EPFL

Thèse n° 9925

Positive Definite Completions and Continuous Graphical Models

Présentée le 25 septembre 2023

Faculté des sciences de base
Chaire de statistique mathématique
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Kartik WAGHMARE

Acceptée sur proposition du jury

Prof. T. Mountford, président du jury
Prof. V. Panaretos, directeur de thèse
Prof. T. Gneiting, rapporteur
Prof. M. Drton, rapporteur
Prof. A. Davison, rapporteur

École polytechnique fédérale de Lausanne

2023

Contents

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my advisor, Prof. Victor M. Panaretos. His constant encouragement and unwavering belief in my abilities were instrumental. He always made time for our discussions and shared advice that proved invaluable.

Secondly, I am grateful to the members of the thesis committee: Prof. Thomas Mountford, for acting as the president; and Prof. Tilmann Gneiting, Prof. Mathias Drton, and Prof. Anthony Davison, for agreeing to serve as experts, reading the original draft of this thesis, and offering generous advice on its revision.

I am eternally indebted to my parents, Prof. Govind and Vaishali Waghmare, for their unconditional support and encouragement of my academic pursuits.

I wish to thank my friends and colleagues at the Chair of Mathematical Statistics: Alessia, Almond, Ho Yun, Laya, Leonardo, Neda, Soham, Tomas Masak, and Tomas Rubin. I am particularly grateful to Tomas Masak and Leonardo for fostering a spirited and delightfully competitive environment in our office. Additionally, my heartfelt thanks extend to friends and colleagues outside the chair: Aditya P., Aditya H., Sylvie, Juraj, Marcello, Tim, Bhushan, Mario, Jonathan, Sonia, Vignesh, Soumaya, Akash, Charles, Aishwarya, Moritz, Jana, Alex, Matias, Lorenzo, Amit, Akhil, Delphine, Anna, Linda, Yoav, Motilal, Spandan, Nitin, and others. They broadened my horizons, introducing me to fresh perspectives on work and life, and were consistently excellent company.

I am also grateful to Dr. Guillaume Obozinski, Prof. Sofia Olhede, and Dr. Maria Süveges, under whom I served as a teaching assistant. I learned a great deal about teaching in general and statistics in particular from them.

Finally, I would like to thank the Swiss taxpayers whose generous funding made this research possible.

Lausanne, August 2, 2023 Kartik Waghmare

Abstract

This thesis concerns the theory of positive-definite completions and its mutually beneficial connections to the statistics of function-valued or continuously-indexed random processes, better known as *functional data analysis*. In particular, it dwells upon the reproducing kernel character of covariances.

In the introduction, we attempt to summarize the basic ideas and thoughts upon which the thesis is built.

Chapter 1 deals with the problem of covariance completion and develops an intuitive and interpretable approach to the problem of covariance estimation for fragmented functional data based on the concept of canonical completion. For a suitably restricted class of domains, we describe how the canonical completion may be constructed and use it to produce a characterization of the set of all completions. Furthermore, we identify necessary and sufficient conditions for uniqueness of completion and for the exact recovery of the true covariance.

Chapter 2 considers the problem of positive-definite completion in its generality and represents a purely mathematical treatment of the subject compared to Chapter 1. We study the problem for many classes of domains and present results concerning existence and uniqueness of solutions, their characterization and the existence and uniqueness of a special solution called canonical completion. We prove many new variational and algebraic characterizations of the canonical completion. Most importantly, we show the existence of canonical completion for the class of *band-like* domains. This leads to the existence of a canonical extension in the context of the classical problem of extensions of positive-definite functions, which is shown to correspond to a strongly continuous one-parameter semigroup and consequently, to an abstract Cauchy problem.

Chapter 3 presents a rigorous generalization of undirected Gaussian graphical models to arbitrary, possibly uncountable index sets. We prove an inverse zero characterization for these models, analogous the one known for multivariate graphical models and develop a procedure for their estimation based on the notion of resolution. The utility of the

concept and method is illustrated using real data applications and simulation studies.

Chapter 4 concerns the problem of recovering conditional independence relationships between a finite number of jointly distributed second-order Hilbertian random elements in a sparse high-dimensional regime with a particular interest in multivariate functional data. We propose an infinite-dimensional generalization of the multivariate graphical lasso and prove model selection consistency under natural assumptions. The method can be motivated from a coherent maximum likelihood philosophy

Chapter 5 discusses ways in which the results of this thesis can be strengthened or completed and its ideas and conclusions extended. With the only exception of Chapter 5, all chapters are independent self-contained articles and can be read in an arbitrary order although the given order is recommended.

Keywords: positive-definite completion, canonical completion, functional data analysis, covariance completion, continuous-time graphical models, reproducing kernel.

Résumé

Le sujet de cette thèse est la théorie des complétions positives-définies et ses liens mutuellement bénéfiques avec la statistique des processus aléatoires à valeurs fonctionnelles ou à indices continus, mieux connus sous le nom d'analyse des données fonctionnelles. En particulier, elle s'attarde sur le caractère de noyau reproducteur des covariances.

Dans l'introduction, nous tentons de résumer les idées et les réflexions fondamentales sur lesquelles repose la thèse.

Le chapitre 1 traite du problème de la complétion de la covariance et développe une approche intuitive et interprétable du problème de l'estimation de la covariance pour les fonctions fragmentées, basée sur le concept de complétion canonique. Pour une classe de domaines raisonnablement restreinte, nous décrivons comment la complétion canonique peut être construite et nous l'utilisons pour produire une caractérisation de l'ensemble de toutes les complétions. En outre, nous identifions conditions nécessaires et suffisantes pour l'unicité de la complétion et pour la récupération de la vraie covariance.

Le chapitre 2 considère le problème de la complétion positive-définie dans toute sa généralité et représente un traitement purement mathématique du sujet par rapport au chapitre 1. Nous étudions le problème pour de nombreuses classes de domaines et présentons des résultats concernant l'existence et l'unicité des solutions, leur caractérisation et l'existence et l'unicité d'une solution spéciale appelée complétion canonique. Nous prouvons de nombreuses nouvelles caractérisations variationnelles et algébriques de la complétion canonique. Surtout, nous montrons l'existence d'une complétion canonique pour la classe des domaines similaires à une bande. Il en résulte l'existence d'une extension canonique dans le contexte du problème classique des extensions des fonctions définies-positives, que nous montrons correspond à un semigroupe fortement continu et par conséquent, à un problème de Cauchy abstrait.

Le chapitre 3 présente une généralisation rigoureuse des modèles graphiques gaussiens non orientés à des ensembles d'indices arbitraires, potentiellement indénombrables. Nous prouvons une caractérisation de l'inverse du zéro pour ces modèles, analogue à celle connue pour les modèles graphiques multivariés et nous développons une procédure pour leur estimation basée sur la notion de résolution. L'utilité du concept et de la méthode est

illustrée à l'aide de nombreuses applications de données réelles et d'études de simulation.

Le chapitre 4 concerne le problème de la récupération des relations d'indépendance conditionnelle entre un nombre fini d'éléments aléatoires hilbertiens du second ordre distribués conjointement dans un régime à haute dimension éparse, avec un intérêt particulier pour les données fonctionnelles multivariées. Nous proposons une généralisation en dimension infinie du lasso graphique multivarié et prouvons la cohérence de la sélection de modèle sous des hypothèses naturelles. La méthode peut être motivée à partir d'une philosophie cohérente du maximum de vraisemblance.

Le chapitre 5 discute des façons dont les résultats de cette thèse peuvent être renforcés ou complétés et ses idées et conclusions étendues. À l'exception du chapitre 5, tous les chapitres sont des articles indépendants et autonomes et peuvent être lus dans un ordre arbitraire, bien que l'ordre indiqué soit recommandé.

Mots-clés : complétion positive-définie, complétion canonique, analyse des données fonctionnelles, complétion de la covariance, modèles graphiques à temps continu, noyau reproducteur.

Introduction

> As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality" it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art*. This need not be bad, if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste. But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much "abstract" inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up...
>
> In any event, whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the re-injection of more or less directly empirical ideas. I am convinced that this was a necessary condition to conserve the freshness and the vitality of the subject and that this will remain equally true in the future.

John von Neumann, *The Mathematician* (1947)

This thesis is centered around the theory of positive-definite completions and its connections to the statistics of function-valued or continuously-indexed random processes, better known as *functional data analysis*. The problem of positive-definite completion is a general formulation in terms of reproducing kernels which encompasses many classical problems such as positive-definite completions of partially specified matrices and extensions of positive-definite functions which have been studied in analysis (Krein, 1940) and linear algebra (Grone et al., 1984), and is intimately related to several others in operator theory (Gohberg et al., 1989) and probability (Parthasarathy and Varadhan, 1964).

Let X be a set and $\Omega \subset X \times X$. Given a function $K_\Omega : \Omega \to \mathbb{R}$, we are interested in determining whether it can be extended to a reproducing kernel K on X, which is to say,

a function $K : X \times X \to \mathbb{R}$ which satisfies $K(x, y) = K(y, x)$ for $x, y \in X$ and

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0$$

for every $n \geq 1$, $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$ and $\{x_j\}_{j=1}^n \subset X$. In other words, the matrix $[K(x_i, x_j)]_{i,j=1}^n$ is positive semidefinite for every $n \geq 1$ and $\{x_j\}_{j=1}^n \subset X$. If extension is indeed possible, then we are interested in studying the properties of the *completions* K. In particular, we are interested in the properties of a special solution $K_\star$ to the above problem, called the *canonical completion*.

Reproducing kernels on a set X can naturally be thought of as covariances of a Gaussian process indexed by X. The covariance of a Gaussian process being the canonical completion $K_\star$ endows the process with a peculiar conditional independence structure. In fact, the process can be regarded as a (possibly infinite) undirected Gaussian graphical model with the graph (X, Ω) which one obtains by considering X as the vertex set and Ω as the edge set. We shall see that this insight renders the beautiful paradigm of graphical modelling, which is very popular in multivariate statistics, accessible to functional data analysis. So far, this development had been obstructed by the fact that covariances of function-valued or continuously-indexed random processes are rather poorly endowed in terms of algebraic and analytic structure, in that, they do not have (well-behaved) inverses, a behaviour which is categorically different from covariance matrices that we encounter in multivariate statistics.

It is here that Aronszajn's theory of reproducing kernels rescues us with its radically different way of looking at things. Instead of treating our covariances as arrays of numbers or linear transformations of vectors in the usual way, it proposes that we think of them as corresponding to certain Hilbert spaces of functions on the index set. The correct analogue of the inverse of a covariance for our setting is then the inner product associated with the Hilbert space of the covariance.

Statistical Origins

The completion problem arose in connection with a covariance estimation problem for functional data. Consider a real-valued stochastic process $Y = \{Y_t\}_{t \in I}$ on an interval $I \subset \mathbb{R}$. We are interested in estimating the covariance of Y but instead of complete independent samples $Y^j \sim Y$ which are observed over the entire interval, we are given partially observed samples or *fragments* $Y^j|_J$ for $Y^j \sim Y$, where J is a much smaller subinterval of I. Consequently, familiar methods only permit us to estimate the covariance of Y over the region $\Omega \subset I \times I$ instead of $I \times I$. To produce an estimate of the covariance of Y over $I \times I$, we need to extrapolate the partial estimate from the region Ω to the region $\Omega^c = (I \times I) \setminus \Omega$ where the covariance can't be estimated. The extrapolation procedure

must ensure that the resulting estimate is at least approximately positive-definite. If we think of I as X and the partial estimate of the covariance as K_Ω, this can be understood as an approximate form of the positive definite completion problem.

Our first construction of canonical completion was also achieved by reflecting on how one might approach the problem of linear prediction when the entire covariance of the process is not available. In fact, the basic idea is so simple that we shall venture to give it here.

For simplicity, assume that $\Omega = (I_1 \times I_1) \cup (I_2 \times I_2)$ where I_1 are I_2 subintervals of I such that $I_1 \cup I_2 = I$. Furthermore, assume that we are given $K_\Omega : \Omega \to \mathbb{R}$ such that the restrictions $K_1 = K_\Omega|_{I_1 \times I_1}$ and $K_2 = K_\Omega|_{I_2 \times I_2}$ are reproducing kernels. We shall construct a Gaussian process $Y_\star = \{Y_\star(t)\}_{t \in I}$ whose covariance $K_\star$ is a completion of K_Ω. Let $Y_1 = \{Y_1(t)\}_{t \in I_1}$ and $Y_2 = \{Y_2(t)\}_{t \in I_2}$ be Gaussian processes with mean zero and covariance K_1 and K_2, respectively. Also, let $J = I_1 \cap I_2$.

Define $Y_\star = \{Y_\star(t)\}_{t \in I}$ as follows:

$$Y_\star(t) = \begin{cases} Y_1(t) & \text{for } t \in I_1, \\ \hat{Y}(t) + Z(t) & \text{for } t \in I_2 \setminus I_1, \end{cases}$$

where

$$\hat{Y}(t) = \mathbb{E}[Y_2(t) \mid Y_2(u) = Y_1(u) \text{ for } u \in J]$$

and $Z = \{Z(t) : t \in I_2 \setminus I_1\}$ is an independent Gaussian process with mean zero and covariance $K_Z(s,t) = \mathbb{E}[W(s)W(t)]$ for $s, t \in I_2 \setminus I_1$ with

$$W(t) = Y_2(t) - \mathbb{E}[Y_2(t) \mid Y_2(u) \text{ for } u \in J]$$

for $t \in I_2 \setminus I_1$. Taking conditional expectations in this manner requires some additional theoretical justification. However, if we were to interpret them in a formal way, some, slightly tedious, verification would reveal that the covariance $K_\star$ of $Y_\star$ is indeed a completion of K_Ω. It turned out that this relatively innocuous construction has many remarkable properties, perhaps most important of them being that $Y_\star$ can be thought of as a graphical model with the graph Ω. For this reason, $K_\star$ is called canonical completion.

It is worth pointing out that the canonical completion for the special case of symmetric Toeplitz matrices made its first appearance also in connection with the statistical problem of autocovariance estimation for stationary time series in the work of J. P. Burg (Burg, 1975).

Synopsis

In Chapter 1, we refine the said construction by abstracting away the Gaussian processes with curves of vectors in a reproducing kernel Hilbert space and extend it to more general domains. Furthermore, we develop a theory of covariance completion which comprehensively answers many of the questions posed by the problem of covariance estimation with functional fragments. In particular, we show how the knowledge of the "graph" of the underlying process is crucial to the problem.

In Chapter 2, we attempt to exploit the rich structure provided by the theory of reproducing kernels to its fullest extent towards studying the completion problem. We study the completion problem for a variety of domains and derive many new and interesting results in addition to recovering several old ones with great ease. In particular, we prove the existence of a canonical extension to the classic positive-definite extension problem. The relatively innocuous original construction of the canonical completion has been refined beyond recognition and been complemented by many different characterization of an algebraic or variational nature. In comparison to Chapter 1, Chapter 2 is a completely mathematical treatment of the subject without any statistical preoccupations whatsoever.

Chapter 1 tells us how to augment covariance estimation for a process using the knowledge of its graph. In Chapter 3, we take the opposite perspective and attempt to solve the inverse problem: recovering the graph of the process from its covariance. One of the strengths of the reproducing kernel formalism is that it can describe the graphical structure of a Gaussian process in terms of the covariance without requiring the covariance to have a well-behaved inverse. This allows us to characterize the graph of a continuous-time Gaussian process in terms of its covariance and develop a method for its estimation.

Prior to coming up with the reproducing kernel approach, we attempted to solve the completion problem for reproducing kernels using what can be understood as a maximum entropy approach. Basically, for matrices there is a functional, namely the matrix determinant, which has this extraordinary property that maximizing the functional over the space of all (not necessarily positive-definite) completions yields the canonical completion. We attempted to construct a similar functional for kernels using the Fredholm determinant. Unfortunately, it became apparent that for the method to work, it would be necessary for the kernels to satisfy some very stringent and unnatural conditions. Thankfully, the tools acquired in pursuing this approach, such as Radon-Nikodym derivatives of Gaussian measures and Fredholm determinants, could be easily salvaged and appropriated for solving a quite different problem of recovering conditional independence graphs for multivariate functional data. Chapter 4 is a product of this line of research.

1 The Completion of Covariance Kernels

Es gibt nichts Praktischeres als eine gute Theorie.

There is nothing more practical than a good theory.

Ludwig Boltzmann

Abstract[1]

We consider the problem of positive-semidefinite continuation: extending a partially specified covariance kernel from a subdomain Ω of a rectangular domain $I \times I$ to a covariance kernel on the entire domain $I \times I$. For a broad class of domains Ω called *serrated domains*, we are able to present a complete theory. Namely, we demonstrate that a canonical completion always exists and can be explicitly constructed. We characterise all possible completions as suitable perturbations of the canonical completion, and determine necessary and sufficient conditions for a unique completion to exist. We interpret the canonical completion via the graphical model structure it induces on the associated Gaussian process. Furthermore, we show how the estimation of the canonical completion reduces to the solution of a system of linear statistical inverse problems in the space of Hilbert-Schmidt operators, and derive rates of convergence. We conclude by providing extensions of our theory to more general forms of domains, and by demonstrating how our results can be used to construct covariance estimators from sample path fragments of the associated stochastic process. Our results are illustrated numerically by way of a simulation study and a real example.

[1]The chapter has been adapted from the article: Waghmare, K. G. and Panaretos, V. M. (2022). The completion of covariance kernels. *The Annals of Statistics*, 50(6):3281 – 3306.

1.1 Introduction

Consider a bivariate function $K_\Omega : \Omega \to \mathbb{R}$ where Ω is a subset of $I \times I$ for some bounded interval $I \subset \mathbb{R}$. An extension of K_Ω to a covariance kernel on I is called a *completion*. Under appropriate conditions on Ω and K_Ω, we would like to answer the following questions: Does a completion always exist? Is there a canonical completion and can we construct it explicitly? Can we characterise the set of all completions? Can we give necessary and sufficient conditions for a unique completion to exist? Is a unique completion necessarily canonical?

Such questions are arguably very natural from a mathematical point of view, with connections to the trigonometric moment problem and the continuation of characteristic functions, via Bochner's theorem. They are well understood for covariance matrices and for stationary/isotropic kernels. It appears that the study of positive-definite completions was initiated by Carathéodory (1907), who showed that every positive-definite function on a subset $\{j \in \mathbb{Z} : |j| \leq n\}$ of $\mathbb{Z}$ extends to a positive-definite function on $\mathbb{Z}$. A continuous analogue of this result was proved by Krein (1940) for continuous positive-definite functions on a symmetric interval of the real line, and the problem of uniqueness as well as that of description of all extensions in case of non-uniqueness was considered in the same context by Krein and Langer (2014). Higher-dimensional versions of the problem were considered by Calderón and Pepinsky (1952) and Rudin (1963), whose results were extended to discontinuous kernels by Artjomenko (1941b) and Gneiting and Sasvári (1999). A short survey of these developments can be found in Sasvári (2006). The case of positive-definite completions of banded matrices was considered by Dym and Gohberg (1981). The general case was treated by Grone et al. (1984). Many of these results have been further extended to matrices with operator entries in Gohberg et al. (1989) and Paulsen et al. (1989). An extensive survey concerning the importance of positive-definite functions and kernels can be found in Stewart (1976).

Our interest is to obtain a theory for the case where I is an interval and K_Ω is not constrained to satisfy invariance properties such as stationarity. Besides the intrinsically mathematical motivation for developing such extensions, we are motivated by *the problem of covariance estimation from sample path fragments*. Namely, estimating the covariance of a (potentially non-stationary) second-order process $X = \{X_t : t \in I\}$ on the basis of i.i.d. sample paths $\{X_j\}$ censored outside subintervals $\{I_j\}$, i.e. on the basis of *fragments* $X_j|_{I_j}$ drawn from $X_{I_j} = \{X_t : t \in I_j\}$ for a collection of subintervals $\{I_j\}$ of I. Because of the fragmented nature of the observations, one is only able to estimate a restriction K_Ω of K to a symmetric region, say $\Omega \subset I \times I$ centered around the diagonal. Nevertheless, one needs an estimator of the full covariance K, as this is necessary for further statistical analysis – tasks like dimension reduction, regression, testing, and classification require the complete covariance. The problem thus reduces to ascertaining how and under what conditions one can estimate K from an estimate $\hat{K}_\Omega$ of K_Ω. This problem arises in a range of contexts, as documented in the references in the next paragraph. For instance,

in longitudinal studies where a continuously varying random quantity (e.g. bone mineral density or systolic blood pressure) is measured on each study subject over a short time interval (see Section 1.11 for a presentation and analysis of such an example), or in the modeling of hourly electricity pricing, where price functions are only partially observed.

Kraus (2015) originally introduced and studied a simpler version of this problem, where some samples were observable over the entire domain, hence resulting in reduced rather than no information outside Ω. Delaigle and Hall (2016) were the first to attack the genuinely fragmented problem, by imposing a (discrete) Markov assumption. Though their approach also yielded a completion, it was more focused on predicting the missing segments. Similarly, Kneip and Liebl (2020) focused on how to optimally reconstruct the missing segments using linear prediction. The problem has been recently revisited with a firm focus on the identifiability and estimation of the complete covariance itself, see Descary and Panaretos (2019b), Delaigle et al. (2021) and Lin et al. (2021). At a high level, they all proceed by (differently) penalized least squares fitting of a finite-rank tensor product expansion over the region Ω, which is then used to extrapolate the covariance beyond Ω. While there are substantial differences in their set up and technique, common to all three approaches is the pursuit of sufficient conditions on the process X for identifiability to hold, i.e. for a uniquely existing completion. Imposing such conditions *a priori* ensures that extrapolation is sensible. Starting with a strong condition in Descary and Panaretos (2019b) (namely, analyticity), these sufficient conditions have progressively been weakened, albeit not to the point of attaining conditions that are also necessary.

We shall pursue a different approach to the problem, which we believe sheds more insight, and ultimately leads to necessary and sufficient conditions for uniqueness. Rather than start by focusing on uniqueness, we will aim at a comprehensive description of the set of all valid completions from a broad class of domains Ω called *serrated domains*. Specifically, we will exhibit that a *canonical completion* can always be explicitly and uniquely constructed (Section 1.3). Canonicity will be clearly interpreted by means of a graphical model structure on the associated Gaussian process (Section 1.4). We will then obtain necessary and sufficient conditions for a unique completion to exist, and discuss how these relate to the problem of identifiability (Section 1.5). Furthermore, we will constructively characterise the set of completions as suitable perturbations of the canonical completion (i.e. show how any other valid completion can be built using the canonical completion; see Section 1.6) and parametrize this set in terms of contractions between certain L^2 spaces (Theorem 1.7). As for the statistical side of the question related to fragments, since a canonical solution always exists uniquely, and is equivalent to the unique completion when uniqueness holds, it is always an identifiable and interpretable target of estimation. We thus consider how to estimate it based on an estimator of the observed partial covariance, say $\hat{K}_\Omega$, and provide rates of convergence in Section 1.7. We then show how our results can be adapted to more general domains Ω in Section 1.8. This allows us to give a treatment of the statistical problem of covariance estimation from sample path fragments in Section 1.9. The finite sample performance of

our statistical methodology is investigated by means of an extensive simulation study (Section 1.10) and a data analysis (Section 1.11). The proofs of our results are collected in the Supplementary Material.

Our general perspective is inspired by the work of Grone et al. (1984) and Dym and Gohberg (1981) on matrices. Our methods, however, are very different, because algebraic tools such as determinants and matrix factorization that are elemental to those works are unavailable in the kernel case. Instead, we generalize the concept of canonical extension Dym and Gohberg (1981) (or determinant-maximizing completion Grone et al. (1984)) to a general kernel version by demonstrating and exploiting its intimate connection to Reproducing Kernel Hilbert Spaces (RKHS) and graphical models for random processes (Theorem 1.4). An apparent consequence is that our necessary and sufficient conditions for a partial covariance to complete uniquely (Theorem 1.5) seem to be novel even in the context of matrices.

1.2 Background and Notation

To set the context of the problem, we delineate the functions K_Ω that are admissible as partial covariances, and the types of domains Ω under consideration. Recall that $K : I \times I \to \mathbb{R}$ is a covariance kernel on I if

1. $K(s,t) = K(t,s)$ for $s,t \in I$, and
2. $\sum_{i,j=1}^{n} \alpha_i \alpha_j K(t_i,t_j) \geq 0$ for $n \geq 1$, $\{t_i\}_{i=1}^{n} \subset I$ and $\{\alpha_i\}_{i=1}^{n} \subset \mathbb{R}$.

We shall denote the set of covariances on I by $\mathcal{C}$. We shall say that $\Omega \subset I \times I$ is a *symmetric domain* if $(s,t) \in \Omega$ if $(t,s) \in \Omega$ for $s,t \in I$ and $\{(t,t) : t \in I\} \subset \Omega$. Since a covariance is always defined over square domains, it is natural to define partial covariances as follows:

Definition 1.1 (Partial Covariance)**.** *Let I be a set and $\Omega \subset I \times I$ be a symmetric domain. A function $K_\Omega : \Omega \to \mathbb{R}$ is called a partial covariance on Ω if for every $J \subset I$ such that $J \times J \subset \Omega$, the restriction $K_J = K_\Omega|_{J\times J}$ is a covariance on J.*

In the above definition, the set J need not be an interval.

Remark 1.1 (On Notation)**.** *Whenever we write K_J for some $J \subset I$, we will always understand that this refers to the restriction $K_\Omega|_{J\times J}$ of the partial covariance K_Ω to the square $J \times J \subseteq \Omega$.*

A completion of the partial covariance K_Ω will be a function $K : I \times I \to \mathbb{R}$ such that

$$K \in \mathcal{C} \text{ and } K|_\Omega = K_\Omega. \tag{1.1}$$

The set of all possible completions of K_Ω will be denoted by

$$\mathcal{C}(K_\Omega) = \{K \in \mathcal{C} : K|_\Omega = K_\Omega\}.$$

Note that our definition of partial covariance does not *a priori* assume that K_Ω arises as the restriction of a covariance K on I. Rather, it defines K_Ω intrinsically on Ω. In this sense, our setting is more general than the functional fragment setting. Consequently, $\mathcal{C}(K_\Omega)$ is not automatically non-empty. Notice however that if $K_1, K_2 \in \mathcal{C}(K_\Omega)$ then $\alpha K_1 + (1-\alpha)K_2 \in \mathcal{C}(K_\Omega)$ for every $\alpha \in (0,1)$. $\mathcal{C}(K_\Omega)$ is thus *convex*. It is also poitwise bounded so long as $\sup_{t\in I} K_\Omega(t,t) < \infty$ because for every $K \in \mathcal{C}(K_\Omega)$ we have $|K(s,t)| \leq \sqrt{K(s,s)K(t,t)} \leq \sup_{t\in I} K_\Omega(t,t)$. It follows from convexity that $\mathcal{C}(K_\Omega)$ can either be an empty set, a singleton or have an (uncountably) infinite number of elements. Finally, the elements of $\mathcal{C}(K_\Omega)$ inherit the regularity properties of K_Ω. In particular, if $K_\Omega \in C^{k,k}(\Omega)$, then $K \in C^{k,k}(I \times I)$, where $C^{k,k}(\Delta)$ for a domain $\Delta \subset I \times I$ denotes the set of functions $F : \Delta \to \mathbb{R}$ such that the partial derivatives $\partial_y^j \partial_x^i F(x,y)$ and $\partial_x^i \partial_y^j F(x,y)$ exist for $0 \leq i,j \leq k$. This is a direct consequence of the fact that the process X is k-differentiable in quadratic mean if and only its covariance's partial derivatives $\partial_y^j \partial_x^i K$ and $\partial_x^i \partial_y^j K$ exist for $0 \leq i,j \leq k$ at the diagonal $\{(x,x) : x \in I\}$ (see Loeve (2017) or Saitoh and Sawano (2016)). Indeed, if $K_\Omega \in C^{k,k}(\Omega)$, then the corresponding process X is k-differentiable in quadratic mean and hence $K \in C^{k,k}(I \times I)$.

