequences of differences in male and female vocal tract dimensions. *Journal of the Acoustical Society of America*, 109(5):2153–2164, 2001.

[132] Y. Samuelsson. Gender effects on phonetic variation and speaking styles A literature study. Gslt speech technology term paper, Department of Linguistics, Stockholm University, autumn 2006.

[133] R. E. Turner and R. D. Patterson. An analysis of the size information in classical formant data: Peterson and barney (1952) revisited. *J. Acoust. Soc. Jpn*, 33:585–589, 2003.

[134] M.Saitou and T.Tsumura. Relation between timbre and depth of frequency fluctuation -comparison between listening condition through headphone and free field-. *J. Acoust. Soc. Jpn.*, MA 90-5:3–10, 1990.

[135] E. L. Crow and K. Shimizu, editors. *Lognormal Distributions: Theory and Applications*. Markel Dekker Inc, New York, 1988.

[136] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2003.

[137] J.J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin Harcourt, 1979.

[138] L. Menard, J.-L. Schwarz, and J. Aubin. Invariance and variablity in the production of height feature in french vowels. *Speech Communication*, 50(1):14–28, 2008.

[139] U. G. Goldstein. *An Articulatory Model for the Vocal Tracts of Growing Children*. PhD thesis, Massachusetts Institute of Technology, 1980.

[140] S. Lee, A. Potamianos, and S. Narayanan. Acoustics of childrenÕs speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.*, 105(3), March 1999.

[141] I. Shimizu. *Singing Synthesis Controlled Fundamental Frequency and Spectrum Envelope*. Master thesis, Japan Advanced Institute of Science and Technology, 2003.

[142] D. L. Philipona and J. K. O'Regan. Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, 23:331–339, 2006.

[143] F.E. Clements. *Research Methods in Ecology*. University Publishing Company, Lincoln, Nebraska, 1905.

[144] C. Goodwin. Participation, stance, and affect in the organization of activities. *Discourse and Society*, 18(1):53–73, 2007.

[145] R. Jakobson. *Essais de Linguistique Général, II Rapports Internes er Externes du Language*. Les Editions de Minuits, Paris, 1973.

[146] C.P. Snow. *The Two Cultures And The Scientific Revolution*. Cambridge University Press, 1959.

[147] A. Wray, editor. *The Transition to Language*. Oxford University Press, USA, 2002.

[148] N. Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, IT-2(3):113–124, 1956.

[149] J. Petitot. Sur la signification linguistique de la théorie des catastrophes. *Mathématiques et Sciences Humaines*, tome 79:37–74, 1982.

[150] A. J. Lotka. *Elements of Mathematical Biology (formerly published under the title Elements of Physical Biology)*. Dover, New York, 1958.

[151] V. Volterra. Variazioni e fluttuazioni del numero d'individui in specie d' animali conviventi. *Mem. Acad. Lincei.*, 2:31–113, 1926.

[152] J. A. Vano, J. C. Wildenberg, M. B. Anderson, J. K. Noel, and J. C. Sprott. Chaos in low-dimensional lotka-volterra models of competition. *Nonlinearity*, 19:2391–2404, 2006.

[153] T.M. Liggett. *Interacting Particle Systems*. Springer-Verlag, New York, 1985.

[154] L. Malson and J. Itard. *Les Enfants Sauvages, Mythe et Réalité Suivi de Mémoire et Rapport sur Victor de l'Aveyron*. Bibliothèques, 2002.

[155] S.S. Mufwene. *Language Evolution: Contact, Competition and Change*. Continuum International Publishing Group Ltd, 2008.

[156] L.J. Calvet. *La Guerre des Langues : Et les Politiques Linguistiques*. Payot, 1987.

[157] T.E. Harris. Contact interactions on a lattice. *Ann. Probab.*, 2:969–988, 1974.

[158] Y. Sakakibara. *Let's Sing Languages (Kotoba wo Utae! Kodomotachi, in japanese)*. Chikuma-Shobo, Tokyo, 1985.

[159] http://www.lexhippo.gr.jp/hippo.php.

[160] Hippo Family Club. *Anyone can Speak 7 Languages*. Lex Institute / Hippo Family Club, 1997. ISBN 0-9643504-5-9.

[161] D. K. Mauzy and R. S. Milne. *Singapore Politics Under the People's Action Party*. Routledge, 2002.

[162] R. Vasil. *A Citizen's Guide to Government and Politics in Singapore*. Talisman Pub., Singapore, 2004.

[163] R. Vasil. *Governing Singapore : Interviews with the new leaders*. Times Books, Singapore, 1988.

[164] L. J. Abu-Raddad, P. Patnaik, and J. G. Kublin. Dual infection with hiv and malaria fuels the spread of both diseases in sub-saharan africa. *Science*, 314(5805):1603–1606, 2006.

[165] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 1989.

[166] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1(2005.0010.), 2005.

[167] K. Oda and H. Kitano. A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology*, 2(2006.0015.), 2006.