In some cases, we will need to work with the covariance *operators* associated with the corresponding covariance kernels. For a measure μ on the Borel sets of I, and $S \subseteq I$ we define the Hilbert space $L^2(S)$ to be the set of all $f : S \to \mathbb{R}$ such that $\int_S f^2(x)\, \mu(dx) < \infty$ with associated inner product

$$\langle f, g \rangle_2 = \int_S f(x) g(x) \mu(dx), \qquad f, g \in L^2(S).$$

Since continuity of K_Ω implies continuity of any completion thereof, any completion K induces a Hilbert-Schmidt integral operator $\mathbf{K} : L^2(I) \to L^2(I)$ given by

$$\mathbf{K} f(x) = \int_I K(x,y) f(y) \mu(dy), \qquad \mu - \text{a.e.},$$

i.e. an operator with K as its integral kernel. Similarly, any restriction $K|_{S\times S}$ on a square domain induces an integral operator $\mathbf{K}|_S : L^2(S) \to L^2(S)$ by way of

$$\mathbf{K}|_S\, g(x) = \int_S K(x,y) g(y)\ \mu(dy), \qquad \mu|_S - \text{a.e.}$$

The operator norm $\|\cdot\|_\infty$ and Hilbert-Schmidt norm $\|\cdot\|_2$ of an operator $\mathbf{K} : L^2(S_1) \to$

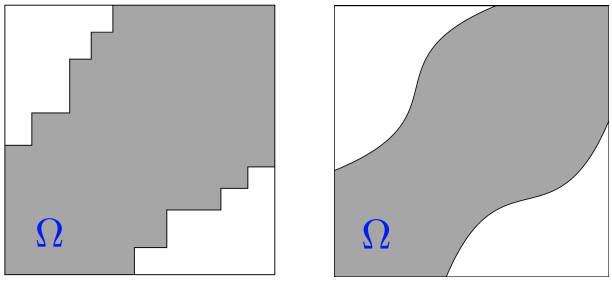


Figure 1.1: A serrated domain (left) and a nearly serrated domain (right).

$L^2(S_2)$, $S_1, S_2 \subseteq I$, with continuous kernel $K : S_1 \times S_2 \to \mathbb{R}$ will be defined via

$$\|\mathbf{K}\|_\infty^2 = \sup_{f \in L^2(S_1)\setminus\{0\}} \frac{\int_{S_2} \left(\int_{S_1} K(u,v) f(v) \mu(dv) \right)^2 \mu(du)}{\int_{S_1} f^2(u) \mu(du)}$$

and

$$\|\mathbf{K}\|_2^2 = \int_{S_1} \int_{S_2} K^2(u,v) \mu(du) \mu(dv).$$

The positive root of an operator $\mathbf{A}$ will be denoted by $|\mathbf{A}| = (\mathbf{A}\mathbf{A}^*)^{1/2}$. We denote the space of Hilbert-Schmidt operators from $L^2(S_1)$ to $L^2(S_2)$ as $\mathcal{S}_2(S_1, S_2)$. The image of a subset $\mathcal{S}_1 \subseteq L^2(S_1)$ via the operator $\mathbf{K} : L^2(S_1) \to L^2(S_2)$ will be simply denoted as $\mathbf{K}\mathcal{S}_1 = \{\mathbf{K}f : f \in \mathcal{S}_1\}$. We shall use the same convention for operator multiplication, for example, we denote $\mathbf{K}\mathcal{S}_2(S_3, S_1) = \{\mathbf{K}\mathbf{A} : \mathbf{A} \in \mathcal{S}_2(S_3, S_1)\}$.

Given a Hilbert-Schmidt operator $\mathbf{K} : L^2(S_1) \to L^2(S_1)$ with integral kernel $K : S_1 \times S_1 \to \mathbb{R}$, we define the Reproducing Kernel Hilbert Space (RKHS) of K (equivalently of $\mathbf{K}$) as the Hilbert space $\mathcal{H}(K) = \mathbf{K}^{1/2} L^2(S_1)$, endowed with the inner product

$$\langle f, g \rangle_{\mathcal{H}(K)} := \langle \mathbf{K}^{-1/2} f, \mathbf{K}^{-1/2} g \rangle_{L^2(S_1)}, \qquad f, g \in \mathcal{H}(K).$$

As for the types of symmetric domains Ω under consideration, our main focus will be on serrated domains:

Definition 1.2 (Serrated Domain)**.** *Let $I \subset \mathbb{R}$ be a bounded interval. A domain $\Omega \subseteq I \times I$ is called serrated if it can be written as a union $\Omega = \cup_j (I_j \times I_j)$ for $\{I_j\}$ a finite cover of I comprised of subintervals $I_j \subseteq I$.*

Informally, a serrated domain consists of a collection of squares of varying sizes, strung symmetrically along the diagonal in a manner that covers it (see Figure 1.1). When restricting attention to matrices or stationary kernels, serrated domains reduce to the types of domains on which the problem has been previously studied. In the functional fragments problem, the observation of a finite collection of path fragments $X_j|_{I_j}$ leads to partial covariance information on the serrated domain $\Omega = \cup_j (I_j \times I_j)$. By taking

sequences of covers consisting of progressively more squares, serrated domains can approximate a very rich class of symmetric domains that we call *nearly serrated* (see Figure 1.1 and Definition 1.4 for a rigorous definition). In the next sections, we develop an essentially complete theory of completion for serrated domains. Then, Section 1.8 demonstrates how our results on serrated domains can be used to obtain results for nearly serrated domains.

1.3 The Canonical Completion

Recall that the set of completions $\mathcal{C}(K_\Omega)$ of a partial covariance K_Ω can be empty, a singleton, or uncountably infinite. We will now show that for Ω a serrated domain, $\mathcal{C}(K_\Omega)$ is never empty. We will do so by explicitly constructing a completion $K_\star$, that will be subsequently argued to be canonical.

It is instructive to commence with the *2-serrated case*, i.e. when $\Omega = (I_1 \times I_1) \cup (I_2 \times I_2)$ for two intervals $\{I_1, I_2\}$ such that $I_1 \cup I_2 = I$, depicted in Figure 1.2 (left). Define a function $K_\star : I \times I \to \mathbb{R}$ as follows:

$$K_\star(s,t) = \begin{cases} K_\Omega(s,t), & (s,t) \in \Omega \\ \Big\langle K_\Omega(s,\cdot), K_\Omega(\cdot,t) \Big\rangle_{\mathcal{H}(K_{I_1 \cap I_2})}, & (s,t) \notin \Omega. \end{cases} \tag{1.2}$$

Here, $K_{I_1 \cap I_2} = K_\Omega|_{(I_1 \cap I_2)^2}$ is the restriction of the partial covariance K_Ω to the square $(I_1 \cap I_2) \times (I_1 \cap I_2)$ and $\mathcal{H}(K_{I_1 \cap I_2})$ is the RKHS of $K_{I_1 \cap I_2}$. It is implicit in the notation $\langle K_\Omega(s,\cdot), K_\Omega(\cdot,t) \rangle_{\mathcal{H}(K_{I_1 \cap I_2})}$ that the domain of $K_\Omega(s,\cdot)$ and $K_\Omega(\cdot,t)$ is automatically restricted to $I_1 \cap I_2$ within that inner product, as depicted in Figure 1.2 (right).

Remark 1.2. *The reproducing kernel inner product in Equation 1.2 can be seen as the infinite-dimensional equivalent of matrix multiplication formulas appearing in maximum entropy matrix completion (Johnson, 1990) and low rank matrix completion (Descary and Panaretos, 2019a).*

Our first result is now:

Theorem 1.1 (Canonical Completion from a 2-Serrated Domain)**.** *Given any partial covariance K_Ω on a 2-serrated domain $\Omega \subseteq I \times I$, the function $K_\star : I \times I \to \mathbb{R}$ defined in* (1.2) *is a well-defined covariance that constitutes a valid completion, i.e.*

$$K_\star \in \mathcal{C}(K_\Omega).$$

In particular, if K_Ω admits a unique completion, then this must equal $K_\star$.

The second part of the theorem hints at why we refer to the completion $K_\star$ as the *canonical completion* of K_Ω. We will provide a more definitive reason in Section 1.4, but

first we will use the formula from the 2-serrated case in order to extend our result to a general serrated domain.

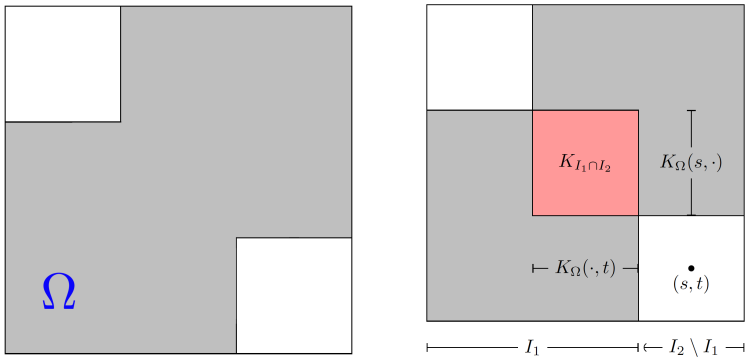


Figure 1.2: A two serrated domain (left) and a heuristic illustration of the formula for $K_\star$.

We will do this iteratively. Intuitively, if we have a general serrated domain generated by a cover of m subintervals $\{I_1, ..., I_m\}$, we can apply the 2-serrated formula to any pair of successive squares $\{I_p^2, I_{p+1}^2\}$, to reduce to the problem to one of completion from a serrated domain generated the reduced set of $m-1$ subintervals $\{I_1, ..., I_{p-1}, I_p \cup I_{p+1}, I_{p+2}, ..., I_m\}$ (see Figure 1.3). Repeating the same prescription, we can eventually complete K_Ω to a covariance on I. To be more precise, let $\Omega = \cup_{j=1}^m (I_j \times I_j)$ be an m-serrated domain and for notational ease assume that the indices of the $\{I_j\}$ correspond to their natural partial ordering as intervals. Define the intersection of any two successive squares as

$$J_p = (I_p \times I_p) \cap (I_{p+1} \times I_{p+1})$$

and the corresponding restriction of K_Ω as $K_{J_p} = K_\Omega|_{J_p \times J_p}$. Next define the square of the union of the intervals $\{I_1, ..., I_p\}$ as

$$U_p = (I_1 \cup \ldots \cup I_p) \times (I_1 \cup \ldots \cup I_p).$$

Finally, define the serrated domain generated by the cover $\{\cup_{j=1}^p I_j, I_{p+1}, ..., I_m\}$ as

$$\Omega_p = U_p \bigcup \left\{ \cup_{j=p+1}^m (I_j \times I_j) \right\}$$

noting that $\Omega_1 = \Omega$.

The following algorithm uses the formula from the 2-serrated case to extend K_Ω to a partial covariance on Ω_2, then Ω_3, and so on, until completion to covariance on $I^2 = \Omega_m$:

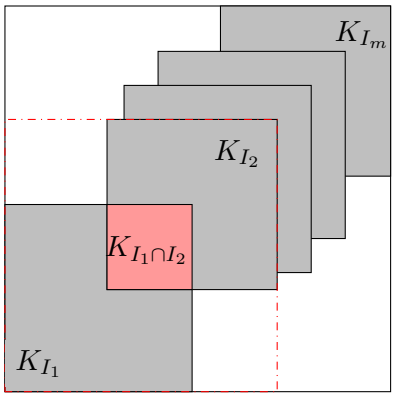


Figure 1.3: Illustration of the iterative completion procedure for a general serrated domain.

Table 1 m-Serrated Completion by Successive 2-Serrated Completions

1. Initialise with the partial covariance $K_1 = K_\Omega$ on $\Omega_1 = \Omega$.

2. For $p \in \{1, ..., m-1\}$ define the partial covariance K_{p+1} on Ω_{p+1} as

$$K_{p+1}(s,t) = \begin{cases} K_p(s,t), & (s,t) \in \Omega_p \\ \left\langle K_p(s,\cdot), K_p(\cdot,t) \right\rangle_{\mathcal{H}(K_{J_p})}, & (s,t) \in \Omega_{p+1} \setminus \Omega_p. \end{cases}$$

3. Output the covariance $K_\star = K_m$ on $I \times I = \Omega_m$.

Of course, there is nothing special about the application of the iterative completion in ascending order. We could have set up our notation and algorithm using a descending order starting with $\{I_m^2, I_{m-1}^2\}$, or indeed using an arbitrary order, starting from any pair of successive squares $\{I_p^2, I_{p+1}^2\}$ and moving up and down to neighbouring squares. Our second result shows that, no matter the chosen order, the algorithm will output the same valid completion $K_\star \in \mathcal{C}(K_\Omega)$:

Theorem 1.2 (Canonical Completion from a General Serrated Domain)**.** *The recursive application of the 2-serrated formula as described in Algorithm* (1) *to a partial covariance K_Ω on a serrated domain Ω yields the same valid completion $K_\star \in \mathcal{C}(K_\Omega)$, irrespective of the order it is applied in. In particular, if K_Ω admits a unique completion, then this must equal $K_\star$.*

Notice that Theorems 1.1 and 1.2 make no assumption on K_Ω except that it be a partial covariance. In particular, K_Ω need not be continuous or even bounded.

Example 1.1 (Brownian Motion)**.** *As an example, consider the following partial covariance on a 2-serrated subdomain of* $[0,1]^2$*:*

$$K_\Omega(s,t) = s \wedge t, \quad (s,t) \in \Omega = \underbrace{([0,2/3] \times [0,2/3])}_{I_1 \times I_1} \cup \underbrace{([1/3,1] \times [1/3,1])}_{I_2 \times I_2}.$$

Clearly, this can be completed to the covariance of standard Brownian motion on $[0,1]^2$*,*

$$K(s,t) = s \wedge t, \qquad (s,t) \in [0,1]^2.$$

To see what our completion algorithm yields, we note that the restriction $K_{[1/3,2/3]}$ *yields the RKHS inner product*

$$\langle f, g \rangle_{\mathcal{H}(K_{[1/3,2/3]})} = \frac{1}{(1/3)} f(1/3) g(1/3) + \int_{1/3}^{2/3} f'(u) g'(u)\, du,$$

,

$$K_\star(s,t) = 3 \cdot (s \wedge 1/3)(t \wedge 1/3) + \int_{1/3}^{2/3} 0\, du = 3 \cdot (1/3) \cdot t = t = s \wedge t, \qquad \textit{since } t < s.$$

Iterating, we can directly see that the extension of a partial covariance that has the form $s \wedge t$ *on an arbitrary domain by means of Algorithm 1 will also yield the covariance of Brownian motion.*

The example illustrates that Algorithm 1 yields the "right" answer in an important special case. The next section demonstrates that this is no accident, and that the completions given in Theorems 1.1 and 1.2 are indeed canonical in a strong sense.

1.4 Canonicity and Graphical Models

We will now interpret the canonical completion via the conditional independence structure it induces on the associated Gaussian process. Recall that an undirected graph G on a set I is an ordered pair $G = (I, \Omega)$ where I is called the vertex set and $\Omega \subseteq I \times I$ is the edge set such that $(s,t) \in \Omega$ if and only if $(t,s) \in \Omega$. We shall often refer to the graph (I, Ω) as Ω. Notice that if I is an interval of the real line, then a symmetric domain Ω induces an uncountable graph on I with Ω serving as the edge set.

We shall say that $S \subset I$ *separates* $s, t \in I$ with respect to the graph (I, Ω) if every path from s to t comprised of edges in Ω is intercepted by S, that is, for every $\{t_i\}_{i=1}^r \subset I$ with $r \geq 1$ such that $t_1 = s$, $(t_i, t_{i+1}) \in \Omega$ for $1 \leq i < r$ and $t_r = t$, we have that $t_j \in S$ for some $1 < j < r$.

A graph (I, Ω) induces a conditional independence structure on a Gaussian process

$X_I := \{X_t : t \in I\}$ much in the same way as in the finite dimensional case.

Definition 1.3 (Graphical Models on Gaussian Processes)**.** *The Gaussian process $X = \{X_t : t \in I\}$ is said to form an undirected graphical model over the graph $\Omega \subseteq I \times I$, if for every $s, t \in I$ separated by $J \subset I$, we have*

$$\mathrm{Cov}(X_s, X_t | X_J) \equiv \mathbb{E}\left[(X_s - \mathbb{E}\left[X_s|X_J\right])(X_t - \mathbb{E}\left[X_t|X_J\right])|X_J\right] = 0 \quad a.s. \tag{1.3}$$

Equation (1.3) implies that $\mathbb{E}\left[X_s X_t|X_J\right] = \mathbb{E}\left[X_s|X_J\right]\mathbb{E}\left[X_t|X_J\right]$ almost surely. Taking the expectation gives

$$\mathbb{E}\left[X_s X_t\right] = \mathbb{E}\left[\mathbb{E}\left[X_s|X_J\right]\mathbb{E}\left[X_t|X_J\right]\right], \tag{1.4}$$

i.e. the covariance of X_s and X_t coincides with that of their best predictors given X_J when J separates s and t. Notice that from (1.4), it follows that

$$\mathbb{E}\left[\mathbb{E}\left[X_s|X_J\right]\mathbb{E}\left[X_t|X_J\right]\right] = \mathbb{E}\left[\mathbb{E}\left[X_s\mathbb{E}\left[X_t|X_J\right]|X_J\right]\right] = \mathbb{E}\left[X_s\mathbb{E}\left[X_t|X_J\right]\right]$$

and thus, $\mathbb{E}\left[(X_t - \mathbb{E}[X_t|X_J])X_s\right] = 0$ which implies that $\mathbb{E}[X_t|X_J] = \mathbb{E}[X_t|X_J, X_s]$ by the projection theorem. Similar reasoning yields,

$$\mathbb{E}\left[f(X_t)|X_J, X_s\right] = \mathbb{E}\left[f(X_t)|X_J\right], \tag{1.5}$$

which is reminiscent of Markov processes, where

$$\mathbb{E}\left[f(X_t)|\{X_u : u \leq v\}\right] = \mathbb{E}\left[f(X_t)|X_v\right]. \tag{1.6}$$

Indeed, the undirected graphical model structure induced by Ω is a natural generalization of the Markov property, but with a notion of separation stemming from the graph structure rather than simple time ordering. In the terminology of Markov random fields, Definition 1.3 is equivalent to the global Markov property with respect to Ω.

Theorem 1.3. *Let K_Ω be a partial covariance on a serrated domain $\Omega \subset I$. The canonical completion $K_\star$ is the only completion of K_Ω such that the associated Gaussian process $\{X_t : t \in I\}$ forms an undirected graphical model with respect to the graph $G = ([0, 1], \Omega)$.*

Said differently, $K_\star$ is the only completion of K_Ω that possesses the global Markov property with respect to the edge set Ω. Intuitively, the canonical completion is the unique completion to rely exclusively on correlations intrinsic to the "observed" set Ω: it propagates the "observable" correlations of K_Ω to the rest of I via the Markov property, without introducing any extrinsic "unobserved" correlations. By contrast, any other completion will introduce correlations extrinsic to those observed via K_Ω. This last statement is considerably refined in Section 1.6, where we characterise all possible completions as perturbations of the canonical completion.

In closing this section, we give a result going in the opposite direction: namely we show that a Gaussian process admits a graphical model structure w.r.t. a serrated Ω if and only if it has a covariance that satisfies the defining equations (1.2) of a canonical completion. This is a result that is of interest in its own right, since it characterises the set of all Gaussian process graphical models compatible with Ω. Because this characterization is pointwise in nature, it provides an arguably more convenient way of expressing conditional independence relations in a Gaussian process than, say, cross-covariance operators defined by Baker (1973). To state it rigorously, define the set of covariances

$$\mathcal{G}_\Omega = \left\{ K \in \mathcal{C} : K(s,t) = \Big\langle K(s,\cdot), K(\cdot,t) \Big\rangle_{\mathcal{H}(K_J)} \text{ for all } J \subset I \text{ separating } s,t \in I \text{ in } \Omega \right\}.$$

We can now state:

Theorem 1.4. *Let $\{X_t : t \in I\}$ be a Gaussian process with covariance K. Then, X forms an undirected graphical model with respect to a serrated Ω if and only if $K \in \mathcal{G}_\Omega$.*

There is actually no reason to restrict attention to Gaussian processes, and we did this solely for interpretability: for a Gaussian process, the condition $K \in \mathcal{G}_\Omega$ can be interpreted in terms of conditional independence. But we can more generally define a second-order graphical model as long as we focus solely on conditional uncorrelatedness rather than conditional independence – just take Definition 1.3 and drop the word "Gaussian", while replacing "graphical model" by "second order graphical model".

1.5 Necessary and Sufficient Conditions for Unique Completion

We will now state necessary and sufficient conditions guaranteeing unique completion from a serrated domain $\Omega \subset I$. And we will argue that identifiability can occur even without enforcing the existence of a unique extension. For this, we need some additional notation. Given $A \subset B \subset \Omega$, let K_B/K_A be the Schur complement of K_B with respect to K_A,

$$(K_B/K_A)(s,t) = K_B(s,t) - \Big\langle K_B(s,\cdot), K_B(\cdot,t) \Big\rangle_{\mathcal{H}(K_A)}$$

, that is, the covariance of the *residuals* $\{X_t - \Pi(X_t|X_A) : t \in B \setminus A\}$, where $\Pi(W|Z)$ is the best linear predictor of W given Z. We now have:

Theorem 1.5 (Unique Completion from a Serrated Domain)**.** *Let K_Ω be a partial covariance kernel on a serrated domain $\Omega = \cup_{p=1}^m I_p \times I_p \subset I$ corresponding to m intervals $\{I_p\}_{p=1}^m$ covering I. The following two statements are equivalent:*

(I) K_Ω admits a unique completion on I, i.e. $\mathcal{C}(K_\Omega)$ is a singleton.

(II) there exists an $r \in \{1, \ldots, m\}$*, such that*

$$K_{I_p}/K_{I_p \cap I_{p+1}} = 0, \text{ for } 1 \leq p < r \quad \text{and} \quad K_{I_{q+1}}/K_{I_q \cap I_{q+1}} = 0, \text{ for } r \leq q < m.$$

Condition (II) is strictly weaker than any of the sufficient conditions that have previously been stated in the literature on functional fragments, such as Theorem 1 in Delaigle et al. (2021) and Proposition 2 in Descary and Panaretos (2019b). Consequently, none of those conditions is necessary in the context of a serrated domain (for a discussion of more general domains, see Section 1.8). Furthermore, an appealing feature of (II) is that it is checkable at the level of K_Ω in a concrete manner by constructing a finite number of Schur complements (in fact the number is linear in m).

Theorem 1.5 elucidates just how restrictive it is to a priori assume that a unique completion exists. When the Schur complements involved in (II) vanish, one can start with the associated process $\{X_t : t \in I_r\}$ restricted to I_r, and iteratively *perfectly* predict each segment $\{X_t : t \in I_j\}$ by means of best linear prediction. Consequently, the entire process $\{X_t : t \in I\}$ is generated as the image of its restriction $\{X_t : t \in I_r\}$ via a *deterministic linear operator*. Indeed which interval(s) $\{I_j\}_{j=1}^m$ generate(s) the process can be discovered by checking the equations given in (II).

Note, however, that being able to identify K from $K|_\Omega$ does not require assuming that $K|_\Omega$ completes uniquely – all we need is a way to select one element from $\mathcal{C}(K|_\Omega)$. For example, to obtain identifiability, it would be much less restrictive to assume the admittance of a (second order) graphical model with respect to (I, Ω). The set of covariances $\mathcal{G}_\Omega$ corresponding to such processes is potentially very large, and encompasses highly "non-deterministic" dependence structures. Assuming that $K \in \mathcal{G}_\Omega$ will then yield identifiability given $K|_\Omega$ via Theorem 1.3, which can be re-interpreted in this notation as stating

$$\mathcal{C}(K|_\Omega) \cap \mathcal{G}_\Omega = \{K_\star\}$$

Since a unique completion is automatically canonical, it must also lie in $\mathcal{G}_\Omega$. Therefore, the assumption $K \in \mathcal{G}_\Omega$ is *strictly weaker* than the uniqueness assumption, while still guaranteeing identifiability. As noted earlier, in the last paragraph of Section 1.4, one can easily define a "second-order graphical model" structure with conditional uncorrelatedness replacing conditional independence, so imposing the assumption $K \in \mathcal{G}_\Omega$ in no way entails assuming Gaussianity. The family $\mathcal{G}_\Omega$ can also be thought of as a covariance selection model of the kind first proposed by Dempster (1972) for multivariate normal distributions, so that imposing the condition $K \in \mathcal{G}_\Omega$ amounts to doing continuous-domain parameter reduction.

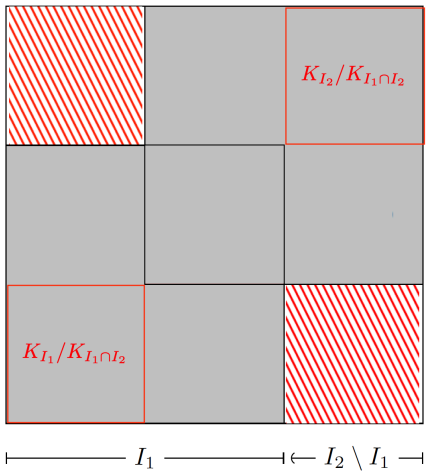


Figure 1.4: Illustration of Theorem 1.6. The 2-serrated domain Ω is shaded in grey, and the central square is $(I_1 \cap I_2)^2$. The set $\mathcal{C}(K_\Omega)$ is spanned as $K_\star + C$, where C ranges over cross-covariances supported on the union of the two squares shaded in red, and compatible with the covariances $K_{I_1} \setminus K_{I_1\cap I_2}$ and $K_{I_2} \setminus K_{I_1\cap I_2}$ (outlined in red).