[168] H. Kitano, editor. *Foundations of Systems Biology*. The MIT Press, 2001.

[169] M. Tokoro, editor. *Open Systems Science: From Understanding Principles to Solving Problems*. IOS Press, 2010.

[170] R. Descartes. *Discours de la Méthode*. Editions Flammarion, 1991.

Publications

1. M.Funabashi and K. Aihara, "**Modeling Birdsong Learning with a Chaotic Elman Network**" *Artificial Life and Robotics* (2007), Vol 11, Number 2, pp. 162-166

2. M. Funabashi, A. Tamaki and Y. Sakakibara, "**Methodology for Dialectic between Linguistic Theory and Mathematical Modeling and Its Application to Modeling Multilingual Environment with Contact Process**" in *Rapport d'activité du Groupe de Travail Ecologie, Linguistique et Anthropologie (GdT-ELA). Paris: ISCPIF/RNSC/EHESS (Appel à Idées07)* (2008)

3. M. Funabashi and S. Aoki, "**Balancing Autonomy and Environmental Response with Hierarchical Chaotic Dynamics**" in *Proceedings of IEEE International Conference on Robotics and Biomimetics*, (2008)

4. M. Funabashi, D. Chavalarias, and J.-P. Cointet, "**Order-wise Correlation Dynamics in Text Data**" in *Complex Network*, Studies in Computational Intelligence 207, Springer (2009) pp.161-172

5. M. Funabashi, "**Network Decomposition: An Information Theoretical Approach**" in *Proceedings of European Conference of Complex Systems* (2009)

Glossary

AIC

Akaike information criterion. An index to evaluate the plausibility of statistical model. Generally the maximum likelihood augments as the number of model parameters increases, though it risks to have overfitting problem. Information criterion is invented to balance between the complexity of the model and the goodness of fitting.

Attractor Ruin

The state of attractor region that went through the interior crisis and possesses both attracting and repelling directions of orbit. The plural presence of attractor ruin is considered as the definition of chaotic itinerancy. I. Tsuda insists that attractor ruins are generated by the crisis of Milnor attractors. According to K. Kaneko, in globally coupled map with maximum symmetry, the augmentation of the variable dimensions around 7 ± 2 degrees of freedom automatically realizes the situation that the number of attractors overpasses the possible partition number of basin, therefore each basin is crashed in confined state space and becomes attractor ruin.

Chaos

The mathematical definition of chaos still varies according to researchers, though the common properties can be summarized as follows:

- Complex dynamics similar to random sequence appear from rather simple equation.
- Maximum Lyapunov exponent is positive, therefore only short-time prediction is possible.
- High sensitivity to initial condition, or a tiny change of initial condition results in global difference of long-term dynamics.
- Contains topologically homologous structure with baker's transformation.

Chaotic Itinerancy

Considered as an universal class of dynamics in high-dimensional dynamical systems that the orbits follow intermittent chaotic transition among relatively low-dimensional synchronous states. It was first discovered by K. Ikeda in optical turbulence in 1989, then independently by K. Kaneko and I. Tsuda after 1990. Its importance is stressed in many biological and other systems.

Dynamical System

Systems that change its state according to the time development with defined rules, or mathematical models to describe such systems. Generally, we define the elements that affect the state of system as variables, and describe the interactions between variables with differential or difference equation.

Information Geometry

A field of applied mathematics combining statistics and geometry created principally by S. Amari, initially to simulate the information propagation in neural network. Information transmission in neural network often shows stronger correlation between relatively distanced neurons than closer ones. To analyze such dynamics, Riemannian geometry is more appropriate to treat an ensemble of statistical models. Information geometry is an attempt to interpret information theory and statistics from differential geometry, and to provide a more integrative theoretical framework.

Functor

In this thesis defined as some function of plural variables with one-dimensional return. Named as functor to symbolize the function as an object and consider the correspondence between them, taking after the "function object" of C++ language. The expansion to relate with the concept of functor in category theory may be possible but not treated in this thesis.

Fisher Information Matrix

Represents the amount of information that observable variables possess with respect to the unknown parameters of likelihood function. Mathematically defined as the variance of the score, which is the partial derivatives of the logarithm of the likelihood function with respect to the parameters. The first idea to utilize Fisher information as Riemannian metric in tangent space of statistical manifold was proposed by C.R. Rao. Geometrical structure introduced with the use of Fisher information remains invariant under non-linear transformation of parameters and variables, as long as the parameters uniquely specify each distribution and the variables are sufficient statistics.

Minimum Descriptor

In this thesis defined as some function of physical, chemical and biological index that represents important characteristics of morphogenetic process in embryology. Proposed by Nadine Peyriéras in Embryomics project to interpret the high-dimensional primary data with biological context.

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other French or foreign examination board.

The thesis work was conducted from 9/2006 to 3/2010 under the supervision of Prof. Jean Petitot and Paul Bourgine at CREA, Ecole Polytechnique.

Paris, 31/3/2010

Masatoshi Funabashi