1.6 Characterisation of All Completions

We will now show how the elements of $\mathcal{C}(K_\Omega)$ can be spanned by suitable perturbations of the canonical completion, when Ω is serrated. Again, it is instructive to commence with the 2-serrated case (see the left plot in Figure 1.2, p. 12).

Theorem 1.6 (Characterisation of Completions in the 2-Serrated Case)**.** *Let $\Omega = (I_1 \times I_1) \cup (I_2 \times I_2)$ be a 2-serrated subdomain in $I \times I$. The function $K : I \times I \to \mathbb{R}$ is a completion of $K_\Omega : \Omega \to \mathbb{R}$ if and only if*

$$K = K_\star + C$$

where $C : (I_1 \cup I_2)^2 \to \mathbb{R}$ satisfies $C(s,t) = 0$ for $(s,t) \in I_1^2 \cup I_2^2$ and is otherwise such that the function $L : [(I_1 \setminus I_2) \cup (I_2 \setminus I_1)]^2 \to \mathbb{R}$ given by

$$\begin{aligned} L|_{(I_1\setminus I_2)^2} &= K_{I_1}/K_{I_1\cap I_2}, \quad L|_{(I_1\setminus I_2)\times(I_2\setminus I_1)} = C|_{(I_1\setminus I_2)\times(I_2\setminus I_1)}, \\ L|_{(I_2\setminus I_1)^2} &= K_{I_2}/K_{I_1\cap I_2}, \quad L|_{(I_2\setminus I_1)\times(I_1\setminus I_2)} = C|_{(I_2\setminus I_1)\times(I_1\setminus I_2)} \end{aligned}$$

is a covariance.

Said differently, in the 2-serrated case $\Omega = (I_1 \times I_1) \cup (I_2 \times I_2)$ one has

$$K \in \mathcal{C}(K_\Omega) \iff K = K_\star + C$$

where $K_\star$ is the canonical completion and C is a valid perturbation. The set of all valid perturbations C is given by the cross-covariances $C|_{(I_1\setminus I_2)\times(I_2\setminus I_1)}$ (with $C|_{(I_2\setminus I_1)\times(I_1\setminus I_2)}$

determined by symmetry, i.e. $C(s,t) = C(t,s)$) corresponding to all possible couplings (Y_t, W_t) of the Gaussian processes

$$\begin{aligned} \{Y_t : t \in I_1 \setminus I_2\}, \quad Y &\sim N(0, K_{I_1}/K_{I_1 \cap I_2}), \\ \{W_t : t \in I_2 \setminus I_1\}, \quad W &\sim N(0, K_{I_2}/K_{I_1 \cap I_2}), \end{aligned}$$

over the indicated region, and are zero elsewhere.

Selecting valid perturbations C is straightforward: it basically amounts to the functional analogue of "assigning a correlation to two variances". At the same time, notice that any non-zero perturbation C introduces arbitrary correlations that were never observed (i.e. are entirely extrinsic to the partial covariance K_Ω). This observation crystallises some of the remarks made in the closing of Section 1.4, i.e. that the canonical completion is unique in not introducing any arbitrary correlations extrinsic to K_Ω.

We will now re-interpret the last result through the lens of operator theory – this perspective will allow a fruitful extension of our characterisation to general serrated domains. First, we note that if K_Ω is continuous on Ω, then so are all elements of $\mathcal{C}(K_\Omega)$ on I and L is also continuous on $[(I_1 \setminus I_2) \cup (I_2 \setminus I_1)]^2$ (for the latter, see Remark A.3 in Section A.2 in the Supplementary Material). As a result, we can think of L as the kernel of a covariance operator. Let

$$\mathbf{L}_1 : L^2(I_1 \setminus I_2) \to L^2(I_1 \setminus I_2) \quad \text{and} \quad \mathbf{L}_2 : L^2(I_2 \setminus I_1) \to L^2(I_2 \setminus I_1)$$

denote the integral operators induced by the covariance kernels

$$L|_{(I_1 \setminus I_2) \times (I_1 \setminus I_2)} = K_{I_1}/K_{I_1 \cap I_2} \quad \text{and} \quad L|_{(I_2 \setminus I_1) \times (I_2 \setminus I_1)} = K_{I_2}/K_{I_1 \cap I_2}.$$

Moreover, let

$$\mathbf{L}_{12} : L^2(I_2 \setminus I_1) \to L^2(I_1 \setminus I_2)$$

denote the integral operator corresponding to the kernel

$$L|_{(I_2 \setminus I_1) \times (I_1 \setminus I_2)}.$$

Finally, define

$$\mathbf{L} : L^2(I_1 \setminus I_2) \times L^2(I_2 \setminus I_1) \to L^2(I_1 \setminus I_2) \times L^2(I_2 \setminus I_1)$$

to be a linear operator defined via its action:

$$\mathbf{L}(f,g) = (\mathbf{L}_1 f + \mathbf{L}_{12} g, \mathbf{L}_{12}^* f + \mathbf{L}_2 g).$$

Clearly, L is a completion if and only if $\mathbf{L}$ is positive semidefinite. Notice that $\mathbf{L}_1$ and $\mathbf{L}_2$ are trace-class and positive semidefinite, and as a result L is trace-class if it is

positive-semidefinite. Now, $\mathbf{L}$ is positive semidefinite if and only if there is Gaussian measure μ_{12} on the Hilbert space $L^2(I_1 \setminus I_2) \times L^2(I_2 \setminus I_1)$ with zero mean and covariance operator $\mathbf{L}$ which has two Gaussian measures μ_1 and μ_2 with zero mean and covariance operators $\mathbf{L}_1$ and $\mathbf{L}_2$ as marginals. According to Theorem 2 of Baker (1973), the possible values of $\mathbf{L}$ are precisely the ones given when setting

$$\mathbf{L}_{12} = \mathbf{L}_1^{1/2} \Psi \mathbf{L}_2^{1/2}$$

for $\Psi : L^2(I_1 \setminus I_2) \to L^2(I_2 \setminus I_1)$ a bounded linear map with operator norm $\|\Psi\|_\infty \leq 1$. In summary, if K_Ω is continuous, Theorem 1.6 can be re-interpreted at the level of operators. Namely, in block notation, the operator $\mathbf{K}$ has a kernel in $\mathcal{C}(K_\Omega)$ if and only if

$$\mathbf{K}f = \mathbf{K}_\star f + \underbrace{\left(\begin{array}{c|c|c} 0 & 0 & (\mathbf{L}_1^{1/2}\Psi\mathbf{L}_2^{1/2})^* \\ \hline 0 & 0 & 0 \\ \hline \mathbf{L}_1^{1/2}\Psi\mathbf{L}_2^{1/2} & 0 & 0 \end{array}\right)}_{\mathbf{C}} \left(\begin{array}{c} f|_{I_1 \setminus I_2} \\ f|_{I_1 \cap I_2} \\ f|_{I_2 \setminus I_1} \end{array}\right). \tag{1.7}$$

Here $\mathbf{K}_\star$ is the operator with the canonical completion $K_\star$ as its kernel, and as Ψ ranges over the ball $\|\Psi\|_\infty \leq 1$, the expression above generates all possible operator completions. Choosing $\Psi = 0$ obviously yields the canonical completion. Note that the operator $\mathbf{C}$ in Equation (1.7) is precisely the operator corresponding to the (cross-covariance) integral kernel C as described earlier.

We will use this operator perspective to obtain a characterisation in the general case, where $\Omega = \cup_{j=1}^m (I_j \times I_j)$ is an m-serrated domain. This will require some additional notation to avoid excessively cumbersome expressions. For $1 \leq p < m$ define the following sets:

$$J_p = I_p \cap I_{p+1}, \quad D_p = I_{p+1} \setminus I_p, \quad S_p = \left[\cup_{j=1}^p I_j\right] \setminus I_{p+1}, \quad R_p = D_p \times S_p, \quad R'_p = S_p \times D_p.$$

See Figure 1.5 for a visual interpretation.

Let K be a covariance on I with associated operator $\mathbf{K}$. For every $1 \leq j \leq m$, let $\mathbf{K}_j : L^2(I_j) \to L^2(I_j)$ be the Hilbert-Schmidt operator induced by the integral kernel $K_{I_j} = K|_{I_j \times I_j}$. And for $1 \leq p < m$, let $\mathbf{J}_p \in \mathcal{S}_2(J_p, J_p)$ and $\mathbf{R}_p \in \mathcal{S}_2(D_p, S_p)$ be Hilbert-Schmidt operators induced by the integral kernels $K_{J_p} = K|_{J_p \times J_p}$ and $K_{R_p} = K|_{S_p \times D_p}$ respectively.

We will now show that $\mathbf{K}$ can always be written in a sort of "block notation", i.e. in terms of $\{\mathbf{K}_j\}_j$, $\{\mathbf{J}_p\}_p$ and $\{\mathbf{R}_p\}_p$. This will allow us to generalise the type of expression Equation (1.7) to the m-serrated case.

Lemma 1.1. *Given any $f \in L^2(I)$ and continuous kernel $K : I \times I \to \mathbb{R}$ with associated*

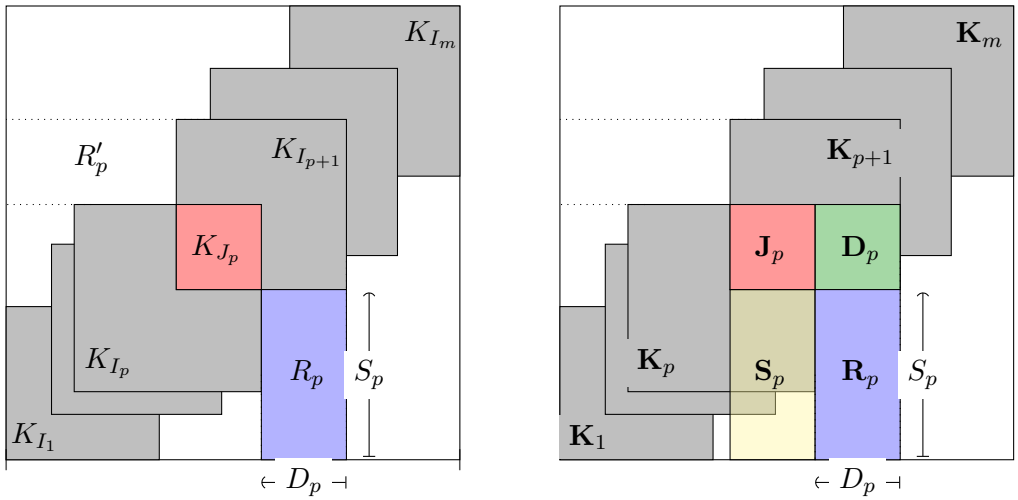


Figure 1.5: Illustration of the sets D_p, S_p, R_p and R'_p (left) and the corresponding operators (right).

operator $\mathbf{K}$*, the mapping* $f \mapsto \mathbf{K}f$ *can be represented blockwise as*

$$\mathbf{K}f(t) = \sum_{j:t\in I_j} \mathbf{K}_j f|_{I_j}(t) + \sum_{p:t\in S_p} \mathbf{R}_p f|_{D_p}(t) + \sum_{p:t\in D_p} \mathbf{R}_p^* f|_{S_p}(t) - \sum_{p:t\in J_p} \mathbf{J}_p f|_{J_p}(t) \text{ a.e.}$$

Consequently, in order to characterize any integral operator corresponding to a completion of K_Ω, it suffices to characterize the operators $\{\mathbf{R}_p\}_p$. These are the only "missing pieces", as the rest is known from K_Ω (see Figure 1.5).

To this end, for $1 \leq p < m$, we define $\mathbf{D}_p \in \mathcal{S}_2(J_p, D_p)$ and $\mathbf{S}_p \in \mathcal{S}_2(S_p, J_p)$ to be the Hilbert-Schmidt operators induced by the integral kernels $K_\Omega|_{D_p\times J_p}$ and $K_\star|_{J_p\times S_p}$ respectively (with $K_\star$ the canonical completion, as always).

Now we have all the ingredients to characterise all completions from a serrated domain:

Theorem 1.7 (Characterisation of Completions from a General Serrated Domain)**.** *Let* K_Ω *be a continuous partial covariance on a serrated domain* Ω *of* m *intervals. Then* $K : I \times I \to \mathbb{R}$ *with* $K|_\Omega = K_\Omega$ *is a completion of* K_Ω *if and only if the integral operator* $\mathbf{K}$ *corresponding to* K *is of the form*

$$\mathbf{K}f(t) = \sum_{j:t\in I_j} \mathbf{K}_j f_{I_j}(t) + \sum_{p:t\in S_p} \mathbf{R}_p f_{D_p}(t) + \sum_{p:t\in D_p} \mathbf{R}_p^* f_{S_p}(t) - \sum_{p:t\in J_p} \mathbf{J}_p f_{J_p}(t) \text{ a.e.}, \quad (1.8)$$

where for $1 \leq p < m$*,*

$$\mathbf{R}_p = \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p\right] + \mathbf{U}_p^{1/2}\Psi_p\mathbf{V}_p^{1/2}, \quad (1.9)$$

with

$$\mathbf{U}_p = \mathbf{K}_{S_p} - \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right], \qquad \mathbf{V}_p = \mathbf{K}_{D_p} - \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p^*\right]$$

and $\Psi_p : L^2(D_p) \to L^2(S_p)$ *is a bounded linear map with* $\|\Psi_p\| \leq 1$.

The only degrees of freedom in Equation (1.9) stem from the m contractions $\{\Psi\}_{p=1}^m$. All other operators involved in Equation (1.9) (and in the right hand side of Equation (1.8)) are uniquely defined via K_Ω (or equivalently via $K_\star$). Allowing these to range over the unit balls

$$\|\Psi_p\|_\infty \leq 1, \quad \Psi_p : L^2(D_p) \to L^2(S_p), \qquad p = 1, \ldots, m$$

we trace out the set $\mathcal{C}(K_\Omega)$ and get an idea of what the different possibilities of the actual covariance may look like. Substituting $\Psi_p = 0$ for all $1 \leq p < m$ returns the integral operator corresponding to the canonical completion $K_\star$ of K_Ω. Since all other elements of Equation (1.9) are fully determined by K_Ω, it is clear that the choice of $\{\Psi_p\}$ is arbitrary, and any non-zero choice will introduce information extrinsic to observed correlation patterns – extending the intuition build in the 2-serrated case relating to the canonicity of $K_\star$.

Theorem 1.7 also complements Theorem 1.2, in that it expresses the canonical completion as the solution of a system of equations rather than the output of an algorithm. This manner of specification is slightly weaker, in that it assumes continuity of K_Ω, whereas Theorem 1.2 makes no such assumption. On the other hand, it provides a characterisation of the canonical solution in a form that lends itself for the problem of *estimation*, treated in the next section.

1.7 Estimation of the Canonical Completion

In this section, we consider the problem of estimation of the canonical completion $K_\star$ when we only have access to an estimator of the partial covariance K_Ω. From a purely analytical sense, we are studying the *stability* of the canonical completion $K_\star$ of K_Ω with respect to perturbations of the partial covariance K_Ω. From the statistical point of view, this relates to the problem that arises in the context of covariance recovery from functional fragments: when we observe fragments $X_j|_{I_j}$ of i.i.d. realisations of a second-order process $\{X_t : t \in I\}$ for a collection $\{I_j\}$ of subintervals $I_j \subset I$ covering I. Because of the fragmented nature of the observations, we only have covariance information on the serrated domain $\Omega = \cup_j I_j \times I_j$, or equivalently we only can identify the partial covariance K_Ω corresponding to the restriction of $\mathrm{Cov}\{X_s, X_t\} = K(s,t)$ to the serrated domain Ω.

In this context, we posit that it makes good sense to choose the canonical completion $K_\star$ as the target of estimation. This is because the canonical completion is always an

identifiable and *interpretable* object:

1. When a unique completion exists, it must be the canonical one. So choosing the canonical completion allows us to *adapt* to uniqueness.

2. When multiple completions exist, the canonical completion remains identifiable, and is least presumptuous – it relies solely on the available data.

Targeting the canonical completion without any attempt to enforce uniqueness is qualitatively very different from previous approaches to covariance recovery from fragmented paths. Those approaches imposed uniqueness by way of assumptions (indeed assumptions implying very rigid consequences, as demonstrated in Section 1.5). Once uniqueness is a priori guaranteed, any estimator $\hat{K}$ whose restriction $\hat{K}|_\Omega$ is consistent for K_Ω will be valid – so, for instance, one can safely extrapolate an estimator $\hat{K}_\Omega$ of K_Ω by means of a basis expansion or matrix completion. But when uniqueness fails to hold, such "extrapolation" estimators yield arbitrary completions, indeed completions that likely will not even converge asymptotically, but rather oscillate in some open neighbourhood of $\mathcal{C}(K_\Omega)$. On the other hand, the adaptivity (to uniqueness) and stability/interpretability (under non-uniqueness) of the canonical completion comes at a price: to be able to target the canonical completion $K_\star$ we need an estimator that is not merely consistent for K_Ω on Ω, but one that (asymptotically) also satisfies the system of equations in Theorem 1.7 (with Ψ_p identically zero). This has consequences on the rates of convergence, which can no longer be as fast as the rates of estimating the partial covariance K_Ω.

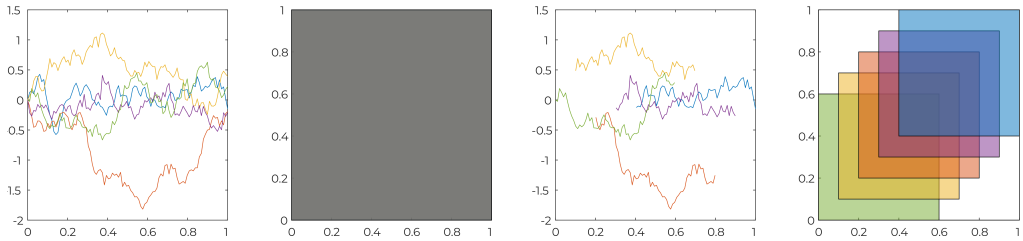


Figure 1.6: Illustration of the problem of covariance recovery from fragments: (from the left) fully observed sample paths of a process on the unit interval $I = [0, 1]$; the region I^2 on which the covariance can be estimated in the fully observed case; partially observed versions of the same sample paths; and, the region on which the covariance can be estimated from the sample paths of the corresponding colour.

1.7.1 Definition of the Estimator

Courtesy of Theorem 1.7, the specification of $K_\star$ reduces to that the solution of the following system of linear equations:

$$\begin{aligned} \mathbf{J}_p^{1/2}\mathbf{X}_p &= \mathbf{S}_p^*, \\ \mathbf{J}_p^{1/2}\mathbf{Y}_p &= \mathbf{D}_p, \end{aligned} \qquad \text{for } 1 \le p < m. \tag{1.10}$$

Notice that the operator $\mathbf{J}_p^{1/2}$ is compact because $\mathbf{J}_p$ is. It follows that the canonical completion $K_\star$ does not depend continuously on the partial covariance K_Ω. In practice we only have access to an estimator $\hat{K}_\Omega$ of K_Ω. Therefore, the *operator* of the inverse problem, i.e. $\mathbf{J}_p^{1/2}$, as well as the *data* of the inverse problem, in the form of $\mathbf{D}_p$ and $\mathbf{S}_p$, are inexactly specified.

We will thus define our estimator as the solution of a regularized empirical version of the system. Let $\hat{K}_\Omega$ be an estimator of K_Ω. Let $\hat{\mathbf{K}}_p$, $\hat{\mathbf{D}}_p$ and $\hat{\mathbf{J}}_p$ be the Hilbert-Schmidt operators with the kernels $\hat{K}_\Omega|_{I_p \times I_p}$, $\hat{K}_\Omega|_{J_p \times D_p}$ and $\hat{K}_\Omega|_{J_p \times J_p}$, respectively.

Finally, motivated by the definition

$$\mathbf{R}_p = \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p\right]$$

and using a truncated inverse of $\hat{\mathbf{J}}_p$, we define the regularised empirical version of $\mathbf{R}_p$ as

$$\hat{\mathbf{R}}_p = \sum_{k=1}^{N_p} \frac{1}{\hat{\lambda}_{p,k}} \cdot \hat{\mathbf{S}}_p \hat{e}_{p,k} \otimes \hat{\mathbf{D}}_p^* \hat{e}_{p,k} \tag{1.11}$$

where $\hat{\lambda}_{p,k}$ and $\hat{e}_{p,k}$ denote the kth eigenvalue and eigenfunction of $\hat{\mathbf{J}}_p$, N_p is the truncation or regularization parameter, and $\hat{\mathbf{S}}_p$ has kernel $\hat{K}_\star|_{S_p \times J_p}$. Notice that the definition is recursive:

- $\hat{\mathbf{R}}_p$ depends on $\hat{\mathbf{S}}_p$ and thus on $\hat{\mathbf{R}}_i$ for $i < p$.
- $\hat{\mathbf{S}}_1$ is fully determined by $\hat{K}_\Omega$, and $\hat{\mathbf{S}}_{p+1}$ is fully determined by $\hat{K}_\Omega$ and $\hat{\mathbf{R}}_p$.

In particular, though the kernel of $\hat{\mathbf{S}}_p$ can *a posteriori* be seen to equal $\hat{K}_\star|_{S_p \times J_p}$, this does not mean that it depends *a priori* on $\hat{K}_\star$ (i.e., there is no vicious circle in the definition).

We can now define our estimator $\hat{K}_\star : I \times I \to \mathbb{R}$ of $K_\star$ to be the integral kernel of the

Hilbert-Schmidt operator $\hat{\mathbf{K}}_\star : L^2(I) \to L^2(I)$ defined via the action

$$\hat{\mathbf{K}}_\star f(t) = \sum_{j:t\in I_j} \hat{\mathbf{K}}_j f|_{I_j}(t) + \sum_{p:t\in S_p} \hat{\mathbf{R}}_p f|_{D_p}(t) + \sum_{p:t\in D_p} \hat{\mathbf{R}}_p^* f|_{S_p}(t) - \sum_{p:t\in J_p} \hat{\mathbf{J}}_p f|_{J_p}(t). \quad (1.12)$$

Equivalently, we can define $\hat{K}_\star \in L^2(I\times I)$ recursively as follows: $\hat{K}_\star|_\Omega = \hat{K}_\Omega$ and $\hat{K}_\star|_{R_p}$ is the kernel associated with the Hilbert-Schmidt operator $\hat{\mathbf{R}}_p$ defined recursively via (1.11).

1.7.2 Rate of Convergence

We will now characterize the rate of convergence of $\hat{K}_\star$ to $K_\star$ in terms of the spectral properties of the partial covariance K_Ω, and the rate of convergence of the partial covariance estimator $\hat{K}_\Omega$ we have used as a basis, to the partial covariance K_Ω itself. Concerning the spectral properties of K_Ω, let $\{\lambda_{p,k}\}_{k=1}^\infty$ be the eigenvalues of $\mathbf{J}_p$ and define $A_{p,k}$ as:

$$A_{p,k} = \left\| \sum_{j=k+1}^{\infty} \frac{\mathbf{S}_p e_{p,k} \otimes \mathbf{D}_p^* e_{p,k}}{\lambda_{p,k}} \right\|_2^2$$

where $\{e_{p,k}\}_{k=1}^\infty$ are the eigenfunctions of $\mathbf{J}_p$. Notice that $A_{p,0}$ is simply the Hilbert-Schmidt norm of $\mathbf{R}_p$ and $A_{p,k}$ represents the error in approximating $\mathbf{R}_p$ by using a rank-k truncated inverse of $\mathbf{J}_p^{1/2}$ instead of $\mathbf{J}_p^{1/2}$ in the expression

$$\mathbf{R}_p = \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p\right].$$

Consequently, $A_{p,k}$ must necessarily converge to 0 as $k\to\infty$. The following result gives the rate of convergence for the case when the eigenvalues and approximation errors decay at a polynomial rate.

Theorem 1.8 (Consistency and Rate of Convergence)**.** *Let K_Ω be a partial covariance on a serrated domain Ω of m intervals and $\hat{K}_\Omega$ be an estimator thereof. Let $\hat{K}$ be defined as in Equation* (1.12)*. Assume that for every $1 \le p < m$, we have $\lambda_{p,k} \sim k^{-\alpha}$ and $A_{p,k} \sim k^{-\beta}$. If the error in the estimation of K_Ω satisfies*

$$\|\hat{K}_\Omega - K_\Omega\|_{L^2(\Omega)} = O_{\mathbb{P}}(1/n^\zeta)$$

where n is the number of fragments, then the error in the estimation of the canonical completion satisfies, for every $\varepsilon > 0$,

$$\|\hat{K}_\star - K_\star\|_{L^2(I\times I)} = O_{\mathbb{P}}(1/n^{\zeta\gamma_{m-1}-\varepsilon})$$

provided the truncation parameters $\mathbf{N} = (N_p)_{p=1}^{m-1}$ *scale according to the rule*

$$N_p \sim n^{\gamma_p/\beta}$$

where $\gamma_{m-1} = \frac{\beta}{\beta+2\alpha+3/2}\left[\frac{\beta}{\beta+\alpha+1/2}\right]^{m-2}$.

Remark 1.3 (Plug-in Interpretation)**.** *The theorem can be seen as a plug-in rate of convergence theorem. We plug-in the "baseline" rate of convergence of* $\hat{K}_\Omega \to K_\Omega$*, and get a rate for* $\hat{K}_\star \to K_\star$*. Note that the tuning of the truncation parameters also depends on the baseline rate of convergence. Baseline rates are readily available for sparse, dense, and complete observation regimes.*

Notice that the rate of convergence γ_{m-1} strictly decreases as a function of the number of intervals m, but can get arbitrarily close to 1 for a large enough rate of decay of approximation errors β. Moreover, an increase in the rate of decay of eigenvalues α is accompanied by a decrease in the rate of convergence. If K_Ω is r-times differentiable then the same applies to the kernels $K_\Omega|_{J_p\times J_p}$ of $\mathbf{J}_p$ implying $\lambda_{p,k}$ is $o(1/n^{r+1})$ for every $1 \le p < m$ and thus $\alpha = r + 1$. Thus, all other things being equal, an increase in the smoothness of K_Ω also tends to decrease the rate of convergence —which is not surprising from an inverse problems perspective.

1.8 Beyond Serrated Domains

Our theory has thus far concentrated on domains that are *serrated* in the sense of Definition 1.2. We now turn our attention to a much larger class of domains, namely domains that can be approximated to an arbitrary level of precision by serrated domains. Recall that for subsets X and Y of a metric space (M, d), the Hausdorff distance is defined as

$$d_H(X,Y) = \Big[\sup_{x\in X}\inf_{y\in Y} d(x,y)\Big] \vee \Big[\sup_{y\in Y}\inf_{x\in X} d(x,y)\Big].$$

We define the class of *nearly serrated domains* as the Hausdorff "closure" of the set of serrated subdomains of $I \times I$:

Definition 1.4 (Nearly Serrated Domain)**.** *We say that* $\widetilde{\Omega} \subset I \times I$ *is a nearly serrated domain if for every* $\epsilon > 0$*, there exist serrated domains* Ω_ϵ *and* Ω^ϵ *such that* $\Omega_\epsilon \subset \Omega \subset \Omega^\epsilon$ *and* $d_H(\Omega, \Omega_\epsilon), d_H(\Omega, \Omega^\epsilon) < \epsilon$*, where* d_H *is the Hausdorff metric induced by the Euclidean metric on* $I \times I \subset \mathbb{R}^2$*.*

Notice that every serrated domain is nearly serrated according to the above definition. Of particular importance is the case when Ω is a strip of width $w > 0$ around the diagonal, that is, $\Omega = \{(s,t) \in I \times I : |s - t| \le w/\sqrt{2}\}$. This occurs asymptotically in the problem of covariance recovery from fragments, when each sample path is observable which over

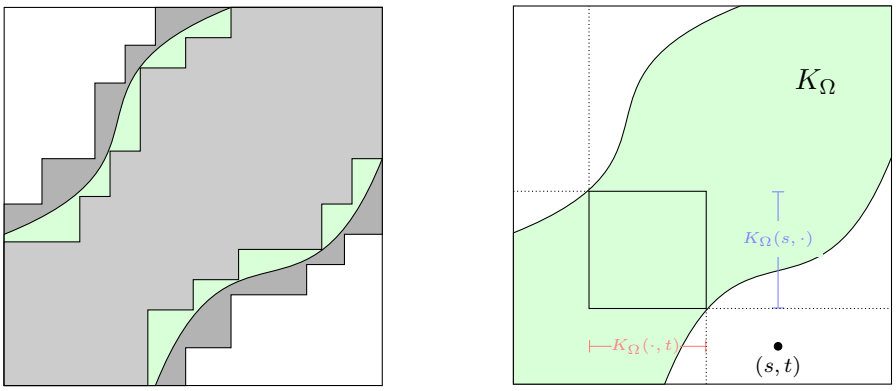


Figure 1.7: Left: Illustration of a nearly serrated domain (in green) as enveloped by two serrated domains (in light and dark grey). Right: a point (s,t) escaping the scope of Equation (1.2).

"uniformly distributed" intervals of constant length $w/\sqrt{2}$.

It should be clear from Figure 1.7 that we cannot exploit Equation 1.2 to recover the canonical completion of a partial covariance on a nearly serrated but not serrated domain, as we did for serrated domains previously. This is because for such domains there are points (s,t) for which the cross-covariances $K_\Omega(s,\cdot)$ and $K_\Omega(\cdot,t)$ are not available, nor can they be iteratively calculated from the part of the covariance that is known. Thus one cannot evaluate their inner product $\left\langle K_\Omega(s,\cdot), K_\Omega(\cdot,t)\right\rangle_{\mathcal{H}(K_{I_1\cap I_2})}$ as required in Equation 1.2. It was precisely because the domain was serrated that we were able to recover the value of the canonical completion over successively larger regions as we did in Algorithm (1).

Additionally, since we cannot apply Algorithm 1, it is unclear what it means for a completion of a partial covariance K_Ω on a nearly serrated domain to be canonical. Here we must lean on our graphical models interpretation to define the canonical completion. We shall say that a covariance K is a canonical completion of K_Ω if it is a completion i.e. $K|_\Omega = K_\Omega$ and $K \in \mathcal{G}_\Omega$ as defined in Section 1.4.

Our focus will, therefore, be to obtain results pertaining to uniqueness/canonicity of completions from nearly serrated domains $\widetilde{\Omega}$ by means of serrated subdomains $\Omega \subset \widetilde{\Omega}$ or superdomains $\widetilde{\Omega} \subset \Omega$. Our first result gives a sufficient condition for unique completion from a nearly serrated domain:

Theorem 1.9 (Checking Uniqueness via Serrated Subdomains)**.** *Let $K_{\widetilde{\Omega}}$ be a partial covariance on a nearly serrated domain $\widetilde{\Omega}$ and let $\Omega \subset \widetilde{\Omega}$ be a serrated domain. If the restriction $K_{\widetilde{\Omega}}\big|_\Omega$ admits a unique completion, then so does $K_{\widetilde{\Omega}}$.*

Theorem 1.9, via our necessary and sufficient conditions for uniqueness on serrated domains (Theorem 1.5), yields sufficient conditions for unique completion from a banded domain that are strictly weaker than any previously known set of sufficient conditions.

In the serrated case, a unique completion is necessarily canonical. A natural question is whether this remains the case for nearly serrated domains. The answer is in the affirmative:

Theorem 1.10 (Unique Completions are Canonical)**.** *If the partial covariance $K_{\widetilde{\Omega}}$ on a nearly serrated domain $\widetilde{\Omega}$ has a unique completion on $I \times I$, this completion is canonical.*

Theorem 1.10 shows that targeting canonical completions remains a sensible strategy in the context of nearly serrated domains – they remain interpretable and yield the "correct answer" in the presence of uniqueness. That is, of course, if we know how to construct them. Our last result pertains to this matter:

Theorem 1.11 (Construction of Canonical Completions)**.** *A covariance $K_\star$ on I can be recovered as the canonical completion of its restriction $K_\star|_\Omega$ on a serrated domain Ω if and only if it is the canonical completion of a partial covariance on some nearly serrated domain $\widetilde{\Omega} \subset \Omega$.*

In particular, if a unique completion of $K|_{\widetilde{\Omega}}$ exists then it equals the canonical completion of $K|_\Omega$ for a (in fact any) serrated $\Omega \supset \widetilde{\Omega}$. Alternatively, if the process X with covariance K forms a second-order graphical model with respect to the nearly serrated $\widetilde{\Omega}$, then K can be obtained by means of Algorithm 1 applied to $K|_{\Omega'}$, for any $\Omega' \supset \widetilde{\Omega}$. This is possible because if $\widetilde{\Omega} \subset \Omega$, then every separator of Ω also separates $\widetilde{\Omega}$. As a result, if the "separation equation" is satisfied by $(s,t) \in (\widetilde{\Omega})^c$ for separators of $\widetilde{\Omega}$, then it is also satisfied for separators of Ω. Thus $\widetilde{\Omega} \subset \Omega$ implies $\mathcal{G}_{\widetilde{\Omega}} \subset \mathcal{G}_\Omega$.

1.9 Covariance Estimation from Sample Path Fragments

We now elucidate how one can make combined use of our results from Section 1.7 and Section 1.8, in order to address the problem of covariance estimation from sample path fragments in a general context. Let X be a second-order process on the unit interval I with the covariance K. Suppose that for n intervals $I_i \subset I$ we observe n sample path fragments $X_i = X_i|_{I_i}$, where $X_i \sim X$ independently. Now define the domain

$$\Omega_\infty = \limsup_{k \to \infty} (I_k \times I_k),$$

as the set of pairs $(x,y) \in I \times I$ such that $I_i \ni \{x,y\}$ infinitely often. The sequence $\{X_i\}_{i=1}^n$ enables us to consistently estimate the restriction $K_\Omega := K|_\Omega$ of K on any $\Omega \subset \Omega_\infty$. Call such a consistent estimator $\hat{K}_\Omega$.

However, we wish to estimate the complete K, not merely its restrictions to $\Omega \subset \Omega_\infty$. This requires K to be identifiable from Ω_∞. One means to securing identifiability is to impose unique extendability, i.e. assume that $\mathcal{C}(K_\Omega) = \{K\}$ for some Ω whose elements are covered infinitely often by the sequence of rectangles $I_k \times I_k$ (i.e. we have the inclusion[2] $\Omega \subset \Omega_\infty$). But uniqueness was seen to be overly restrictive (see Theorem 1.5). We therefore wish to avoid this route to identifiability. Instead, following the development in Section 1.4 we will secure identifiability by assuming that the underlying process X is globally Markov with respect to some domain Ω (i.e. $K \in \mathcal{G}_\Omega$) whose elements are covered infinitely often by the sequence the rectangles $I_k \times I_k$ (i.e. $\Omega \subset \Omega_\infty$). This is a substantially weaker assumption (due to Theorem 1.10), and arguably much more intuitive.

Proceeding thus, let $\Omega \subset \Omega_\infty$ be some nearly serrated domain with respect to which X is global Markov. By Theorem 1.4, the true covariance K is the canonical completion from Ω, and by Theorem 1.11, it is also the canonical completion from any serrated domain containing Ω. Thus we can identify K directly from K_Ω as the canonical completion of K_{Ω_m} for some m-serrated Ω_m (with some $m < \infty$) satisfying $\Omega \subseteq \Omega_m \subseteq \Omega_\infty$. The inclusions will always be possible for some $m < \infty$ provided the boundaries $\partial\Omega$ and $\partial\Omega_\infty$ are everywhere distinct (i.e. $\|u - v\| > 0$ for all $u \in \partial\Omega$ and $v \in \partial\Omega_\infty$).

Now we distinguish two cases:

(i) Ω_∞ **is serrated.** If intervals $\{I_j\}_{j=1}^n$ are sampled from a finite cover of I, then Ω_∞ will be a serrated domain. This represents a fixed domain setting, in that for all sufficiently large n the observation domain becomes almost surely fixed.

(ii) Ω_∞ **is nearly serrated.** If the intervals $\{I_j\}_{j=1}^n$ are sampled from an infinite cover of I, then Ω_∞ will be a nearly serrated domain. This represents a variable domain setting, as our observation domain will continue evolving as n grows.

In case (i), we are squarely within the context of Section 1.7 and can use $\hat{K}_\Omega$ to directly define the estimator $K_\star$ used in Equation 1.12.

In case (ii), we choose and fix an m-serrated approximation of the observable region $\Omega_m \subset \Omega$. We then construct the estimator $\hat{K}_\star$ in Equation 1.12 based on $\hat{K}_{\Omega_m}$. So long as Ω_m contains Ω and m is held fixed, the estimator thus constructed will converge to K as per Theorem 1.8. A good choice of Ω_m involves a trade-off between covering a large subregion of Ω_∞ (to use as much of the observable domain as possible) and keeping m small (to limit the number of inverse problems solved). This approach is illustrated

[2]Note that in general the inclusion $\Omega_\infty \supset \Omega$ is to be taken as strict. The "critical" equality case would be rather exceptional, because Ω_∞ is defined by the censoring mechanism, whereas the Ω is a population quantity. Stipulating that the structure of X varies with or is tailored to the censoring mechanism would be contrived.

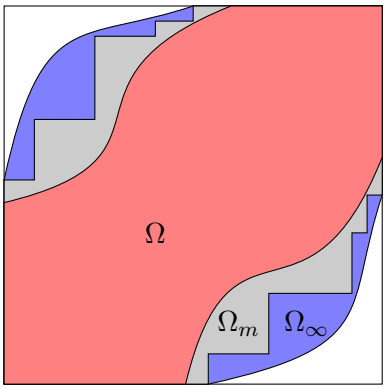


Figure 1.8: The regions Ω (red), Ω_m (gray), and Ω_∞ (blue).

through a data analysis in Section 1.11, and its performance is investigated via extensive simulations in Section 1.10 (with special focus on the effect of the choice of m).

Remark 1.4 (On m vs n – Practical Considerations)**.** *From a practical point of view, the choice of a serrated approximation Ω_m to Ω_∞ does not entail any significant loss. In practice, Ω_∞ is in fact unknown, and the de facto domain of observation is the n-serrated domain $\cup_{j=1}^n (I_j \times I_j)$. Nevertheless, statistical considerations suggest that we should not use the full observation domain, namely:*

1. *The domain $\cup_{j=1}^n (I_j \times I_j)$ generally "overfits" $\Omega = \limsup_j (I_j \times I_j)$. Regions of $I \times I$ that are more densely populated by observations are better proxies for Ω_∞ (meaning regions comprised of pairs $(x, y) \in I \times I$ such that $\#\{k \leq n : \{x, y\} \subset I_k\}$ is large). This suggests choosing Ω_m with m distinctly smaller than n.*

2. *When the fragments are observed discretely and smoothing is used to construct $\hat{K}_{\Omega_m}$, we still use data in the region $\cup_{j=1}^n (I_j \times I_j) \setminus \Omega_m$ as part of the local averaging, even though we desist from estimating outside Ω_m. Hence we do not necessarily discard information, but rather focus on a smaller region on which we can estimate more efficiently: because this region is more densely populated by observations and furthermore because we avoid boundary effects.*

Remark 1.5 (On m vs n – Asymptotic Considerations)**.** *As argued earlier, m need not grow with n for consistent estimation. We can nevertheless ask at what rate one might choose to let m grow with n, in the spirit of trading off more error due to a higher number of inverse problems to solve in exchange for more data. Theorem 1.8 can partially inform heuristics on this. Suppose that the error $\hat{K}_\Omega$ is $n^{-\alpha}$–consistent for K_Ω, $\alpha \in (0, 1)$ – which is certainly the case under complete observation. Take $\gamma_{m-1} = O(\eta^m)$ where $\eta \in (0, 1)$. Plugging these into Theorem 1.8 would suggest that*

$$\|\hat{K}_\star - K_\star\|_{L^2(I\times I)} \preceq n^{-\alpha\eta^m} = \exp\left[-\alpha\eta^m \log n\right] \to 1$$

This would indicate that m should not grow any faster than $O(\log\log n)$, effectively leading to m being practically constant: the increase in ill-posedness overwhelms any gain by adding more data.

1.10 Simulation Study

We consider three covariances:

$$K_1(s,t) = \sum_{j=1}^{4} \frac{\phi_j(s)\phi_j(t)}{2^{j-1}}, \qquad K_2(s,t) = s \wedge t, \qquad K_3(s,t) = 10ste^{-10|s-t|^2}$$

where $\phi_1(t) = 1$, $\phi_2(t) = \sqrt{3}(2t-1)$, $\phi_3(t) = \sqrt{5}(6t^2-6t+1)$ and $\phi_4(t) = \sqrt{7}(20t^3-30t^2+12t-1)$. The first covariance is both finite-rank and analytic, the second is infinite-rank and non-analytic and the third is infinite-rank and analytic. For the first and the third covariance, every restriction to a serrated domain admits a unique completion (due to analyticity), which is not the case for the second covariance (by Lemma A.2 in the Supplementary Material). Define the domains Ω_j as follows:

$$\begin{aligned}
\Omega_1 &= [0,3/5]^2 \cup [2/5,1]^2 \\
\Omega_2 &= \Omega_1 \cup [1/5,4/5]^2 \\
\Omega_3 &= \Omega_2 \cup [1/10,7/10]^2 \cup [3/10,9/10]^2 \\
\Omega_4 &= \Omega_3 \cup [1/20,13/20]^2 \cup [3/20,15/20]^2 \cup [5/20,17/20]^2 \cup [7/20,19/20]^2 \\
\Omega_5 &= \Omega_4 \cup [1/40,25/40]^2 \cup [3/40,27/40]^2 \cup [5/40,29/40]^2 \cup [7/40,31/40]^2 \\
&\qquad \cup [9/40,33/40]^2 \cup [11/40,35/40]^2 \cup [13/40,37/40]^2 \cup [15/40,39/40]^2.
\end{aligned}$$

The number of intervals m for the domains is 2, 3, 5, 9 and 17, respectively. The variable domain simulations of Delaigle et al. (2021) roughly correspond to an implicit choice of $m = 17$ (i.e. Ω_5).

The computations have been implemented in the `R` programming language (R Core Team, 2019) with the exception of those involving the estimator proposed in Delaigle et al. (2021) which was implemented in MATLAB. The implementation of our estimator in R can be found in the covcomp package (Waghmare, 2022).

1.10.1 General Simulation Study

For the covariances $K = K_1, K_2$ and K_3, we study the performance of our estimator (1.12) for the domains Ω_2 ($m = 3$) and Ω_4 ($m = 9$), and the number of fragments $n = 100$ and 500, for two different sampling regimes:

(a) *Regularly Observed Fragments.* We simulate n fragments corresponding to Ω over a regular grid of size 100 over the unit interval. And, we estimate the covariance over Ω using the pairwise empirical covariance estimator given by: for s, t such that $n(s,t) = \#\{j : s,t \in U_j\} \geq 10$,

$$\hat{K}_\Omega(s,t) = \frac{1}{n(s,t)} \sum_{j:s,t\in U_j} X_j(s)X_j(t) \tag{1.13}$$

where $U_j \subset I$ denotes the support of the fragment X_j.

(b) *Sparsely Observed Fragments.* We generate n fragments as before in (a) but retain only ~ 6 points for every fragment chosen randomly and discard the rest. We estimate the covariance over Ω by locally linear kernel smoothing. For K_1 and K_2, this is achieved using the fdapace package (Carroll et al., 2021) in R under the default parameters. For K_2, the same method is unsuitable due to non-differentiability at the diagonal, and so we use the reflected triangle estimator proposed in Jouzdani and Panaretos (2021) instead.

Using the estimate of K_Ω, we construct the completion using the method described in Section 1.7. We do this 100 times for every combination of covariance, domain and number of samples. We calculate the median and mean absolute deviation for the error in the form of integrated squared error in estimating K over the observed region Ω and its complement Ω^c to which it is extended using the completion procedure. The results are summarized in Table 1.1. Note that for high sample size combinations in the sparse observation case of $K = K_2$, our computational resources proved to be inadequate for using the available implementation of the reflected triangle estimator to complete the computation. For such cases, the results provided are for $n = 200$ and have been marked with an asterisk.

The choice of truncation parameter can be made using a scree plot or the fraction of variance explained (FVE) criterion given by

$$N_p = \min\{r \geq 1 : \sum_{j=1}^{r} \hat{\lambda}_j(\mathbf{J}_p) > 0.95 \cdot \mathrm{tr}(\mathbf{J}_p)\}.$$

Here, we choose the truncation parameters manually to illustrate how the nature of the covariance affects the choice of the truncation parameter. For a finite rank covariance, the truncation parameter should be close to but not exceeding the rank. Therefore, for K_1, we choose $N_p = 4$. For infinite rank covariances exhibiting fast eigenvalue decay, such as K_2 and K_3, small values of N_p such as 2 or 3 work well. Accordingly, we choose $N_p = 2$ for them. This choice also seems to work slightly better in practice.

The results are summarised in Table 1.1. Naturally, the error in the estimation of K_Ω and $K_\star$ tends to decrease as N increases in all cases. The error in estimating K_Ω tends

to increase as m increases, even relative to the norm of K_Ω. The same applies to the error in estimating $K_\star$ in the case of regular observations, however for sparse observations there does not seem to be a clear relationship.

Table 1.1: Results of General Simulation Study

Median ± Mean Absolute Deviation of of Integrated Squared Errors.

Parameters			Regular Observations		Sparse Observations		Squared Norms	
K	m	N	$\int_\Omega \lvert\hat{K}_\Omega - K_\Omega\rvert^2$	$\int_{\Omega^c} \lvert\hat{K}_\star - K\rvert^2$	$\int_\Omega \lvert\hat{K}_\Omega - K_\Omega\rvert^2$	$\int_{\Omega^c} \lvert\hat{K}_\star - K\rvert^2$	$\int_\Omega \lvert K\rvert^2$	$\int_{\Omega^c} \lvert K\rvert^2$
K_1	3	100	0.0901 ±0.0604	0.0318 ±0.0325	0.1395 ±0.0823	0.1489 ±0.1630	1.3080	0.0381
	3	500	0.0152 ±0.0099	0.0100 ±0.0112	0.0397 ±0.0182	0.0930 ±0.1136	1.3080	0.0381
	9	100	0.1781 ±0.1159	0.1980 ±0.1947	0.1729 ±0.0893	0.0749 ±0.0713	1.3225	0.0236
	9	500	0.0301 ±0.0197	0.0482 ±0.0465	0.0567 ±0.0310	0.0664 ±0.0633	1.3225	0.0236
K_2	3	100	0.0058 ±0.0052	0.0007 ±0.0006	0.0046 ±0.0043	0.0010 ±0.0011	0.1573	0.0094
	3	500	0.0016 ±0.0013	0.0002 ±0.0001	0.0034 ±0.0028*	0.0005 ±0.0005*	0.1573	0.0094
	9	100	0.0128 ±0.0078	0.0013 ±0.0008	0.0068 ±0.0067	0.0009 ±0.0010	0.1614	0.0053
	9	500	0.0025 ±0.0017	0.0003 ±0.0002	0.0037 ±0.0035*	0.0005 ±0.0006*	0.1614	0.0053
K_3	3	100	0.3657 ±0.2879	0.0231 ±0.0230	0.6279 ±0.3913	0.0420 ±0.0391	5.7930	0.0021
	3	500	0.0758 ±0.0604	0.0198 ±0.0119	0.1594 ±0.0943	0.0278 ±0.0220	5.7930	0.0021
	9	100	0.6543 ±0.4339	0.0463 ±0.0329	0.7780 ±0.6108	0.0569 ±0.0536	5.7949	0.0001
	9	500	0.1082 ±0.0764	0.0231 ±0.0098	0.2417 ±0.1508	0.0410 ±0.0266	5.7949	0.0001

*values computed for $n = 200$.

1.10.2 Estimator Performance versus m

We now turn to studying the dependence of the error of estimating $K_\star$ on the number of intervals m corresponding to Ω. To this end, we generate $n = 100$ fragments of the covariances $K = K_1, K_2$ and K_3 corresponding to the domains $\Omega \in \{\Omega_j\}_{j=1}^5$ over a grid of length 100, estimate the partial covariance on the corresponding domain using the pairwise empirical covariance estimator as defined by Equation 1.13 and apply the completion algorithm. We then compute the ratio of relative errors (RRE) defined as the ratio of the median ISE in estimating K_{Ω^c} to that of K_Ω both relative to the norms of the respective quantities they are estimating. In other words,

$$\text{RRE} = \frac{\int_{\Omega^c} |\hat{K}_\star - K|^2 / \int_{\Omega^c} |K|^2}{\int_\Omega |\hat{K}_\Omega - K_\Omega|^2 / \int_\Omega |K|^2}. \tag{1.14}$$

The results have been summarized as boxplots in Figure 1.9. We observe that the median RRE does not vary much in response the number of intervals m for any of the covariance scenarios as we move from smaller values such as $m = 2, 3$ to larger values such as $m = 17$. The most noticeable effect appears to be in the finite rank case, where the interquantile range of the RRE increases with m, even if the median is relatively stable. Importantly, we observe that the increase in error (across scenarios) is nowhere so large so as to affect the utility of the estimation procedure and the empirical performance appears more

optimistic than what predicted by Theorem 1.8.

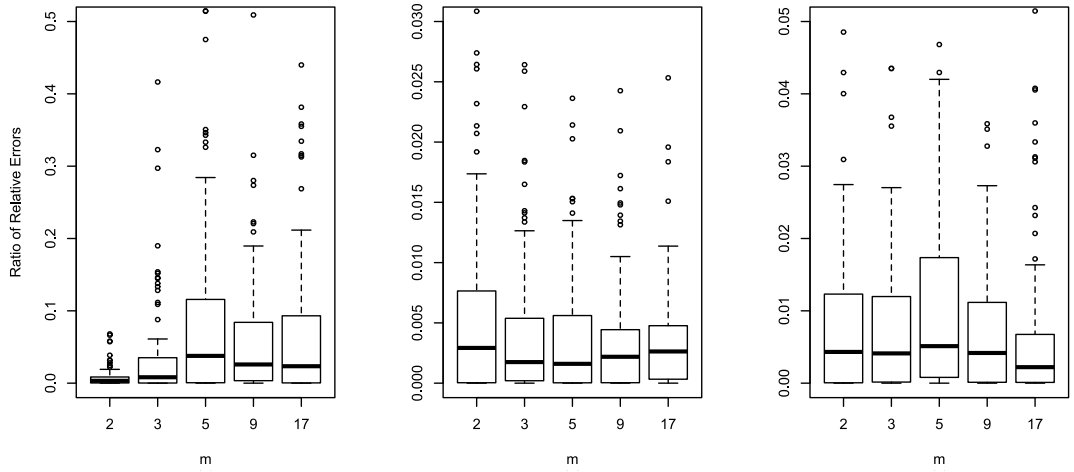


Figure 1.9: Boxplot of Ratio of Relative Errors (RRE) vs. the number of intervals m for K_1 (left), K_2 (middle) and K_3 (right).

1.10.3 Comparative Simulations

In order to benchmark the performance of our estimator $\hat{K}_\star$, we compare to that of the estimator $\hat{K}_p$ proposed in Delaigle et al. (2021). For different choices of K, m and n we generate fragments on a regular grid of size 50 on the unit interval. We then estimate the covariance on Ω using locally linear kernel smoothing and then apply the completion procedure. We do this 100 times and calculate the median and mean absolute deviation of the integrated squared error. We do the same for the estimator $\hat{K}_p$. The results are summarized by Table 1.2. As can be expected, neither estimator dominates, and performance varies according to scenario. The scenarios involving K_1 and K_3 feature covariances that are analytic and exactly or effectively low rank. As expected, $\hat{K}_p$ has better performance here, since these two settings admit unique extension and their infinite smoothness combined with their low (effective) rank is ideally suited for truncated series extrapolation. That being said, the performance of $\hat{K}_\star$ remains competitive, with errors of similar magnitude in these two scenarios. On the other hand, $\hat{K}_\star$ outperforms $\hat{K}_p$ by an order of magnitude in scenario K_2, which is a low regularity scenario without a unique completion. One would summarise that $\hat{K}_\star$ behaves like a "robust" estimator: competitively in "easy" scenarios, but substantially better otherwise. Another overarching observation (in line with intuition and theoretical results) is that the performance of $\hat{K}_\star$ is tied to the performance of the estimator of K_Ω – in some cases (see e.g. scenario K_3), the larger errors relative to $\hat{K}_p$ might have more to do with the quality of estimation on Ω itself, than the with completion procedure.

Table 1.2: Results of Comparative Simulation Study

Median ± Mean Absolute Deviation of of Integrated Squared Errors.

Parameters			Our estimator $\hat{K}_\star$		The estimator $\hat{K}_p$	
K	m	N	$\int_\Omega \lvert\hat{K}_\Omega - K_\Omega\rvert^2$	$\int_{\Omega^c} \lvert\hat{K}_\star - K\rvert^2$	$\int_\Omega \lvert\hat{K}_p - K_\Omega\rvert^2$	$\int_{\Omega^c} \lvert\hat{K}_p - K\rvert^2$
K_1	3	100	0.0779 ±0.0541	0.0984 ±0.1080	0.0614 ±0.0402	0.0742 ±0.0407
	3	500	0.0157 ±0.0103	0.0288 ±0.0295	0.0146 ±0.0086	0.0145 ±0.0099
	9	100	0.1250 ±0.0773	0.1400 ±0.1450	0.0982 ±0.0479	0.1115 ±0.0721
	9	500	0.0279 ±0.0204	0.0622 ±0.0760	0.0225 ±0.0123	0.0224 ±0.0143
K_2	3	100	0.0058 ±0.0053	0.0007 ±0.0005	0.0049 ±0.0041	0.0055 ±0.0049
	3	500	0.0016 ±0.0014	0.0001 ±0.0001	0.0009 ±0.0006	0.0010 ±0.0008
	9	100	0.0129 ±0.0073	0.0012 ±0.0007	0.0081 ±0.0065	0.0059 ±0.0052
	9	500	0.0029 ±0.0020	0.0003 ±0.0002	0.0020 ±0.0017	0.0012 ±0.0009
K_3	3	100	0.0232 ±0.0202	0.0057 ±0.0059	0.0023 ±0.0018	0.0025 ±0.0022
	3	500	0.0053 ±0.0039	0.0009 ±0.0008	0.0006 ±0.0004	0.0006 ±0.0004
	9	100	0.0340 ±0.0254	0.0064 ±0.0054	0.0050 ±0.0035	0.0044 ±0.0034
	9	500	0.0058 ±0.0037	0.0013 ±0.0010	0.0010 ±0.0009	0.0008 ±0.0005

1.11 Illustrative Data Analysis

Following Delaigle et al. (2021), we apply our method to the spine bone mineral density (BMD) data described in Bachrach et al. (1999). We consider measurements of 117 females taken between the ages of 9.5 and 21 years. The measurements for every subject are taken over a short period of time, comprising in each case an interval far shorter than the age-range interval. Hence, the measurements on each subject constitute independent sparsely observed fragments, see Figure 1.10 (left), and yield information only on a partial covariance. Nevertheless, if we wish to conduct further analyses such as classification, regression, prediction, or even dimension reduction, we need access to a complete covariance.

We plot all those pairs of ages for which we have measurements on the same subject, see Figure 1.10 (right). Based on the plot, we infer that the covariance can be estimated reasonably well over the serrated domain Ω (colored in red) corresponding to the intervals $[9.5, 13.5]$, $[11.5, 15.5]$, $[13.5, 17.5]$, $[15, 19.5]$ and $[17, 21]$. We then estimate the covariance on Ω using locally linear kernel smoothing through the fdapace package and use the completion algorithm to estimate the covariance over the entire region, see Figure 1.11.

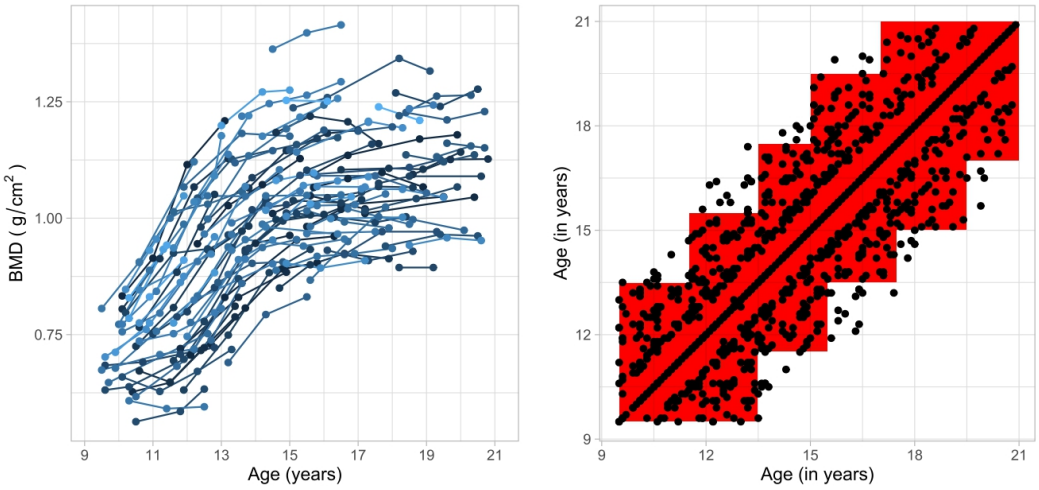


Figure 1.10: (left) Sparsely observed spine BMD curves for 117 females (right) Scatter plot of pairs of ages for which simultaneous observations are available.

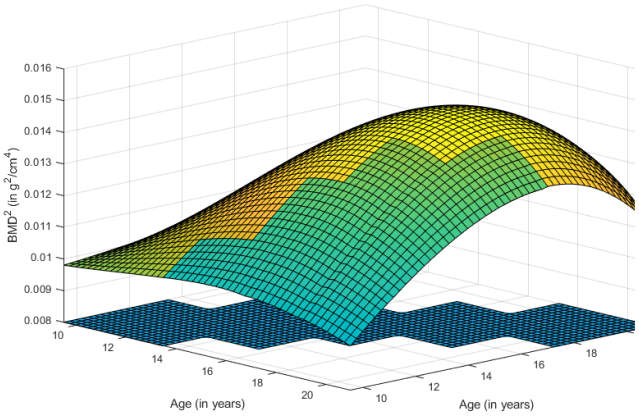


Figure 1.11: Completed covariance of the BMD data.

2 Positive-Definite Completions

I think that it is a relatively good approximation to truth... that mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science.

John von Neumann, *The Mathematician* (1947)

Abstract

We study the positive-definite completion problem for a variety of domains and prove results concerning the existence and uniqueness of solutions and their characterization. Most importantly, we study a special solution called the canonical completion which is the reproducing kernel analogue of the determinant-maximizing completion known to exist for matrices, and establish many results concerning its existence and uniqueness, which include many interesting algebraic and variational characterizations.

Most importantly, we prove the existence of a canonical completion for domains which are equivalent to the band. This extends to the existence of a canonical extension in the context of the classical extension problem of positive-definite functions.

2.1 Introduction

Let X be a set and $\Omega \subset X \times X$. Given a function $K_\Omega : \Omega \to \mathbb{R}$, we consider the problem of extending K_Ω to $X \times X$ such that the resulting extension $K : X \times X \to \mathbb{R}$ is a reproducing kernel, which is to say $K(x, y) = K(y, x)$ for $x, y \in X$ and

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0$$

for every $n \geq 1$, $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$ and $\{x_j\}_{j=1}^n \subset X$. We shall refer to this as a *completion problem* of K_Ω and the extensions K, which can be regarded as its solutions, shall be called *completions*.

The problem has been studied before in the literature for certain special cases by several well-known mathematicians. For finite X, the problem can be understood as that specifying the unspecified entries of a partially specified matrix so as to make it positive semidefinite. In this form, the problem has been studied for the *band* case, where $\Omega = \{(i,j) : |i-j| \leq p\}$ for $X = \{j : 1 \leq j \leq n\}$ for some $n > 1$, by H. Dym and I. Gohberg in Dym and Gohberg (1981), where they derived necessary and sufficient conditions on K_Ω for the existence of a completion K and established the existence of a unique special completion which maximizes the determinant of the matrix $[K(i,j)]_{i,j\in X}$ among all completions K and is the unique completion with the property that the ijth entry of the inverse of the matrix $[K(i,j)]_{i,j\in X}$ vanishes if $(i,j) \notin \Omega$. Grone et al. (1984) studied the problem for general Ω and proved the existence and uniqueness of the special completion assuming the existence of a completion. Necessary and sufficient conditions for the existence of a completion for general Ω were derived by Paulsen et al. (1989). A complete characterization of completions for the band case was arrived at in Gohberg et al. (1989) and the results were also extended to matrices of operators, which can be thought of as operator-valued kernels in our setting (see Bakonyi and Woerdeman (2011) and Paulsen and Raghupathi (2016)).

For infinite X, the completion problem has been studied mostly in the form of the extension problem for positive-definite functions. In this setting, $X = \mathbb{Z}$ or $\mathbb{R}$ usually and one is concerned with positive-definite extensions $\tilde{F}$, that is $\tilde{F} : X \to \mathbb{R}$ such that $\tilde{K}(x,y) = \tilde{F}(x-y)$ for $x, y \in X$ is a reproducing kernel, of positive-definite functions F on $\{x \in X : |x| < a\}$ for some $a > 0$, which is to say that $K(x,y) = F(x-y)$ for $0 \leq x, y < a$ is also a reproducing kernel. In our language, this means that Ω is the band $\{(x,y) : |x-y| < a\} \subset X \times X$ and we only consider *stationary* completions, that is, completions of the form $\tilde{K}(x,y) = \tilde{F}(x-y)$, of stationary K_Ω, which is to say $K_\Omega(x,y) = F(x-y)$ for some $F : \{x \in X : |x| < a\} \to \mathbb{R}$. For $X = \mathbb{Z}$, it was shown by Carathéodory (1907), that every positive-definite function F on $\{x \in \mathbb{Z} : |x| < a\}$ for some integer $a > 0$, admits a positive-definite extension to $\mathbb{Z}$. The analogous result for $X = \mathbb{R}$ was proved by Krein (1940) for continuous F, and later by Artjomenko (1941a) without the continuity assumption. Necessary and sufficient conditions for uniqueness of extension were derived by Keich (1999). A short historical survey of further developments can be found in Sasvári (2006). An analogue of the special solution from the matrix case for $X = \mathbb{Z}$ arose in the work of Burg (1975) concerning spectral estimation for stationary time series. However, no such analogue for $X = \mathbb{R}$ has been studied in the existing literature to the best of our knowledge.

In this chapter, we study the positive-definite completion problem in greater generality, and in particular, without requiring stationarity or finiteness of X. Needless to say, this is

a non-trivial problem because the methods used for proving the classical results discussed previously, such as matrix determinants and factorization or unitary representation, do not generalize in an obvious way to arbitrary reproducing kernels. The problem has not been studied in such a setting before in the existing literature, with the exception of Waghmare (2022). Our approach mostly involves a generous use of simple tools from the theory of reproducing kernels such as contraction maps and inner products of reproducing kernel Hilbert spaces (Paulsen and Raghupathi, 2016), and some results from the theory of tensor products of Hilbert spaces (Ryan, 2002; Treves, 2016), Γ-convergence (Braides, 2002; Dal Maso, 1993) and strongly continuous one-parameter semigroups (Davies, 1980; Engel and Nagel, 2000).

A few of the results presented here (Section 2.5 in particular) appeared in Waghmare (2022) for the special case of X being an interval of $\mathbb{R}$. This chapter represents a more mathematically mature treatment of the subject and the results here are more general and elegant.

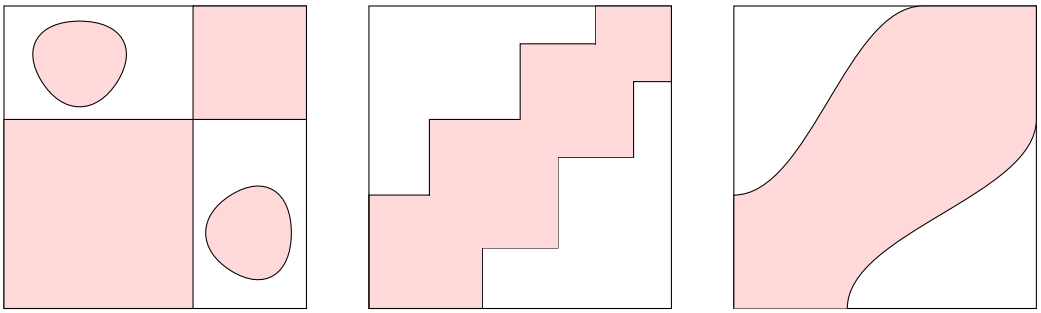
2.1.1 Contributions

We also study the general characteristics of the set of completions and derive a surprisingly simple characterization of its extreme points in terms of their reproducing kernel Hilbert space.

For domains which are, in a certain sense, *large* (see Figure 2.1), we show that positive-definite completion is equivalent to solving a linear equation in the projective tensor product space of certain reproducing kernel Hilbert spaces. As a consequence, we characterize the set of completions in terms of bounded extensions of a linear functional on the tensor product space.

For the class of *serrated* domains, we prove the existence of a unique canonical completion and given an iterative formula involving certain contraction maps for computing it. We derive a particularly simple closed form expression for the inner product of its reproducing kernel Hilbert space. Furthermore, we present many interesting variational characterizations of the canonical completion. Finally, we prove partial analogues of the determinant maximization and inverse zero characterizations. All of these results can be generalized to a even more expansive class of domains we call junction-tree domains.

For $X = \mathbb{R}$, we establish the existence of the analogue of the special completion from the matrix case, which we call the *canonical completion*, for continuous K_Ω on domains Ω which are, in a sense, *band-like*. Importantly, we prove the existence of a *canonical extension* $F_\star$ to $\mathbb{R}$ of a positive-definite function F on $(-a, a) \subset \mathbb{R}$ for some $a > 0$, thus demonstrating the existence of an analogue of the determinant-maximizing special completion from the matrix case for positive-definite functions on $\mathbb{R}$. The extension is shown to correspond to a certain strongly continuous semigroup on a reproducing kernel



(a) A large domain. (b) A serrated domain. (c) A regular domain.

Figure 2.1: Domains. The red region represents Ω.

Hilbert space and consequently, can be thought of as the solution of an abstract Cauchy problem in that space. Under certain technical conditions, we also show the uniqueness of the canonical extension and recover the generator of its semigroup as the closure of a certain operator, which basically amounts to recovering the canonical extension.

2.1.2 Interpretations and Connections to other Problems

Reproducing kernels are everywhere in analysis and probability. In some contexts, they arise purely by virtue of being the essence of positive-definiteness, for example, as positive semidefinite matrices, inner products and Mercer kernels; while elsewhere for less obvious reasons, for example, as characteristic functions of distributions. In this section, we discuss how the completion problem relates to these other contexts.

Fourier Transforms and Characteristic Functions

Positive-definite functions occur naturally as Fourier transforms of finite positive Borel measures in probability and analysis and as characteristic functions of random variables in probability theory. Because the correspondence is precise, we can think of positive-definite extensions of a continuous positive-definite function F on an interval $(-a, a)$ for $a > 0$, as corresponding to Borel measures μ on $\mathbb{R}$ which satisfy

$$\int_{-\infty}^{\infty} e^{itx}\, d\mu(x) = F(t) \quad \text{for } t \in (-a, a). \tag{2.1}$$

This can be regarded as a generalization of the Hamburger moment problem, since the moments of a measure are determined by the values of the Fourier transform around the origin. Krein's result implies the existence of a measure μ satisfying (2.1). Our result concerning the existence of a canonical extension $\tilde{F}$ of F points to the existence of a special solution of the above problem.

Gaussian Processes and Graphical Models

There is a well-known bijective correspondence between reproducing kernels and the covariances of Gaussian processes. The completions K of K_Ω thus correspond to zero-mean Gaussian processes $Y = \{Y_x : x \in X\}$ satisfying

$$\mathbb{E}[Y_x Y_y] = K_\Omega(x, y) \quad \text{for } (x, y) \in \Omega. \tag{2.2}$$

In finite dimensions, the differential entropy of a zero-mean Gaussian distribution is proportional the logarithm of the determinant of its covariance matrix. Therefore, for finite X, the canonical completion $K_\star$ corresponds to the Gaussian process Y which maximizes differential entropy under the constraint (2.2). The canonical completion also has an interesting interpretation in terms of the probability density p of Y because the inverse of $[K_\star(i,j)]_{i,j\in X}$ being zero at the entries corresponding to $(i,j) \notin \Omega$ implies that products of the form $t_i t_j$ for $(i,j) \in \Omega$ do not appear in $p(\mathbf{t})$ where $\mathbf{t} = (t_j)_{j\in X}$.

The canonical completion can also be interpreted in this context for possibly infinite X. It corresponds to the Gaussian process satisfying (2.2) which is Markov with respect to Ω in the extended sense of the *global Markov property*:

$$\mathbb{P}[Y_u \in A, Y_v \in B | Y_S] = \mathbb{P}[Y_u \in A | Y_S]\mathbb{P}[Y_v \in B | Y_S] \tag{2.3}$$

where $A, B \subset \mathbb{R}$ and $Y_S = \{Y_s : s \in S\}$. In other words, the random variables Y_u and Y_v for $u, v \in X$ separated by $S \subset X$ are conditionally independent given Y_S. This is analogous to how the future and the past are conditionally independent given the present for an ordinary Markov process. The global Markov property is of natural way of extending Markovianity to processes indexed by vertices of a graph instead of time. Alternatively, we can say that $K_\star$ is the covariance of the Gaussian graphical model Y corresponding to the "graph" Ω satisfying (2.2).

Constrained Embeddings into Hilbert Spaces

Notice that for every completion K of K_Ω, we can write for the generators $k_x \in \mathcal{H}(K)$ of K given by $k_x(y) = K(x, y)$ for $x, y \in X$, that $\langle k_x, k_y \rangle = K_\Omega(x, y)$ for $(x, y) \in \Omega$. Every completion thus corresponds to an embedding $x \mapsto \varphi_x$ of X into a Hilbert space $\mathcal{H}$ satisfying the constraint that

$$\langle \varphi_x, \varphi_y \rangle = K_\Omega(x, y) \quad \text{for } (x, y) \in \Omega. \tag{2.4}$$

In fact, every such embedding into a Hilbert space $\mathcal{H}$ will be equal, up to isometry, to an embedding of the form $x \mapsto k_x$ into the reproducing kernel Hilbert space $\mathcal{H}(K)$ of some completion K of K_Ω. The set $\mathcal{C} = \mathcal{C}(K_\Omega)$ can thus be regarded as the set of solutions to a constrained embedding problem (2.4).

A canonical solution to the completion problem naturally corresponds to a special solution to the embedding problem. In fact, the canonical solution $K_\star$ can be understood as corresponding to an embedding $x \mapsto k_x$ such that the vectors k_x are, in a sense, *maximally dispersed* in $\mathcal{H}(K)$. When X is finite, this is can be easily formalized by choosing the determinant of the matrix $[K(i,j)]_{i,j\in X}$ as the measure of dispersion, which is only natural given that the determinant is proportional to the "volume" of the simplex formed by the vectors $\{k_j : j \in X\}$ in $\mathcal{H}(K)$. Moreover, it vanishes if $\{k_j : j \in X\}$ are linearly dependent. Furthermore, if Ω is the diagonal $\{(j,j) : j \in X\}$, then by Hadamard's inequality, it follows that the determinant is maximized precisely when the vectors $\{k_j : j \in X\}$ are orthogonal to each other, which perfectly conforms with our intuitive understanding of dispersion.

For infinite X and under certain conditions, we prove a local analogue of the determinant maximization principle, which essentially says that every nice perturbation of a canonical solution $K_\star$ tends to increase the determinant in an appropriate sense, thus justifying the interpretation of the canonical completion in terms of dispersion for infinite X.

Metric Embeddings into Hilbert Spaces

According to Schoenberg's embedding theorem, a metric space (X,d), where $d : X \times X \to \mathbb{R}$ is a distance function on X, can be embedded into a Hilbert space if and only if $K_t(x,y) = e^{-td^2(x,y)}$ is a reproducing kernel for every $t > 0$. Naturally, one can think of a partially specified counterpart of the distance function $d : X \times X \to \mathbb{R}_+$ and this gives rise to the notion of a partially specified metric space (X, d_Ω) where $d_\Omega : \Omega \to \mathbb{R}_+$ for some $\Omega \subset X \times X$ is a partially specified distance function. Many natural phenomenon such as molecules can be regarded as partially specified metric spaces because the distances between two points are not always fixed. Every extension of d_Ω to $X \times X$ which is a valid distance function can be thought of as a *conformation* of the partially specified metric space (X, d_Ω). The problem of determining whether (X, d_Ω) admits a conformation that can be embedded into a Hilbert space is equivalent to that of determining whether there exists an extension d of d_Ω to $X \times X$ such that $K_t(x,y) = e^{-td^2(x,y)}$ where $x, y \in X$ is a completion of $K_{t\Omega}(x,y) = e^{-td^2_\Omega(x,y)}$ where $(x,y) \in \Omega$ for every $t > 0$.

2.1.3 Organization of the Chapter

After discussing some preliminaries in Section 2.2, we begin by treating the general properties of completions in Section 2.3. We then proceed by studying the completion problem while gradually increasing the extent of positive-definiteness imposed on K_Ω. In Section 2.3, we impose no assumption on K_Ω. In Section 2.4 we assume that certain restrictions of K_Ω are reproducing kernels. Finally, from Section 2.5 onwards, we deal exclusively with K_Ω for which every restriction to $A \times A \subset \Omega$ for $A \subset X$ is a reproducing kernel. Section 2.6 is dedicated to study of canonical extensions of positive-definite

functions.

2.2 Preliminaries and Notation

2.2.1 Reproducing Kernels

Let X be a set. A *reproducing kernel* K on X is defined as a function $K : X \times X \to \mathbb{R}$ satisfying $K(x, y) = K(y, x)$ for $x, y \in X$ and

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0$$

for every $n \geq 1$, $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$ and $\{x_j\}_{j=1}^n \subset X$. The functions $k_x : X \to \mathbb{R}$ given by $k_x(y) = K(x, y)$ for $x, y \in X$ are called the *generators* of K. The closure of the linear span of the generators under the norm induced by the inner product $\langle k_x, k_y \rangle = K(x, y)$ for $x, y \in X$ is called the reproducing kernel Hilbert space or associated Hilbert space of K and denoted by $\mathcal{H}(K)$ and associated with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}(K)}$ and the induced norm $\| \cdot \|_{\mathcal{H}(K)}$. To avoid cluttering our notation, we shall always omit the subscript and denote the inner product as $\langle \cdot, \cdot \rangle$ and the norm as $\| \cdot \|$, except in cases where there is a possiblity of confusion. Note that for a function f and kernel K on S, $f \in \mathcal{H}(K)$ with $\|f\| \leq C$ if and only if for some $C > 0$,

$$\left| \sum_{i=1}^{m} \alpha_i f(x_i) \right| \leq C \sqrt{\sum_{i,j=1}^{m} \alpha_i \alpha_j K(x_i, x_j)} \tag{2.5}$$

for every $m \geq 1$, $\{x_i\}_{i=1}^m \subset S$ and $\{\alpha_i\}_{i=1}^m \subset \mathbb{R}$.

For $A \subset X$ and $x \in X$, we define $k_{x,A} : A \to \mathbb{R}$ as $k_{x,A}(y) = K(x, y)$ for $y \in A$. Furthermore, for $A \subset X$, we can define the *subkernel* $K_A : A \times A \to \mathbb{R}$ as the restriction $K_A = K|_{A \times A}$. Naturally, K_A is also a reproducing kernel. Its associated Hilbert space is given by $\mathcal{H}(K_A) = \{f|_A : f \in \mathcal{H}(K)\}$ and the functions $k_{x,A}$ for $x \in A$ are its generators. Using (2.5), one can show that the restriction $\mathcal{J}_A : \mathcal{H}(K) \to \mathcal{H}(K_A)$ given by $f \mapsto f|_A$ is a bounded linear map satisfying $\|f|_A\| \leq \|f\|$ for $f \in \mathcal{H}(K)$, where $\|f|_A\|$ is understood as the norm of $f|_A$ in $\mathcal{H}(K_A)$. Its adjoint $\mathcal{J}_A^* : \mathcal{H}(K_A) \to \mathcal{H}(K)$ is given by $\mathcal{J}_A^* g(x) = \langle \mathcal{J}_A^* g, k_x \rangle = \langle g, \mathcal{J}_A k_x \rangle = \langle g, k_{x,A} \rangle$ which is equal to $g(x)$ for $x \in A$. Notice that $\mathcal{J}_A^* k_{x,A} = k_x$ for $x \in A$. In fact, the associated Hilbert space $\mathcal{H}(K_A)$ is isometrically isomorphic to the closed linear subspace in $\mathcal{H}(K)$ spanned by $\{k_x : x \in A\}$ under the inner product induced by the ambient space and the isometry is given by $\mathcal{J}_A$. This result is known as *subspace isometry*. A direct consequence of this result is that

$$\|\Pi_A f\| = \|f|_A\|. \tag{2.6}$$

for $f \in \mathcal{H}(K)$ where Π_A is the projection to the closed linear subspace spanned by $\{k_x : x \in A\}$. Similarly, the orthogonal complement of a subspace is also isomorphic to a certain reproducing kernel Hilbert space. For $B \subset X$, we can define the *Schur complement* $K/K_B : (X \setminus B) \times (X \setminus B) \to \mathbb{R}$ as $K/K_B(x,y) = K(x,y) - \langle k_{x,B}, k_{y,B} \rangle$. K/K_B is a reproducing kernel because, by subspace isometry, we can write $K(x,y) - \langle k_{x,B}, k_{y,B} \rangle = \langle k_x, k_y \rangle - \langle \Pi_B k_x, \Pi_B k_y \rangle = \langle (k_x - \Pi_B k_x), (k_y - \Pi_B k_y) \rangle$. It can be shown that $\mathcal{H}(K/K_B) = \{f \in \mathcal{H}(K) : f|_B = 0\}$ and that it is isometrically isomorphic to the orthogonal complement of $\Pi_B \mathcal{H}(K)$. Furthermore,

$$\|f - \Pi_B f\| = \|g\|_{\mathcal{H}(K/K_B)} \tag{2.7}$$

where $g = (f - \Pi_B f)|_{X \setminus B}$ or equivalently, $g(y) = f(y) - \langle f|_B, k_{y,B} \rangle$ for $y \in X \setminus B$.

2.2.2 Graphs

Experience with the positive-definite completions of partially specified matrices and their connection to Gaussian graphical models suggests that there is great utility to thinking of a domain Ω as an *undirected graph* (X, Ω) on the set of vertices X, with the vertices $x, y \in X$ being *adjacent* iff $(x,y) \in \Omega$. The pair $(x,y) \in \Omega$ can thus be thought of as the *edge* between x and y, which makes Ω the edge set. Since X will almost always be fixed, we shall often omit writing (X, Ω) and simply identify the graph (X, Ω) with its edge set Ω. Notice that for a set $S \subset X$ such that $S \times S \subset \Omega$, every $x, y \in S$ are adjacent. We call such sets *cliques*. For $x, y \in X$, a *path* in Ω between x and y is a finite sequence $\{z_k\}_{k=0}^{n+1} \subset X$ such that $z_0 = x$, $z_{n+1} = y$ and $(z_k, z_{k+1}) \in \Omega$ for $0 \leq k \leq n$. We say that $x, y \in X$ are *connected* in Ω if there is a path in Ω between them and *disconnected* otherwise. We say $S \subset X$ is a *separating set* or a *separator* of Ω, if there exist $x, y \in X \setminus S$ such that for every path $\{z_k\}_{k=0}^{n+1} \subset X$ between x and y, $z_k \in S$ for some $1 \leq k \leq n$, or in other words, every path between x and y passes through S. Alternatively, $S \subset X$ is a separator if $X \setminus S$ is disconnected. We adopt the convention that, if x and y are disconnected, then they are separated by the empty set $\varnothing$.

2.2.3 Domains and Completions

A *domain* Ω on X is a subset of $X \times X$ which is symmetric in that $(x,y) \in \Omega$ if and only if $(y,x) \in \Omega$ and contains the diagonal $\{(x,x) : x \in X\} \subset X \times X$. If $K_\Omega : \Omega \to \mathbb{R}$ is a function, an extension $K : X \times X \to \mathbb{R}$ of K_Ω which is a reproducing kernel on X is called a *positive-definite completion* or simply, a *completion* of K_Ω.

Definition 2.1 (Completion)**.** *Let K_Ω be a function on a domain Ω on X. A reproducing kernel K on X is called a* completion *of K_Ω if the restriction of K to Ω is K_Ω.*

We shall denote the set of completions of a function K_Ω by $\mathcal{C}(K_\Omega)$ or simply $\mathcal{C}$. The

symmetry of the domain Ω merely accounts for the fact that the completions are themselves symmetric by definition, while containing the diagonal ensures that the set of completions $\mathcal{C}$ is bounded (Theorem 2.1).

Of course, not every such function K_Ω admits a completion. A necessary condition is that suitable restrictions of K_Ω be reproducing kernels. A function $K_\Omega : \Omega \to \mathbb{R}$ on a domain $\Omega \subset X \times X$ is called a *partially reproducing kernel* if for every $A \subset X$ such that $A \times A \subset \Omega$, the restriction $K_A = K_\Omega|_{A\times A}$ is a reproducing kernel on A. Naturally, every reproducing kernel K on X is a partially reproducing kernel on the domain $\Omega = X \times X$. We extend the definition of $k_{x,A}$ for partially reproducing kernels K_Ω by defining them for $x \in X$ and $A \subset X$ such that $x \times A \subset \Omega$ as the functions $k_{x,A} : A \to \mathbb{R}$ given by $k_{x,A}(y) = K(x, y)$ for $y \in A$.

If X is a subset of $\mathbb{R}$ or $\mathbb{Z}$, we call a partially reproducing kernel K_Ω *stationary* if for some $F : X \to \mathbb{R}$ we have $K_\Omega(x, y) = F(x - y)$ for $(x, y) \in \Omega$. Note that this includes reproducing kernels K on X as they can be considered as partially reproducing kernels with $\Omega = X \times X$.

2.2.4 Projective Tensor Product

Consider two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ and their tensor product

$$\mathcal{H}_1 \otimes \mathcal{H}_2 = \text{Span}\{f \otimes g : f \in \mathcal{H}_1 \text{ and } g \in \mathcal{H}_2\}.$$

We define the *projective tensor product norm* or more simply, the π-norm $\|\cdot\|_\pi$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$, as

$$\|\tau\|_\pi = \inf\left\{\sum_{i=1}^{\infty} \|f_i\|\|g_i\| : \tau = \sum_{i=1}^{n} f_i \otimes g_i \text{ where } n \geq 1, f_i \in \mathcal{H}_1, g_i \in \mathcal{H}_2 \text{ for } i \geq 1\right\}.$$

The completion of $\mathcal{H}_1 \otimes \mathcal{H}_2$ under $\|\cdot\|_\pi$ is a Banach space called the projective tensor product space of $\mathcal{H}_1$ and $\mathcal{H}_2$, and denoted by $\mathcal{H}_1 \hat{\otimes}_\pi \mathcal{H}_2$. It turns out that the dual of the projective tensor product space of two Hilbert spaces is isometrically isomorphic to the space of bounded linear operators between them (Ryan, 2002, Chapter 2.2; Treves, 2016, Proposition 43.8). In other words,

$$\left[\mathcal{H}_1 \hat{\otimes}_\pi \mathcal{H}_2\right]^* = \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2) \tag{2.8}$$

and we can think of every $\Phi \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ as a bounded linear functional Φ on $\mathcal{H}_1 \hat{\otimes}_\pi \mathcal{H}_2$ in the following sense: $\Phi[f \otimes g] = \langle \Phi f, g\rangle$ for $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$. The expression $\Phi[\tau]$ is well-defined for every $\tau \in \mathcal{H}_1 \hat{\otimes}_\pi \mathcal{H}_2$ as a result of continuous extension. This result provides an alternative expression for the π-norm which we shall call the *duality formula*

given by

$$\|\tau\|_\pi = \sup\{|\Phi[\tau]| : \Phi \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2), \|\Phi\| \leq 1\}. \tag{2.9}$$

The space $\mathcal{H}_1 \hat{\otimes}_\pi \mathcal{H}_2$ can also be thought of as the space $\mathcal{L}_1(\mathcal{H}_1, \mathcal{H}_2)$ of nuclear operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ and of course, vice-versa (Treves, 2016, Proposition 47.2).

2.3 General Properties of Completions

In this section, we study some of the general properties of the set of completions $\mathcal{C}$ such as convexity and compactness, and their consequence.

Theorem 2.1. *The set of completions $\mathcal{C}(K_\Omega)$ is convex and compact in the topology of pointwise convergence.*

Proof. Let $\mathcal{C}_0 = \{K : |K(x,y)| \leq \sqrt{K_\Omega(x,x)K_\Omega(y,y)} \text{ for } x, y \in X\}$. For $K \in \mathcal{C}_0$, the range of $K(x,y)$ is compact by the Heine-Borel theorem for every $x, y \in X$. By Tychonoff's theorem, $\mathcal{C}_0$ is itself compact in the product topology, which is same as the topology of pointwise convergence.

Notice that $K \in \mathcal{C}$ if and only if $K \in \mathcal{C}_0$, $K(x,y) = K_\Omega(x,y)$ for $(x,y) \in \Omega$, $K(x,y) - K(y,x) = 0$ for $x, y \in X$, and

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0$$

for every $n \geq 1$, $\{x_i\}_{i=1}^n \subset X$ and $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$. Because the expressions $K \mapsto K(x,y)$, $K \mapsto K(x,y) - K(y,x)$ and $K \mapsto \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j)$ are continuous linear functionals under the topology of pointwise convergence, $\mathcal{C}$ is a closed subset of $\mathcal{C}_0$ implying that it is compact. □

Note that $\mathcal{C}$ under the topology of pointwise convergence is not, in general, second-countable, and therefore, compactness does not necessarily imply sequential compactness. In Section 2.4.5, we shall show that $\mathcal{C}$ is also sequentially compact under an additional assumption on Ω.

2.3.1 Convexity

The set of completions $\mathcal{C}$ is a compact convex subset of the space of real-valued functions on $X \times X$ which forms a Hausdorff, locally convex topological vector space under the topology of pointwise convergence. By the Krein-Milman theorem, $\mathcal{C}$ is equal to the closed convex hull of $\text{Ext}(\mathcal{C})$, where $\text{Ext}(\mathcal{C})$ denotes the set of extreme points of $\mathcal{C}$. A completion $K \in \mathcal{C}$ is an extreme point of $\mathcal{C}$ if it can not be represented as a proper

linear combination of other completions. In other words, there do not exist completions $K_1, K_2 \in \mathcal{C}$ such that $K = \alpha K_1 + (1-\alpha)K_2$ for $0 < \alpha < 1$. The following result gives remarkably simple characterization of the set of extreme points of $\mathcal{C}$ in terms of their reproducing kernel Hilbert spaces.

Theorem 2.2. *$K \in \mathrm{Ext}(\mathcal{C})$ if and only if for every self-adjoint $\Psi : \mathcal{H}(K) \to \mathcal{H}(K)$,*

$$\langle k_x, \Psi k_y \rangle = 0 \text{ for } (x,y) \in \Omega \quad \Longrightarrow \quad \Psi = \mathbf{0},$$

where $k_x \in \mathcal{H}(K)$ is given by $k_x(y) = K(x,y)$ for $x, y \in X$.

The above result is a direct consequence of the following lemma, the proof of which can be found in Appendix B.

Lemma 2.1. *Let K be a reproducing kernel on X with the associated Hilbert space $\mathcal{H}$. There is a bijective correspondence between $H : X \times X \to \mathbb{R}$ such that $K + H, K - H \geq O$ and self-adjoint contractions $\Psi \in \mathcal{L}(\mathcal{H})$ given by $H(x,y) = \langle \Psi k_x, k_y \rangle$ for $x, y \in X$, where $k_x \in \mathcal{H}(K)$ is given by $k_x(y) = K(x,y)$ for $x, y \in X$.*

2.3.2 Compactness

An important consequence of compactness in the topology of pointwise convergence is that a completion problem admits a solution if and only if so does every finite subproblem. Let $K_{\Omega\mathcal{F}}$ denote the restriction of K_Ω to the set $\Omega \cap (\mathcal{F} \times \mathcal{F})$.

Theorem 2.3. *$\mathcal{C}(K_\Omega)$ is nonempty if and only if so is $\mathcal{C}(K_{\Omega\mathcal{F}})$ for every finite $\mathcal{F} \subset X$.*

Proof. Let a be a finite subset of X and K_a denote a completion of $K_{\Omega a}$. Define

$$K^a = \begin{cases} K_a(x,y) & x, y \in a \\ 0 & \text{otherwise.} \end{cases}$$

The mapping $a \mapsto K^a$ forms a net on the directed set $A = \{a \subset X : a \text{ is finite}\}$ ordered by inclusion. By compactness of $\mathcal{C}_0$, K^a has a convergent subnet, say $K^b = K^{b(a)}$ which converges to some $K \in \mathcal{C}_0$. It turns out that $K \in \mathcal{C}$. Indeed, $K(x,y) = \lim_b K^b(x,y) = K_\Omega(x,y)$ for $(x,y) \in \Omega$, $K(x,y) - K(y,x) = \lim_b \left[K^b(x,y) - K^b(y,x)\right] = 0$ for $x, y \in X$ and

$$\sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j) = \lim_b \left[\sum_{i,j=1}^n \alpha_i \alpha_j K^b(x_i, x_j) \right] \geq 0$$

for $n \geq 1$, $\{x_i\}_{i=1}^n \subset X$ and $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$. The converse is trivial because if K is a completion of K_Ω, then $K|_{\mathcal{F}\times\mathcal{F}}$ is a completion of $K_{\Omega\mathcal{F}}$. □

For finite X, we have the following result of Paulsen which gives necessary and sufficient conditions for the existence of a completion, in the language of matrices.

Theorem 2.4 (Paulsen et al. (1989, Theorem 2.1))**.** *Let $J \subset \{1, \ldots, n\}^2$ for some $n \geq 1$ such that $(j, j) \in J$ for $1 \leq j \leq n$ and $(i, j) \in J$ if $(j, i) \in J$. A partially specified matrix $T = [t_{ij}]_{(i,j)\in J}$ which is symmetric (i.e. $t_{ij} = t_{ji}$ for $(i, j) \in J$) admits a completion if and only if for every positive semidefinite matrix $M = [m_{ij}]_{i,j=1}^n$ such that $m_{ij} = 0$ for $(i, j) \notin J$ we have*

$$\sum_{(i,j)\in J} m_{ij} t_{ij} \geq 0.$$

The above result gives a concrete but somewhat unwieldy criterion for determining whether $\mathcal{C}$ is nonempty. We shall say that K_Ω is *symmetric*, if $K_\Omega(x, y) = K_\Omega(y, x)$ for $(x, y) \in \Omega$.

Corollary 2.1. *Assume that K_Ω is symmetric. $\mathcal{C}(K_\Omega)$ is nonempty if and only if for every finite $F = \{x_i\}_{i=1}^n \subset X$ and positive semidefinite matrix $M = [m_{ij}]_{i,j=1}^n$ such that $m_{ij} = 0$ for $(x_i, x_j) \notin \Omega$ we have*

$$\sum_{(x_i,x_j)\in\Omega} m_{ij} K_\Omega(x_i, x_j) \geq 0.$$

A criterion of this form can be easily used to work out maximum and minimum values that a completion can have at a given point. Define $m, M : X \times X \to \mathbb{R}$ as

$$M(x, y) = \sup\{K(x, y) : K \in \mathcal{C}(K_\Omega)\} \text{ and } m(x, y) = \inf\{K(x, y) : K \in \mathcal{C}(K_\Omega)\}.$$

We fix $K(x, y) = c$ for some $c \in \mathbb{R}$ and formulate a new completion problem on the domain $\Omega \cup \{(x, y), (y, x)\}$ for a new function equal to K_Ω on Ω and c on $\{(x, y), (y, x)\}$. By Corollary 2.1, the function admits a completion if and only if for every finite $F = \{x, y\} \cup \{x_k\}_{k=1}^n \subset X$ and positive semidefinite matrix $M = [m_{ij}]$ where $i, j \in \{x, y\} \cup \{k\}_{k=1}^n$ such that m_{ij}, m_{xj} and m_{iy} are zero when (x_i, x_j), (x, x_j) and (x_i, y) is not in Ω, respectively, we have

$$2m_{xy}c + 2 \sum_{(x,x_j)\in\Omega} m_{xj} K_\Omega(x, x_j) + 2 \sum_{(x_i,y)\in\Omega} m_{iy} K_\Omega(x_i, y) + \sum_{(x_i,x_j)\in\Omega} m_{ij} K_\Omega(x_i, x_j) \geq 0.$$

Define for the pair (M, F) where M and F are as described above,

$$\mathcal{R}_{xy}(M, F) = \frac{-1}{m_{xy}} \left[\sum_{(x,x_j)\in\Omega} m_{xj} K_\Omega(x, x_j) + \sum_{(x_i,y)\in\Omega} m_{iy} K_\Omega(x_i, y) + \frac{1}{2} \sum_{(x_i,x_j)\in\Omega} m_{ij} K_\Omega(x_i, x_j) \right].$$

By working out the values of c for which the above statement is true the following result becomes apparent.

Theorem 2.5. *Assume that K_Ω is symmetric and $\mathcal{C}$ is nonempty. We have*

$$M(x,y) = \inf_{m_{xy}<0} \mathcal{R}_{xy}(M,F) \text{ and } m(x,y) = \sup_{m_{xy}>0} \mathcal{R}_{xy}(M,F).$$

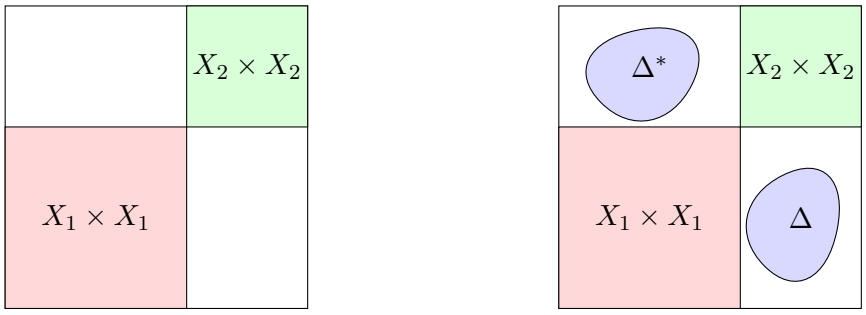
Notice that the value of a completion K at a point $(x,y) \in \Omega^c$ is uniquely determined if and only if $m(x,y) = M(x,y)$. Using this observation, it is not difficult to see why the following result holds.

Theorem 2.6. *Assume that K_Ω is symmetric. $\mathcal{C}$ is a singleton if and only if for every $(x,y) \in \Omega^c$ and $\epsilon > 0$ there exist pairs (M,F) and (M',F') where $M' = [m'_{ij}]$ such that $m_{xy} < 0$, $m'_{xy} > 0$ and*

$$\mathcal{R}_{xy}(M,F) - \mathcal{R}_{xy}(M',F') < \epsilon.$$

2.4 Completion on Large Domains

We say that a domain Ω is *large* if there exist $X_1, X_2 \subset X$ such that $X = X_1 \cup X_2$ and $X_1 \times X_1, X_2 \cup X_2 \subset \Omega$. Let $\Delta = \Omega \cap (X_2 \times X_1)$ and $\Delta^* = \Omega \cap (X_1 \times X_2)$ (see Figure 2.2). We shall assume throughout this section that the restrictions $K_{X_1} = K_\Omega|_{X_1 \times X_1}$ and $K_{X_2} = K_\Omega|_{X_2 \times X_2}$ are reproducing kernels.



(a) $\Delta = \varnothing$ (b) $\Delta \neq \varnothing$

Figure 2.2: Large domain. The colored regions represent Ω.

2.4.1 Contractions and Completions

We begin by considering the special case where Δ is empty (see Figure 2.2a). For every $U \subset X$ such that $U \times U \subset \Omega$ and $u \in U$, we denote $k_{u,U} : U \to \mathbb{R}$ as $k_{u,U}(x) = K_\Omega(x,u)$.

Theorem 2.7 (Contraction Characterization)**.** *Let K_Ω be a partially reproducing kernel on a domain $\Omega = (X_1 \times X_1) \cup (X_2 \times X_2)$ where $X_1, X_2 \subset X$ (see Figure 2.3). There is a bijective correspondence between the completions K of K_Ω and contractions $\Phi : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{X_2})$ satisfying $\Phi k_{x,X_1} = k_{x,X_2}$ for $x \in X_1 \cap X_2$ given by*

$$K(x,y) = \langle \Phi k_{x,X_1}, k_{y,X_2} \rangle \quad \text{for } x \in X_1 \text{ and } y \in X_2. \tag{2.10}$$

If $X_1 \cap X_2 = \varnothing$, then there is a bijective correspondence between the completions K of K_Ω and the contractions $\Phi : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{X_2})$.

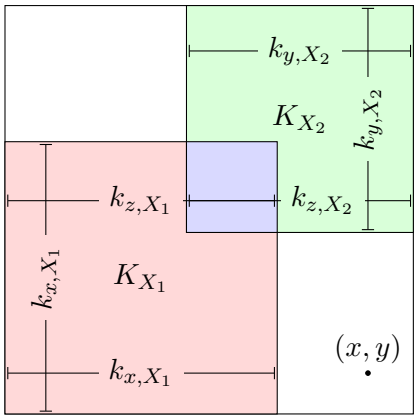


Figure 2.3: The coloured region represents Ω with the kernels K_{X_1} and K_{X_2} being represented by the red and blue regions and blue and green regions, respectively. The functions $k_{u,U}$ are being represented by $\mapsto$ at the position corresponding to their values relative to the kernel.

Proof of Theorem 2.7. Let K be a completion of K_Ω. Define $\Phi_0 : \mathrm{Span}\{k_{x,X_1} : x \in X_1\} \to \mathcal{H}(K_{X_2})$ as $\Phi_0 k_{x,X_1} = k_{x,X_2}$. For $m, n \geq 1$, let $\{x_i\}_{i=1}^m \subset X_1$, $\{y_k\}_{k=1}^n \subset X_2$ and $\{\alpha_i\}_{i=1}^m, \{\beta_k\}_{k=1}^n \subset \mathbb{R}$. By positive definiteness, the expression

$$\sum_{i,j=1}^{m} \alpha_i \alpha_j K(x_i, x_j) + 2 \sum_{i,k=1}^{m,n} \alpha_i \beta_k K(x_i, y_k) + \sum_{k,l=1}^{n} \beta_k \beta_l K(y_k, y_l) \geq 0 \tag{2.11}$$

is non-negative. This can be rewritten in terms of Φ_0 and $f = \sum_{i=1}^m \alpha_i k_{x_i,X_1}$ and $g = \sum_{k=1}^n \beta_k k_{y_k,X_2}$ as follows

$$\langle f, f \rangle + 2\langle \Phi_0 f, g \rangle + \langle g, g \rangle \geq 0$$

Replacing g with $-g$ and using continuity of the inner product, we get for every $g \in \mathcal{H}(K_{X_2})$,

$$|\langle \Phi_0 f, g \rangle| \leq \frac{1}{2} \left[\langle f, f \rangle + \langle g, g \rangle \right].$$

For $\|f\|, \|g\| \leq 1$, we have $|\langle \Phi_0 f, g \rangle| \leq 1$ and thus, $\|\Phi_0 f\| \leq 1$. It follows that for $f \in \mathrm{Span}\{k_{x,X_1} : x \in X_1\}$, $\|\Phi_0 f\| \leq \|f\|$. As a consequence, Φ_0 uniquely extends by continuity to a contraction Φ on $\mathcal{H}(K_{X_1})$ satisfying $\Phi k_{x,X_1} = k_{x,X_2}$ for $x \in X_1 \cap X_2$ by construction.

To show the converse, let $\Phi : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{X_2})$ be a contraction satisfying $\Phi k_{x,X_1} =$

k_{x,X_2} for $x \in X_1 \cap X_2$. Define

$$K(x,y) = \begin{cases} K_\Omega(x,y) & \text{for } (x,y) \in \Omega, \\ \langle \Phi k_{x,X_1}, k_{y,X_2} \rangle & \text{if } x \in X_1 \setminus X_2 \text{ and } y \in X_2 \setminus X_1, \\ \langle \Phi k_{y,X_1}, k_{x,X_2} \rangle & \text{if } y \in X_1 \setminus X_2 \text{ and } x \in B \setminus X_1. \end{cases} \tag{2.12}$$

By construction, K is symmetric. We can write (2.11) using the Cauchy-Schwarz inequality as

$$\begin{aligned} \langle f, f \rangle + 2\langle \Phi f, g \rangle + \langle g, g \rangle &\geq \|f\|^2 - 2\|f\|\|g\| + \|g\|^2 \\ &= (\|f\| - \|g\|)^2 \geq 0. \end{aligned}$$

It follows that K is indeed a completion. Hence proved. □

This is a slightly more general analogue of a standard operator-theoretic result (Bakonyi and Woerdeman, 2011, Lemma 2.4.4) concerning the necessary and sufficient conditions for the non-negativity of a 2×2 operator matrix with only the diagonal entries specified, which was derived by Baker (Baker, 1973, Theorem 1) in the context of joint Gaussian measures on Hilbert spaces.

2.4.2 Existence, Characterization and Uniqueness

We are now prepared to deal with the more general case where Δ can be non-empty. We can assume without any loss of generality that $X_1 \cap X_2 = \varnothing$, by simply taking $X_2 = X \setminus X_1$. Notice that every completion of K_Ω is also a completion of $K_\Omega|_{(X_1 \times X_1) \cup (X_2 \times X_2)}$. By Theorem 2.7, we can write every completion K of K_Ω as in (2.10) for some contraction $\Phi : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{X_2})$ satisfying $\langle \Phi k_{x,X_1}, k_{y,X_2} \rangle = K(x,y)$ for $(x,y) \in \Delta$, which can be thought of as a linear equation in Φ. Indeed, using (2.8) allows us to rewrite it as

$$\Phi[k_{x,X_1} \otimes k_{y,X_2}] = K_\Omega(x,y) \quad \text{for } (x,y) \in \Delta, \tag{2.13}$$

where Φ is a bounded linear functional on the projective tensor product space $\mathcal{H}(K_{X_1}) \hat{\otimes}_\pi \mathcal{H}(K_{X_2})$. It follows that Φ is a bounded extension of the linear functional $\Phi_0 : \mathrm{Span}\{k_{x,X_1} \otimes k_{y,X_2} : (x,y) \in \Delta\} \to \mathbb{R}$ given by

$$\Phi_0[k_{x,X_1} \otimes k_{y,X_2}] = K_\Omega(x,y) \quad \text{for } (x,y) \in \Delta. \tag{2.14}$$

If Φ_0, thus defined, is a linear functional of norm not exceeding 1, then the Hahn-Banach theorem guarantees the existence of an extension Φ of Φ_0 to $\mathcal{H}(K_{X_1}) \hat{\otimes}_\pi \mathcal{H}(K_{X_2})$ such that $\|\Phi\| = \|\Phi_0\|$. In fact, every extension Φ of norm not exceeding 1 will correspond to a completion of K_Ω according to (2.10). On the other hand, if Φ_0 is not well-defined or $\|\Phi_0\| > 1$, then K_Ω does not admit a completion. The following result summarizes our

discussion.

Theorem 2.8 (Existence and Characterization of Completion)**.** *Consider $K_\Omega : \Omega \to \mathbb{R}$ on a domain Ω on X. Assume that there exists a partition $\{X_i\}_{i=1}^2$ of X such that for every i, $X_i \times X_i \subset \Omega$ and $K_{X_i} = K_\Omega|_{X_i \times X_i}$ is a reproducing kernel. The following statements hold:*

1. *The function K_Ω admits a completion to X if and only if* (2.14) *defines a bounded linear functional $\Phi_0 : \mathrm{Span}\{k_{x,X_1} \otimes k_{y,X_2} : (x,y) \in \Delta\} \to \mathbb{R}$ such that $\|\Phi_0\| \leq 1$ or equivalently,*
$$\left|\sum_{j=1}^n \alpha_j K_\Omega(x_j, y_j)\right| \leq \left\|\sum_{j=1}^n \alpha_j k_{x_j,X_1} \otimes k_{y_j,X_2}\right\|_\pi \tag{2.15}$$
for every $n \geq 1$, $\{(x_j, y_j)\}_{j=1}^n \subset \Delta$ and $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$.

2. *There is a bijective correspondence between the completions K of K_Ω and bounded extensions Φ of Φ_0 to $\mathcal{H}(K_{X_1}) \hat{\otimes}_\pi \mathcal{H}(K_{X_2})$ satisfying $\|\Phi\| \leq 1$ given by*
$$\Phi[k_{x,X_1} \otimes k_{y,X_2}] = K(x,y) \quad \text{for } x \in X_1 \text{ and } y \in X_2.$$

In essence, Theorem 2.7 together with the isomorphism (2.8) allowed us to *linearize* the completion problem for K_Ω by framing it as a linear equation (2.13) on a tensor product space.

Equation (2.15) is sometimes called *Helly's theorem* or *extension principle* (see (Narici and Beckenstein, 2010, Theoreom 7.10.1) and (Edwards, 2012, 2.3.1 Theorem)). It is reminiscent of the condition (2.5) for a function to belong to a reproducing kernel Hilbert space. Equation (2.15) can also be used to derive tight lower and upper bounds for the values of completions at points outside Δ. To find the maximum value $M(x,y)$ and minimum value $m(x,y)$ of $K(x,y)$ for some $(x,y) \in (X_1 \times X_2) \setminus \Delta$ over the completions K of K_Ω, we consider an augmented completion problem: let $\tilde{\Omega} = \Omega \cup \{(x,y),(y,x)\}$ and define $K_{\tilde{\Omega}} : \tilde{\Omega} \to \mathbb{R}$ as $K_{\tilde{\Omega}}|_\Omega = K_\Omega$ and $K_{\tilde{\Omega}}(x,y) = K_{\tilde{\Omega}}(y,x) = \nu$. There exists a completion K of K_Ω which satisfies $K(x,y) = \nu$ if and only if $K_{\tilde{\Omega}}$ admits a completion, which is when

$$\left|\nu - \textstyle\sum_{j=1}^n \alpha_j K_\Omega(x_j,y_j)\right| \leq \left\|k_{x,X_1} \otimes k_{y,X_2} - \textstyle\sum_{j=1}^n \alpha_j k_{x_j,X_1} \otimes k_{y_j,X_2}\right\|_\pi$$

for every $n \geq 1$, $\{(x_j,y_j)\}_{j=1}^n \subset \Delta$ and $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$. It follows that

$$M(x,y) = \inf\left\{\textstyle\sum_{j=1}^n \alpha_j K_\Omega(x_j,y_j) + \left\|k_{x,X_1} \otimes k_{y,X_2} - \sum_{j=1}^n \alpha_j k_{x_j,X_1} \otimes k_{y_j,X_2}\right\|_\pi\right\}$$
$$m(x,y) = \sup\left\{\textstyle\sum_{j=1}^n \alpha_j K_\Omega(x_j,y_j) - \left\|k_{x,X_1} \otimes k_{y,X_2} - \sum_{j=1}^n \alpha_j k_{x_j,X_1} \otimes k_{y_j,X_2}\right\|_\pi\right\}$$

where the supremum and infimum are taken over $n \geq 1$, $\{(x_j,y_j)\}_{j=1}^n \subset \Delta$ and $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$. Note that the value of a completion K s uniquely determined at (x,y) if and only if

$m(x,y) = M(x,y)$. If $m(x,y) = M(x,y)$ for every (x,y) outside Δ, then K_Ω admits a unique completion. The following result is now immediate.

Theorem 2.9 (Uniqueness of Completion)**.** *Let K_Ω be as in Theorem 2.8. Then K_Ω admits a unique completion if and only if for every $(x,y) \in X_1 \times X_2 \setminus \Delta$ and $\epsilon > 0$ there exist $n \geq 1$, $\{(x_j, y_j)\}_{j=1}^n \subset \Delta$ and $\{\alpha_j\}_{j=1}^n, \{\beta_j\}_{j=1}^n \subset \mathbb{R}$ such that*

$$\sum_{j=1}^n (\alpha_j - \beta_j) K_\Omega(x_j, y_j) + \left\{ \begin{array}{l} \left\| k_{x,X_1} \otimes k_{y,X_2} - \sum_{j=1}^n \alpha_j k_{x_j,X_1} \otimes k_{y_j,X_2} \right\|_\pi \\ - \left\| k_{x,X_1} \otimes k_{y,X_2} + \sum_{j=1}^n \beta_j k_{x_j,X_1} \otimes k_{y_j,X_2} \right\|_\pi \end{array} \right\} < \epsilon.$$

In particular, this holds if $\mathrm{Span}\{k_{x,X_1} \otimes k_{y,X_2} : (x,y) \in \Delta\}$ *is dense in* $\mathcal{H}(K_{X_1}) \hat{\otimes}_\pi \mathcal{H}(K_{X_2})$.

Remark 2.1. *The linearization approach to completion can be used to determine the existence of completions with given constraints so long as the constraints are linear for Φ. For example, if we want to ascertain whether there exists a completion K of K_Ω for which $K(x,y) = K(x',y')$ for some points $(x,y),(x',y')$ outside Ω, we need only to impose an additional constraint on Φ_0, that is $\Phi_0[k_{x,X_1} \otimes k_{y,X_2} - k_{x',X_1} \otimes k_{y',X_2}] = 0$ and check if $\|\Phi_0\| \leq 1$ as before. We can do the same for a partial derivative $\partial_1 K$ of a completion K which can be expressed as $\partial_1 K(x,y) = \langle \Phi k'_{x,X_1}, k_{y,X_2} \rangle$ for some $k'_{x,X_1} \in \mathcal{H}(K_{X_1})$ under appropriate conditions. This allows us to find the maximum and minimum values of the derivative of a completion at any point.*

2.4.3 Completion on Large Regular Domains

Although Equation (2.15) may appear too unwieldy to be of any use, it is quite straightforward to apply it for bootstrapping on results for finite domains such as those concerning completions of matrices.

Theorem 2.10. *Let Ω be a large regular domain on $X = [0,1]$. Every partially reproducing kernel K_Ω admits a completion.*

Proof. Pick $n \geq 1$, $\{(x_i, y_i)\}_{i=1}^n \subset \Delta$ and $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$. Let $\mathcal{F}_1 = \{x_i\}_{i=1}^n$, $\mathcal{F}_2 = \{y_i\}_{i=1}^n$ and $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$. We consider the completion problem as restricted to $\mathcal{F} \times \mathcal{F}$. According to a classical result (Grone et al., 1984, Theorem 7) concerning the completions of partially specified Hermitian matrices, K_Ω restricted to $\Omega \cap (\mathcal{F} \times \mathcal{F})$ admits an extension to $\mathcal{F} \times \mathcal{F}$. By Theorem 2.8, this means

$$\left| \sum_{i=1}^n \alpha_i K_\Omega(x_i, y_i) \right| \leq \left\| \sum_{i=1}^n \alpha_i k_{x_i,\mathcal{F}_1} \otimes k_{y_i,\mathcal{F}_2} \right\|_\pi .$$

Let $R_1 : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{\mathcal{F}_1})$ and $R_2 : \mathcal{H}(K_{X_2}) \to \mathcal{H}(K_{\mathcal{F}_2})$ denote the restrictions to $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively. Observe that for every contraction $\Phi_F : \mathcal{H}(K_{\mathcal{F}_1}) \to \mathcal{H}(K_{\mathcal{F}_2})$ there

exists a contraction $\Phi : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{X_2})$ such that $\Phi = R_2^* \Phi_F R_1$:

$$\langle \Phi_F k_{x,\mathcal{F}_1}, k_{y,\mathcal{F}_2} \rangle = \langle \Phi_F R_1 k_{x,X_1}, R_2 k_{y,X_2} \rangle = \langle R_2^* \Phi_F R_1 k_{x,X_1}, k_{y,X_2} \rangle.$$

Using the duality formula (2.9), we can write

$$\begin{aligned} \| \textstyle\sum_{i=1}^n \alpha_i k_{x_i,\mathcal{F}_1} \otimes k_{y_i,\mathcal{F}_2} \|_\pi &= \sup\{ | \textstyle\sum_{i=1}^n \alpha_i \langle \Phi_F k_{x_i,\mathcal{F}_1}, k_{y_i,\mathcal{F}_2} \rangle | : \|\Phi_F\| \le 1 \} \\ &\le \sup\{ | \textstyle\sum_{i=1}^n \alpha_i \langle \Phi k_{x_i,X_1}, k_{y_i,X_2} \rangle | : \|\Phi\| \le 1 \} \\ &= \| \textstyle\sum_{i=1}^n \alpha_i k_{x_i,X_1} \otimes k_{y_i,X_2} \|_\pi. \end{aligned}$$

Therefore,

$$\left| \sum_{i=1}^n \alpha_i K_\Omega(x_i, y_i) \right| \le \left\| \sum_{i=1}^n \alpha_i k_{x_i,X_1} \otimes k_{y_i,X_2} \right\|_\pi$$

and the conclusion follows from Theorem 2.8. The converse is trivially true because any completion of K_Ω restricted to $\mathcal{F} \times \mathcal{F}$ is a completion of $K_\Omega|_{\mathcal{F}\times\mathcal{F}}$. □

Of course, we could have derived the result far more easily using Theorem 2.3. But this was good preparation for proving Artjomenko's generalization of Krein's extension theorem which is what follows.

2.4.4 Extension of Positive-definite Functions

Let $F : (-a, a) \to \mathbb{R}$ be a positive-definite function for some $a > 0$. An extension $\tilde{F} : (-2a, 2a) \to \mathbb{R}$ of F is a positive-definite function such that $\tilde{F}|_{(-a,a)} = F$. To express the extension problem of F as a completion problem on a large domain, let $X = [0, 2a)$ with $X_1 = [0, a)$ and $X_2 = [a, 2a)$. Define $K_\Omega : \Omega \to \mathbb{R}$ as $K_\Omega(x, y) = F(x - y)$ for $\Omega = \{(x, y) : |x - y| < a\} \subset X \times X$. The extensions $\tilde{F}$ of F correspond to the stationary completions $\tilde{K}$ of K_Ω. As discussed in Remark 2.1, we can account for the stationarity of K by imposing an additional constaint on Φ_0. Define $\mathcal{S}, \mathcal{T} \subset X_1 \times X_2$ as

$$\begin{aligned} \mathcal{S} &= \mathrm{Span}\{k_{x,X_1} \otimes k_{y,X_2} : y - x < a\}, \text{ and} \\ \mathcal{T} &= \mathrm{Span}\{k_{x,X_1} \otimes k_{y,X_2} - k_{w,X_1} \otimes k_{z,X_2} : y - x = z - w\}. \end{aligned}$$

To show that a stationary completion $\tilde{K}$ exists, we need to show that there exists a contraction Φ such that $\langle \Phi k_{x,X_1}, k_{y,X_2} \rangle = F(y - x)$ and $\Phi[\tau] = 0$ for $\tau \in \mathcal{T}$.

Theorem 2.11. *Every positive-definite function F of $(-a, a) \subset \mathbb{R}$ for some $a > 0$ admits an extension to $(-2a, 2a)$.*

Proof. As before, we construct a grid. Pick $\delta > 0$ and let $n = \max\{j : j\delta < a\}$. Let $\mathcal{F}_1 = \{x_i\}_{i=1}^n$ where $x_i = a - \delta i \in X_1$ for $1 \le i \le n$ and $\mathcal{F}_2 = \{y_j\}_{j=0}^n$ where $y_j = a + \delta j \in X_2$ for $0 \le j \le n$. Let $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ The restriction of K_Ω to $\mathcal{F} \times \mathcal{F}$ can

now be thought of as a partially specified matrix $\mathbf{A} = [A_{ij}]_{i,j=1}^{2n+1}$ where

$$A_{ij} = \begin{cases} F(\delta|i-j|) & \text{for } |i-j| \leq n+1 \\ \text{unspecified.} & \text{for } |i-j| > n+1 \end{cases}$$

By Carathéodory's result, this partially specified matrix admits a positive-definite completion which is also Toeplitz. We can argue as in Theorem 2.10 that

$$|\Phi_0[\sigma]| \leq \|\sigma + \tau\|_\pi \tag{2.16}$$

for $\sigma \in \mathcal{S}_0$ and $\tau \in \mathcal{T}_0$ for dense subsets $\mathcal{S}_0 \subset \mathcal{S}$ and $\mathcal{T}_0 \subset \mathcal{T}$ given by

$$\mathcal{S}_0 = \cup_{\delta>0} \left[\text{Span}\{k_{x_i,X_1} \otimes k_{y_j,X_2} : \delta|j-i| \leq a\}\right]$$
$$\mathcal{T}_0 = \cup_{\delta>0} \left[\text{Span}\{k_{x_i,X_1} \otimes k_{y_j,X_2} - k_{x_k,X_1} \otimes k_{y_l,X_2} : i-j = k-l\}\right]$$

The density follows from the observation that every bounded linear functional which vanishes on $\mathcal{S}_0$ ($\mathcal{T}_0$) vanishes on all of $\mathcal{S}$ ($\mathcal{T}$). The inequality (2.16) implies that Φ_0 is well-defined. Using extension of continuity, we have that (2.16) holds for all $\sigma \in \mathcal{S}$ and $\tau \in \mathcal{T}$. The conclusion now follows from the Hahn-Banach theorem as in Theorem 2.8. □

Since every positive-definite function on $(-a, a)$ can be extended to a positive-definite function on $(-2a, 2a)$ for any $a > 0$, we can iterate the argument and conclude:

Corollary 2.2. *Every positive-definite function $F : (-a, a) \to \mathbb{R}$ for some $a > 0$, admits an extension $\tilde{F}$ to the real line.*

Needless to say, we can derive analogous expressions for the maximum and minimum values of the extension on $(-2a, 2a)$ as well as conditions for uniqueness.

2.4.5 Beyond Large Domains

There does not appear to be an obvious way of extending the linearization technique to "smaller" domains, say if Ω is a domain on X such that $\Omega \supset \cup_i (X_i \times X_i)$ for some partition $\{X_i\}_{i=1}^p$ of X where $p > 2$. However, we can still draw some general conclusions using Theorem 2.7.

Theorem 2.12. *Let Ω be as above and $K_\Omega : \Omega \to \mathbb{R}$ be such that the restrictions $K_{X_i} = K_\Omega|_{x_i \times x_i}$ are reproducing kernels. Then there exists a positive semidefinite operator matrix $[\Phi_{ij}]_{i,j=1}^p$ of contractions $\Phi_{ij} : \mathcal{H}(K_{X_j}) \to \mathcal{H}(K_{X_j})$ such that*

$$K(x, y) = \langle \Phi_{ji} k_{x,X_i}, k_{y,X_j} \rangle \quad \text{for } x \in X_i \text{ and } y \in X_j.$$

A consequence of the Banach-Alaoglu theorem and the above embedding of completions into the product of unit balls in $\mathcal{L}(\mathcal{H}(K_{X_i}), \mathcal{H}(K_{X_j}))$ is the following result.

Theorem 2.13. *If X can be partitioned into a finite or countably infinite number of cliques X_i in Ω, then the set of completions $\mathcal{C}$ of K_Ω is sequentially compact under the topology of pointwise convergence.*

2.5 Canonical Completion

In this section, we study special solutions of the completion problem we shall call canonical completions. We begin by introducing a family of domains for which doing this is particularly simple.

Definition 2.2 (Serrated Domain)**.** *Let X be a set. We say that a domain Ω on X is a n-*serrated domain *if there exists $n \geq 1$ and subsets $\{X_j\}_{j=1}^n$ of X such that* (a) $X = \cup_{j=1}^n X_j$ (b) $X_i \cap X_k \subset X_i \cap X_j$ *for* $1 \leq i < j < k \leq n$ *and* (c) $\Omega = \cup_{j=1}^n (X_j \times X_j)$.

Furthermore, every n-serrated domain is a serrated domain.

We shall derive interesting characterizations of canonical completions in terms of their associated norms and contraction maps. We shall also prove partial analogues of the classical results concerning determinant maximization and inverse zero properties known for matrices. We shall also extend some of these results to a larger families of domains.

2.5.1 Contractions

Let K be a reproducing kernel on X. For $A, B \subset X$, let $K_A = K|_{A\times A}$ and $K_B = K|_{B\times B}$. Define $\Phi_{BA} : \mathcal{H}(K_A) \to \mathcal{H}(K_B)$ as the unique bounded linear map satisfying $\Phi_{BA} k_{x,A} = k_{x,B}$ for $x \in A$. By thinking of $K_{A\cup B}$ as the completion of $K_\Omega = K|_\Omega$ where $\Omega = (A \times A) \cup (B \times B)$, we can deduce using Theorem 2.7 that Φ_{BA} is a contraction. We shall see that these contraction maps can be used to construct completions.

Theorem 2.14 (Properties of Contraction Maps)**.** *Let $A, B \subset X$ and $f \in \mathcal{H}(K_A)$.*

1. *Adjoint.* $\Phi_{BA}^* = \Phi_{AB}$,
2. *Evaluation.* $\Phi_{BA} f(y) = \langle f, k_{y,A} \rangle$ *for* $y \in B$.
3. *Restriction. If $B \subset A$, then* $\Phi_{BA} f = f|_B$,
4. *Minimum Norm Interpolation. If $A \subset B$, then*

$$\Phi_{BA} f = \underset{g\in\mathcal{H}(K_B)}{\arg\min} \{\|g\| : g|_A = f\}$$

Proof. Property (1) follows from writing

$$\langle k_{y,B}, \Phi_{BA} k_{x,A} \rangle = \langle k_{x,B}, k_{y,B} \rangle = K(x,y) = \langle k_{x,A}, k_{y,A} \rangle = \langle \Phi_{AB} k_{y,B}, k_{x,A} \rangle.$$

for every $x \in A$ and $y \in B$. For properties (2) and (3), notice that $\Phi_{BA} f(y) = \langle \Phi_{BA} f, k_{y,B} \rangle = \langle f, \Phi_{AB} k_{y,B} \rangle = \langle f, k_{y,A} \rangle$ for $y \in B$ which is equal to $f(y)$ if $A \subset B$. Finally, to show property (4), let $g \in \mathcal{H}(K_B)$ such that $g|_A = f$. Then

$$\langle g - \Phi_{BA} f, \Phi_{BA} f \rangle = \langle \Phi_{AB} g, f \rangle - \langle f, \Phi_{AB} \Phi_{BA} f \rangle = \langle g|_A, f \rangle - \langle f, f \rangle = 0$$

and we can write $\|g\|^2 = \|g - \Phi_{BA} f\|^2 + \|\Phi_{BA} f\|^2$ which implies that the norm of g is minimum precisely when $g = \Phi_{BA} f$. Hence proved. □

2.5.2 Canonical Completion for 2-Serrated Domains

Consider a set X with subsets X_1 and X_2 such that $X_1 \cup X_2 = X$. Let $\Omega = (X_1 \times X_1) \cup (X_2 \times X_2)$. Let K_Ω be a partially reproducing kernel on Ω. In other words, $K_{X_1} = K_\Omega|_{X_1 \times X_1}$ and $K_{X_2} = K_\Omega|_{X_2 \times X_2}$ are reproducing kernels on X_1 and X_2 respectively. Using Theorem 2.3, we can argue as in Theorem 2.10, that K_Ω admits a completion. Furthermore, by Theorem 2.7, the set $\mathcal{C}$ of completions K is parametrized by contractions $\Phi : \mathcal{H}(K_{X_1}) \to \mathcal{H}(K_{X_2})$ satisfying $\Phi k_{x,X_1} = k_{x,X_2}$ for $x \in X_1 \cap X_2$ according to the relation

$$K(x,y) = \langle \Phi k_{x,X_1}, k_{y,X_2} \rangle$$

for $x \in X_1$ and $y \in X_2$. The case where $x \in X_2$ and $y \in X_1$ is covered by the symmetry of K.

We shall construct a special completion $K_\star$ of K. Notice that for $x \in X_1 \cap X_2$, the contraction map $\Phi_{X_1 \cap X_2, X_1}$ maps k_{x,X_1} to $k_{x,X_1 \cap X_2}$ and the contraction map $\Phi_{X_2, X_1 \cap X_2}$ maps $k_{x,X_1 \cap X_2}$ to k_{x,X_2}. It follows that the product $\Phi_\star = \Phi_{X_2, X_1 \cap X_2} \Phi_{X_1 \cap X_2, X_1}$ satisfies $\Phi_\star k_{x,X_1} = k_{x,X_2}$ for $x \in X_1 \cap X_2$ and is obviously a contraction by virtue of being the product of two contractions. We have thus successfully constructed a member of the family of contractions Φ which parametrizes $\mathcal{C}$. Corresponding to the constructed contraction $\Phi_\star$ is a completion $K_\star$ of K given by

$$K_\star(x,y) = \langle \Phi_\star k_{x,X_1}, k_{y,X_2} \rangle$$

for $x \in X_1$ and $y \in X_2$. If for $x \in X$ and $U \subset X$, we define $k^\star_{x,U} : U \to \mathbb{R}$ as $k^\star_{x,U}(y) = K_\star(x,y)$, then we can also describe $K_\star$ in terms of its generators as

$$k^\star_{x,X_2} = \Phi k_{x,X_1} \text{ for } x \in X_1 \tag{2.17}$$

or equivalently, $k^\star_{y,X_1} = \Phi^* k_{y,X_2}$ for $y \in X_2$. Alternatively, we can express $K_\star$ without

using any contraction maps as

$$\begin{aligned} K_\star(x,y) &= \langle \Phi_\star k_{x,X_1}, k_{y,X_2} \rangle \\ &= \langle \Phi_{X_1\cap X_2,X_1} k_{x,X_1}, \Phi_{X_1\cap X_2,X_2} k_{y,X_2} \rangle \\ &= \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2} \rangle \end{aligned}$$

for $x \in X_1$ and $y \in X_2$. Needless to say, the completion does not change if we switch X_1 and X_2, which implies that is a property of K_Ω only and does not depend on how Ω is parametrized in terms of X_1 and X_2. The following result summarizes the above discussion.

Theorem 2.15. *Let X be a set with subsets $X_1, X_2 \subset X$ such that $X_1 \cup X_2 = X$ and let $\Omega = (X_1 \times X_1) \cup (X_2 \times X_2)$. If K_Ω is a partially reproducing kernel on Ω, then it admits a completion given by*

$$K_\star(x,y) = \langle \Phi_\star k_{x,X_1}, k_{y,X_2} \rangle$$

for $x \in X_1$ and $y \in X_2$, where $\Phi_\star = \Phi_{X_2,X_1\cap X_2}\Phi_{X_1\cap X_2,X_1}$. Furthermore, $K_\star$ can also be expressed as

$$K_\star(x,y) = \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2} \rangle \tag{2.18}$$

for $x \in X_1$ and $y \in X_2$.

Minimum Norm Interpolation

We can acquire deeper insight into the construction of $K_\star$ by understanding the nature of the contraction $\Phi_\star$. Notice that we can rewrite Equation (2.17) as $k^\star_{x,X_2} = \Phi_{X_2,X_1\cap X_2} k_{x,X_1\cap X_2}$ for $x \in X_1$. It follows that $k^\star_{x,X_2}$ for $x \in X_1$ is the minimum norm interpolation of $k_{x,X_1\cap X_2}$ in $\mathcal{H}(K_{X_2})$:

$$k^\star_{x,X_2} = \operatorname*{arg\,min}_{f\in\mathcal{H}(K_{X_2})} \{\|f\| : f|_{X_1\cap X_2} = k_{x,X_1\cap X_2}\} \quad \text{for } x \in X_1. \tag{2.19}$$

Similarly, we can write for $y \in X_2$, that $k^\star_{y,X_1} = \arg\min\{\|g\|\}$ over $g \in \mathcal{H}(K_{X_1})$ such that $g|_{X_1\cap X_2} = k_{y,X_1\cap X_2}$.

Characterization of Completions

In a certain sense, the canonical completion $K_\star$ lies at the center of the set of completions $\mathcal{C}$, which allows us to come up with a simpler characterizations of completions of K_Ω than provided by Theorem 2.8. The following result is a reproducing kernel counterpart of a classic result (Johnson, 1990, Section II) in the theory of matrix completions.

Theorem 2.16. *There is a bijective correspondence between the completions K of K_Ω*

and the contractions $\Psi : \mathcal{H}(K_{X_1}/K_{X_1\cap X_2}) \to \mathcal{H}(K_{X_2}/K_{X_1\cap X_2})$ *given by*

$$K(x,y) = \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2}\rangle + \langle \Psi p_x, q_y\rangle \tag{2.20}$$

for $x \in X_1 \setminus X_2$ *and* $y \in X_2 \setminus X_1$*, where* $p_x(u) = K_{X_1}/K_{X_1\cap X_2}(x,u)$ *and* $q_y(u) = K_{X_2}/K_{X_1\cap X_2}(y,u)$*.*

Proof. Define the partially reproducing kernel D_Ω on Ω given by $D_\Omega(x,y) = K_\Omega(x,y) - \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2}\rangle$ for $(x,y) \in \Omega$. Clearly,

$$D_\Omega(x,y) = \begin{cases} K_{X_1}/K_{X_1\cap X_2}(x,y) & x,y \in X_1 \setminus X_2 \\ 0 & x \text{ or } y \in X_1 \cap X_2 \\ K_{X_2}/K_{X_1\cap X_2}(x,y) & x,y \in X_2 \setminus X_1 \end{cases}$$

which means that D_Ω is indeed a partially reprocuding kernel. Notice that because $D_\Omega(x,y) = 0$ for $x,y \in X_1 \cap X_2$, completing D_Ω is equivalent to completing $D_\Omega|_{\tilde{\Omega}}$ to a kernel on $X \setminus (X_1 \cap X_2)$ where $\tilde{\Omega} = [(X_1 \setminus X_2) \times (X_1 \setminus X_2)] \cup [(X_2 \setminus X_1) \times (X_2 \setminus X_1)]$. The setting of completing $D_\Omega|_{\tilde{\Omega}}$ is equivalent to that of Theorem 2.7 and therefore, the completions D of D_Ω are characterized by $D(x,y) = \langle \Psi p_x, q_y\rangle$ for $x \in X_1 \setminus X_2$ and $y \in X_2 \setminus X_1$ where $\Psi : \mathcal{H}(K_{X_1}/K_{X_1\cap X_2}) \to \mathcal{H}(K_{X_2}/K_{X_1\cap X_2})$ is a contraction.

Notice that there is a bijective correspondence between the completions K of K_Ω and the completions D of D_Ω given by $D(x,y) = K(x,y) - \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2}\rangle$ for $x,y \in X$. Indeed, for every completion K we can write for $x,y \in X$

$$D(x,y) = \langle k_x - \Pi_{X_1\cap X_2} k_x, k_y - \Pi_{X_1\cap X_2} k_y\rangle$$

which satisfies $D|_\Omega = D_\Omega$ and is clearly a reproducing kernel. On the other hand, for every completion D, we have for $x,y \in X$,

$$K(x,y) = D(x,y) + \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2}\rangle.$$

Since$(x,y) \mapsto \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2}\rangle$ $K|_\Omega = K_\Omega$ and D are reproducing kernels so is K and clearly, $K|_\Omega = K_\Omega$. The conclusion follows. □

Using Theorem 2.1, we can obtain a slightly more elegant characterization of $\mathcal{C}$ by simply observing that $\mathcal{C}$ is centered around $K_\star$ and for $H : X \times X \to \mathbb{R}$, $K_\star + H \in \mathcal{C}$ if and only if $K_\star - H \in \mathcal{C}$.

Corollary 2.3. *There is a bijective correspondence between the completions* K *of* K_Ω *and self-adjoint contractions* $\Psi : \mathcal{H}(K_\star) \to \mathcal{H}(K_\star)$ *satisfying* $\langle \Psi k_x^\star, k_y^\star\rangle = 0$ *for* $(x,y) \in \Omega$ *given by*

$$K(x,y) = \langle (\mathbf{I} + \Psi) k_x^\star, k_y^\star\rangle \tag{2.21}$$

for $x, y \in X$, where $k_x^\star \in \mathcal{H}(K_\star)$ is given by $k_x^\star(y) = K_\star(x, y)$.

The Associated Inner Product

We shall now derive the inner product associated with the canonical completion for 2-serrated domains. To this end, the following lemma is useful.

Lemma 2.2. *If $K = K_\star$, then we have $K/K_{X_1} = K_{X_2}/K_{X_1 \cap X_2}$ and $K/K_{X_2} = K_{X_1}/K_{X_1 \cap X_2}$.*

Proof. For $x, y \in X_2 \setminus X_1$

$$\begin{aligned} K/K_{X_1}(x,y) &= K(x,y) - \langle k_{x,X_1}, k_{y,X_1} \rangle \\ &= K(x,y) - \langle \Phi_{X_1, X_1 \cap X_2} \Phi_{X_1 \cap X_2, X_2} k_{x,X_2}, \Phi_{X_1, X_1 \cap X_2} \Phi_{X_1 \cap X_2, X_2} k_{y,X_2} \rangle \\ &= K(x,y) - \langle \Phi_{X_1, X_1 \cap X_2} k_{x, X_1 \cap X_2}, \Phi_{X_1, X_1 \cap X_2} k_{y, X_1 \cap X_2} \rangle \\ &= K(x,y) - \langle \Phi_{X_1 \cap X_2, X_1} \Phi_{X_1, X_1 \cap X_2} k_{x, X_1 \cap X_2}, k_{y, X_1 \cap X_2} \rangle \\ &= K(x,y) - \langle k_{x, X_1 \cap X_2}, k_{y, X_1 \cap X_2} \rangle = K_{X_2}/K_{X_1 \cap X_2}(x,y). \end{aligned}$$

We can argue similarly that $K/K_{X_2} = K_{X_1}/K_{X_1 \cap X_2}$. Hence proved. □

Thus, when $K = K_\star$, the Schur complements K/K_{X_1} and K/K_{X_2} also simplify.

Theorem 2.17. *Let K_Ω and $K_\star$ be as before. The norm associated with $K_\star$ can be expressed as follows:*

$$\|f\|_\star^2 = \|f_{X_1}\|^2 - \|f_{X_1 \cap X_2}\|^2 + \|f_{X_2}\|^2. \tag{2.22}$$

where $\|f_U\|$ denotes the norm of $f_U = f|_U$ in $\mathcal{H}(K_U)$ for $U \subset X$ such that $U \times U \subset \Omega$. Consequently, the inner product is given by

$$\langle f, g \rangle_\star = \langle f_{X_1}, g_{X_1} \rangle - \langle f_{X_1 \cap X_2}, g_{X_1 \cap X_2} \rangle + \langle f_{X_2}, g_{X_2} \rangle. \tag{2.23}$$

Proof. Notice that for $f \in \mathcal{H}$ we can write

$$\|f\|^2 = \|\Pi_{X_1} f + f - \Pi_{X_1} f\|^2 = \|\Pi_{X_1} f\|^2 + \|f - \Pi_{X_1} f\|^2 + 2\langle \Pi_{X_1} f, f - \Pi_{X_1} f \rangle$$

where $\langle \Pi_{X_1} f, f - \Pi_{X_1} f \rangle = 0$ by the projection theorem. By (2.6) and (2.7), $\|\Pi_{X_1} f\| = \|f_{X_1}\|$ and $\|f - \Pi_{X_1} f\| = \|g\|_{\mathcal{H}(K/K_{X_1})}$ where $g \in \mathcal{H}(K/K_{X_1})$ is given by $g(x) = (f - \Pi_{X_1} f)(x)$ for $x \in X_2 \setminus X_1$. Thus,

$$\|f\|^2 = \|f_{X_1}\|^2 + \|g\|^2_{\mathcal{H}(K/K_{X_1})}. \tag{2.24}$$

On a closer look, $g(x) = \langle f - \Pi_{X_1} f, k_x \rangle = f(x) - \langle f_{X_1}, k_{x,X_1,x} \rangle$ where

$$\begin{aligned} \langle f_{X_1}, k_{x,X_1,x} \rangle &= \langle f_{X_1}, \Phi_{X_1,X_1\cap X_2} \Phi_{X_1\cap X_2,X_2} k_{x,X_2} \rangle \\ &= \langle \Phi_{X_1\cap X_2,X_1} f_{X_1}, \Phi_{X_1\cap X_2,X_2} k_{x,X_2} \rangle \\ &= \langle f_{X_1\cap X_2}, k_{x,X_1\cap X_2} \rangle \end{aligned}$$

implying that $g(x) = f(x) - \langle f_{X_1\cap X_2}, k_{x,X_1\cap X_2} \rangle = (f_{X_2} - \Pi_{X_1\cap X_2} f_{X_2})(x)$ for $x \in X_2 \setminus X_1$. By Lemma 2.2, $K/K_{X_1}(x,y) = K_{X_2}/K_{X_1\cap X_2}(x,y)$ for $x, y \in X_2 \setminus X_1$. It follows that the norm of g in $\mathcal{H}(K \setminus K_{X_1})$ can be written as

$$\|g\|^2_{\mathcal{H}(K/K_{X_1})} = \|f_{X_2} - \Pi_{X_1\cap X_2} f_{X_2}\|^2 = \|f_{X_2}\|^2 - \|f_{X_1\cap X_2}\|^2. \tag{2.25}$$

The conclusion follows by substituting (2.25) in (2.24). The inner product formulas can be derived using the observation that $\langle f, g \rangle = \frac{1}{4} \left[\|f+g\|^2 - \|f-g\|^2 \right]$. Hence proved. □

Projections

Let Π_U denote the projection $\mathcal{H}(K)$ to the closed linear subspace spanned by $\{k_{u,X} : u \in U\}$. Using the equivalence between a projection Π_U and restriction to U, we can rewrite (2.22) as

$$\begin{aligned} \langle f, f \rangle &= \|f_{X_1}\|^2 + \|f_{X_2}\|^2 - \|f_{X_1\cap X_2}\|^2 \\ &= \langle \Pi_{X_1} f, f \rangle + \langle \Pi_{X_2} f, f \rangle - \langle \Pi_{X_1\cap X_2} f, f \rangle \\ &= \langle (\Pi_{X_1} + \Pi_{X_2} - \Pi_{X_1\cap X_2}) f, f \rangle. \end{aligned}$$

In fact, we can express that K is the canonical completion purely in terms of these projection operators on $\mathcal{H}(K)$.

Theorem 2.18. *Let K be a completion of a partially reproducing kernel K_Ω on a 2-serrated domain Ω. The following statements are equivalent.*

1. $K = K_\star$,
2. $\mathbf{I} - \Pi_{X_1} - \Pi_{X_2} + \Pi_{X_1\cap X_2} = \mathbf{0}$, *and*
3. $\Pi_{X_1\cap X_2} = \Pi_{X_1}\Pi_{X_2} = \Pi_{X_2}\Pi_{X_1}$.

Proof. We reason as follows:

($1 \implies 2$) By Lemma 2.17, we can write $\|f\|^2$ as

$$\begin{aligned}\langle f, f\rangle &= \|f_{X_1}\|^2 + \|f_{X_2}\|^2 - \|f_{X_1\cap X_2}\|^2 \\ &= \langle \Pi_{X_1} f, f\rangle + \langle \Pi_{X_2} f, f\rangle - \langle \Pi_{X_1\cap X_2} f, f\rangle \\ &= \langle (\Pi_{X_1} + \Pi_{X_2} - \Pi_{X_1\cap X_2}) f, f\rangle.\end{aligned}$$

Thus, $\langle (\mathbf{I} - \Pi_{X_1} - \Pi_{X_2} + \Pi_{X_1\cap X_2}) f, f\rangle$ for $f \in \mathcal{H}$.

($2 \implies 3$) Multiplying both sides of the above equation with Π_{X_1} gives

$$\Pi_{X_1} - \Pi_{X_1}\Pi_{X_1} - \Pi_{X_1}\Pi_{X_2} + \Pi_{X_1}\Pi_{X_1\cap X_2} = -\Pi_{X_1}\Pi_{X_2} + \Pi_{X_1\cap X_2} = \mathbf{0}.$$

Similarly, we can show $\Pi_{X_2}\Pi_{X_1} = \Pi_{X_1\cap X_2}$.

($3 \implies 1$) For $x \in X_1 \setminus X_2$ and $y \in X_2 \setminus X_1$, $K(x,y) = \langle k_x, k_y\rangle = \langle \Pi_{X_1} k_x, \Pi_{X_2} k_y\rangle = \langle \Pi_{X_2}\Pi_{X_1} k_x, k_y\rangle = \langle \Pi_{X_1\cap X_2} k_x, k_y\rangle = \langle k_{x,X_1\cap X_2}, k_{y,X_1\cap X_2}\rangle = K_\star(x,y)$. Similarly, $K(x,y) = K_\star(x,y)$ for $x \in X_2 \setminus X_1$ and $y \in X_1 \setminus X_2$.

Hence proved. □

Separation and Inheritance

It turns out that (2.18) in Theorem 2.15 holds more generally for a separator $S \subset X$ of x, y in Ω, so long as we replace $k_{u,X_1\cap X_2}$ with $k^\star_{u,S}$ for $u = x, y$ and we can write

$$K_\star(x,y) = \langle k^\star_{x,S}, k^\star_{y,S}\rangle. \tag{2.26}$$

Note that $S \subset X$ is a separator if and only if $S \subset X_1 \cap X_2$, which is to say $X_1 \cap X_2$ is the minimal separator of Ω.

There is an alternative way of looking at (2.26). Consider the partially reproducing kernel $K_{\tilde\Omega} = K_\star|_{\tilde\Omega}$ for the 2-serrated domain $\tilde\Omega = \cup_{j=1}^2 (S_j \times S_j)$ where $S_1 = X_1 \cup S$ and $S_2 = X_2 \cup S$. Equation (2.26) is now equivalent to saying that the canonical completion of $K_{\tilde\Omega}$ is same as $K_\star$ and the restriction $K_{\tilde\Omega}$ can be said to *inherit* the canonical completion of K_Ω. In other words, for any 2-serrated domain $\tilde\Omega$ which contains Ω, the canonical completion of $K_\star|_{\tilde\Omega}$ is $K_\star$. We shall now use this insight to prove (2.26).

Theorem 2.19. *If S separates $x \in X_1$ and $y \in X_2$, then $K_\star(x,y) = \langle k^\star_{x,S}, k^\star_{y,S}\rangle$.*

Proof. Define $S_1 = X_1 \cup S$ and $S_2 = X_2 \cup S$. Let $\bar\Omega = \cup_{j=1}^2 (S_j \times S_j)$ and $K_{\bar\Omega} = K_\star|_{\bar\Omega}$. The proof now reduces to showing that the canonical completion of $K_{\bar\Omega}$ is $K_\star$. It suffices to show that the associated inner products are equal. By Theorem 2.17, the inner product

associated with the canonical completion of $K_{\bar{\Omega}}$ is

$$\|f_{S_1}\|^2 - \|f_{S_1 \cap S_2}\|^2 + \|f_{S_2}\|^2 = \|f_{S_1}\|^2 - \|f_S\|^2 + \|f_{S_2}\|^2 \tag{2.27}$$

However, K_{S_1} is itself the canonical completion of $K_{\Omega_1} = K_\Omega|_{\Omega_1}$ where $\Omega_1 = (X_1 \times X_1) \cup (S \times S)$ as can be verifed from Equation (2.18) and therefore, we can write

$$\|f_{S_1}\|^2 = \|f_{X_1}\|^2 - \|f_{X_1 \cap S}\|^2 + \|f_S\|^2 \tag{2.28}$$

by Theorem 2.17. Using the same reasoning for $\Omega_2 = (S \times S) \cup (X_2 \times X_2)$ and $\Omega_3 = [(X_1 \cap S) \times (X_1 \cap S)] \cup [(X_2 \cap S) \times (X_2 \cap S)]$, we can write

$$\|f_{S_2}\|^2 = \|f_S\|^2 - \|f_{X_2 \cap S}\|^2 + \|f_{X_2}\|^2 \tag{2.29}$$

$$\|f_S\|^2 = \|f_{X_1 \cap S}\|^2 - \|f_{X_1 \cap X_2}\|^2 + \|f_{X_2 \cap S}\|^2 \tag{2.30}$$

Substituting Equations (2.28), (2.29) and (2.30) in the expression (2.27) yields

$$\|f_{X_1}\|^2 - \|f_{X_1 \cap X_2}\|^2 + \|f_{X_2}\|^2$$

as desired. The conclusion follows. □

2.5.3 Canonical Completion for Serrated Domains

In the last section, the canonical completion $K_\star$ of a 2-serrated domain was defined by construction. We now give a general definition of the canonical completion of any partially reproducing kernel in terms of separation.

Definition 2.3 (Canonical Completion)**.** *A completion $K_\star$ of a partially reproducing kernel K_Ω is called a* canonical completion*, if we have*

$$K_\star(x, y) = \langle k^\star_{x,S}, k^\star_{y,S} \rangle$$

for every $x, y \in X$ separated by $S \subset X$ in Ω, where $k^\star_{u,U} : U \to \mathbb{R}$ for $u \in X$ and $U \subset X$ is given by $k^\star_{u,U}(v) = K_\star(u, v)$.

Our construction of $K_\star$ for a partially reproducing kernel K_Ω on a 2-serrated domain can be iteratively extended to any serrated domain. Observe that we can extend K_Ω by extending its restriction to $(X_i \times X_i) \cup (X_{i+1} \times X_{i+1})$ using canonical completion to $(X_i \cup X_{i+1}) \times (X_i \cup X_{i+1})$. This results in a partially reproducing kernel on a $(m-1)$-serrated domain and continuing the procedure results in a completion of K_Ω to $X \times X$ in $m - 1$ steps. It turns out that regardless of the order of the 2-serrated completions, one always recovers the same completion which is actually the unique canonical completion of K_Ω in the sense of Defintion 2.3. The proof is not straightforward, but using some

clever argumentation, we shall now reduce this statement to verifying the separation property for a 2-serrated domain.

Theorem 2.20 (Canonical Completion for Serrated Domains)**.** *Let K_Ω be a partially reproducing kernel on a serrated domain Ω on X. Then the following statements hold.*

1. *K_Ω admits a unique canonical completion $K_\star$.*

2. *If $x \in X_i$ and $y \in X_j$ for some $1 \leq i < j \leq n$, then*

$$K_\star(x, y) = \langle [\Phi_{j,j-1} \cdots \Phi_{i+2,i+1}\Phi_{i+1,i}]\, k_{x,X_i}, k_{y,X_j} \rangle$$

   where for $|p-q| = 1$, the mapping $\Phi_{p,q} : \mathcal{H}_q \to \mathcal{H}_p$ is given by $\Phi_{p,q} = \Phi_{X_p, X_p \cap X_q} \Phi_{X_p \cap X_q, X_q}$.

3. *The norm $\|\cdot\|_\star$ associated with the canonical completion $K_\star$ of K_Ω can be expressed as*

$$\|f\|_\star^2 = \sum_{j=1}^{n} \|f_{X_j}\|^2 - \sum_{j=1}^{n-1} \|f_{X_j \cap X_{j+1}}\|^2$$

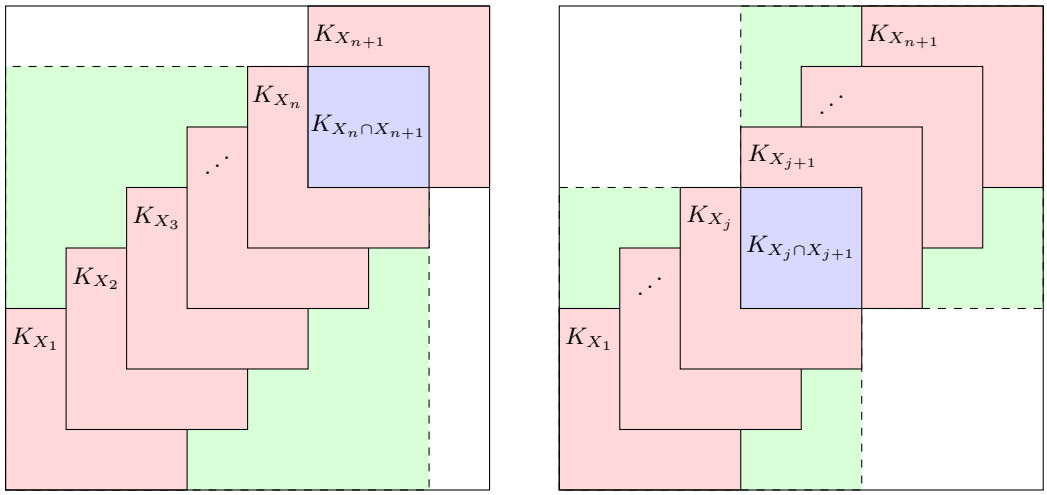
   where $\|f_U\|$ for $U \subset X$ denotes the norm of f_U in $\mathcal{H}(K_U)$.

Proof. Let Ω be an m-serrated domain. We proceed by induction on m. The base case $m = 2$ follows from Theorem 2.15 and Theorem 2.17. Assuming that the statement holds for every $m \leq n$ for some $n \geq 2$, we shall show that it holds for the case $m = n + 1$.

Consider an $(n+1)$-serrated domain $\Omega = \cup_{j=1}^{n+1}(X_j \times X_j)$. Let $\bar{X}_1 = \cup_{j=1}^{n} X_j$, $\bar{X}_2 = X_{n+1}$ and $\Omega_{1n} = \cup_{j=1}^{n}(X_j \times X_j)$, $K_{\Omega_{1n}} = K_\Omega|_{\Omega_{1n}}$ and K_{1n} be the unique canonical completion of $K_{\Omega_{1n}}$. It follows from the induction hypothesis that the restriction to $\bar{X}_1 \times \bar{X}_1$ of a canonical completion $K_\star$ of K_Ω has to be a canonical completion of $K_{\Omega_{1n}}$ and thus, equal to K_{1n}. Define the 2-serrated domain $\bar{\Omega} = (\bar{X}_1 \times \bar{X}_1) \cup (\bar{X}_2 \times \bar{X}_2)$ (see Figure 2.4a) and the partially reproducing kernel $K_{\bar{\Omega}} : \bar{\Omega} \to \mathbb{R}$ as

$$K_{\bar{\Omega}}(x, y) = \begin{cases} K_{1n}(x, y) & \text{if } x, y \in \cup_{j=1}^{n} X_j \\ K_\Omega(x, y) & \text{if } x, y \in X_{n+1}. \end{cases}$$

If $K_\star : X \times X \to \mathbb{R}$ is a canonical completion of $K_{\bar{\Omega}}$ then for $x \in \cup_{j=1}^{n} X_j \setminus X_{n+1}$ and $y \in X_{n+1} \setminus X_n$ we must have $K_\star(x, y) = \langle \bar{k}_{x,X_n \cap X_{n+1}}, \bar{k}_{y,X_n \cap X_{n+1}} \rangle$, where $\bar{k}_{x,X_n \cap X_{n+1}}(y) = K_{\bar{\Omega}}(x, y)$, by taking $S = X_n \cap X_{n+1}$ as the separator in Definition 2.3. It follows that if K_Ω admits a canonical completion, the it must be $K_\star$, which is the canonical completion



(a) The coloured region represents $\bar{\Omega}$.

(b) The coloured region represents $\tilde{\Omega}$.

Figure 2.4: The red and blue region represents Ω.

of $K_{\bar{\Omega}}$. Notice that

$$\begin{aligned} K_\star(x,y) &= \langle \bar{k}_{x,X_n \cap X_{n+1}}, \bar{k}_{y,X_n \cap X_{n+1}} \rangle \\ &= \langle \Phi_{X_n \cap X_{n+1}, X_n} \left[\Phi_{n,n-1} \cdots \Phi_{i+1,i} \right] k_{x,X_i}, [\Phi_{X_n \cap X_{n+1}, X_{n+1}}] k_{y,X_{n+1}} \rangle \\ &= \langle \Phi_{n+1,n} \left[\Phi_{n,n-1} \cdots \Phi_{i+1,i} \right] k_{x,X_i}, k_{y,X_{n+1}} \rangle \end{aligned}$$

for $x \in X_i$ and $y \in X_{n+1}$. Furthermore, we can calculate the associated norm $\|\cdot\|_\star$ associated with $\mathcal{H}(K_\star)$ using Theorem 2.17, as follows

$$\begin{aligned} \|f\|_\star^2 &= \|f_{\bar{X}_1}\|^2 - \|f_{\bar{X}_1 \cap \bar{X}_2}\|^2 + \|f_{\bar{X}_2}\|^2 \\ &= \left[\textstyle\sum_{j=1}^{n} \|f_{X_j}\|^2 - \sum_{j=1}^{n-1} \|f_{X_j \cap X_{j+1}}\|^2 \right] - \|f_{X_n \cap X_{n+1}}\|^2 + \|f_{X_{n+1}}\|^2 \\ &= \textstyle\sum_{j=1}^{n+1} \|f_{X_j}\|^2 - \sum_{j=1}^{n} \|f_{X_j \cap X_{j+1}}\|^2 \end{aligned}$$

for $f \in \mathcal{H}(K_\star)$.

It remains to be shown that $K_\star$ is a canonical completion of K_Ω. Let $x, y \in X$ such that they are separated by $S \subset X$ in Ω. Then x, y must also be separated by a minimal separator $X_i \cap X_{i+1} \subset S$ for some $1 \leq i \leq n$. Let $\tilde{X}_1 = \cup_{j=1}^{i} X_j$, $\tilde{X}_2 = \cup_{j=i+1}^{n+1} X_j$ and $\tilde{\Omega} = (\tilde{X}_1 \times \tilde{X}_1) \cup (\tilde{X}_2 \times \tilde{X}_2)$ (see Figure 2.4b). Consider the partially reproducing kernel $K_{\tilde{\Omega}} : \tilde{\Omega} \to \mathbb{R}$ given by

$$K_{\tilde{\Omega}}(x,y) = \begin{cases} K_{1i}(x,y) & \text{if } x, y \in \tilde{X}_1 \\ K_{i,n+1}(x,y) & \text{if } x, y \in \tilde{X}_2 \end{cases}$$

where K_{1i} and $K_{i,n+1}$ are the canonical completions of $K_\Omega|_{\tilde{\Omega}_1}$ and $K_\Omega|_{\tilde{\Omega}_2}$ where $\tilde{\Omega}_1 =$

$\cup_{j=1}^{i}(X_j \times X_j)$ and $\tilde{\Omega}_2 = \cup_{j=i+1}^{n+1}(X_j \times X_j)$. It is clear that $K_\star$ is the canonical completion of $K_{\tilde{\Omega}}$ from the observation

$$\|f\|_\star^2 = \sum_{j=1}^{n+1} \|f_{X_j}\|^2 - \sum_{j=1}^{n} \|f_{X_j \cap X_{j+1}}\|^2 = \|f_{\tilde{X}_1}\|^2 - \|f_{\tilde{X}_1 \cap \tilde{X}_2}\|^2 + \|f_{\tilde{X}_2}\|^2$$

where

$$\|f_{\tilde{X}_1}\|^2 = \sum_{j=1}^{i} \|f_{X_j}\|^2 - \sum_{j=1}^{i-1} \|f_{X_j \cap X_{j+1}}\|^2$$
$$\|f_{\tilde{X}_2}\|^2 = \sum_{j=i+1}^{n+1} \|f_{X_j}\|^2 - \sum_{j=i+1}^{n} \|f_{X_j \cap X_{j+1}}\|^2$$

and $\|f_{\tilde{X}_1 \cap \tilde{X}_2}\|^2 = \|f_{X_i \cap X_{i+1}}\|^2$ are the quadratic forms associated with the reproducing kernels K_{1i}, $K_{i,n+1}$ and $K_{X_i \cap X_{i+1}}$. Notice that S can now be thought of as a separator of $x, y \in X$ in the 2-serrated domain $\tilde{\Omega}$. By Theorem 2.19, we conclude that

$$K_\star(x, y) = \langle k^\star_{x,S}, k^\star_{y,S} \rangle,$$

where $k^\star_{u,U} : U \to \mathbb{R}$ for $u \in X$ and $U \subset X$ is given by $k^\star_{u,U}(v) = K_\star(u, v)$. Hence proved. □

We can also interpret $K_\star$ as the result of sequence of multiple minimum norm interpolations of the kind that appeared in the study of 2-serrated domains. And similar to 2-serrated domains, the expression for the norm can be evaluated purely in terms of norms of restrictions of the restrictiosn K_{X_j} and $K_{X_j \cap X_{j+1}}$ of K_Ω. The following result can be proved in the same way as Theorem 2.18.

Theorem 2.21 (Projections)**.** *Let K be a completion of a partially reproducing kernel K_Ω on an n-serrated domain Ω. The followings statements are equivalent.*

1. $K = K_\star$,
2. $\mathbf{I} = \sum_{j=i}^{n} \Pi_{X_j} - \sum_{j=i}^{n-1} \Pi_{X_j \cap X_{j+1}}$, *and*
3. *for* $1 \leq j < n$, $A_j = \cup_{k=1}^{j} X_k$ *and* $B_j = \cup_{k=j+1}^{n} X_k$, $\Pi_{X_j \cap X_{j+1}} = \Pi_{A_j} \Pi_{B_j} = \Pi_{B_j} \Pi_{A_j}$

Theorem 2.22. *If $K = K_\star$ and $S_1, S_2 \subset X$ such that $S_1 \cap S_2$ separates $S_1 \setminus S_2$ and $S_2 \setminus S_1$, then*

$$\Pi_{S_1 \cup S_2} = \Pi_{S_1} + \Pi_{S_2} - \Pi_{S_1 \cap S_2} \tag{2.31}$$
$$\Pi_{S_1 \cap S_2} = \Pi_{S_1} \Pi_{S_2} = \Pi_{S_2} \Pi_{S_1}. \tag{2.32}$$

Property (2.32) is somewhat reminiscent of projection valued measures.

Example 2.1 (Matrices with Banded Inverses)**.** *Let $X = \{1, \ldots, m\}$ and $\Omega = \cup_{j=1}^{n}(X_j \times X_j)$ be a serrated domain on X. If K_Ω be a partially reproducing kernel on Ω, we can*

think of it as a matrix $\mathbf{A}_\Omega = [A_{ij}]_{i,j=1}^m$ *where* A_{ij} *is unspecified for* $(i,j) \notin \Omega$. *The inner product associated with a reproducing kernel* K *on finite* X *is given by* $(\mathbf{f}, \mathbf{g}) \mapsto \langle \mathbf{A}^\dagger \mathbf{f}, \mathbf{g} \rangle$ *where* $\mathbf{A}$ *is the kernel* K *in matrix form and* $\mathbf{A}^\dagger$ *denotes the pseudoinverse of* $\mathbf{A}$. *Using this fact and the form we have derived for the inner product associated with the canonical completion, we can write the canonical completion of* K_Ω *in a closed matrix form as*

$$\mathbf{A}^\dagger = \mathbf{A}_1^\dagger - \mathbf{A}_{12}^\dagger + \mathbf{A}_2^\dagger - \cdots - \mathbf{A}_{n-1,n}^\dagger + \mathbf{A}_n^\dagger$$

where $\mathbf{A}_k$ *is the* $m \times m$ *matrix with the* (i,j)*th entry* A_{ij} *for* $i, j \in X_k$ *and* 0 *otherwise, and* $\mathbf{A}_{k,k+1}$ *is the* $m \times m$ *matrix with the* (i,j)*th entry* A_{ij} *for* $i, j \in X_k \cap X_{k+1}$ *and* 0 *otherwise.*

2.5.4 Canonical Completion for Junction Tree Domains

The results for the serrated domains on an interval can be extended to a more general setting where the domains are tree-like in a certain sense. A *tree* is an undirected graph in which there exists a unique path connecting every two vertices. As with other graphs, a tree $\mathcal{T}$ on the set $\{1, \ldots, n\}$ can be treated as a subset of $\{1, \ldots, n\} \times \{1, \ldots, n\}$. We say that Ω can be represented as a *junction tree* if there exists for some $n \geq 1$, a tree $\mathcal{T}$ on $\{1, \ldots, n\}$ and subsets $\{X_j\}_{j=1}^n$ of X such that (a) $X = \cup_j X_j$, (b) $\Omega = \cup_j (X_j \times X_j)$, and (c) for every $(i,j), (j,k) \in \mathcal{T}$, we have $X_i \cap X_k \subset X_j$ (see Figure 2.5). Essentially, this means that the graph Ω admits a tree decomposition into cliques (see Diestel, 2010).

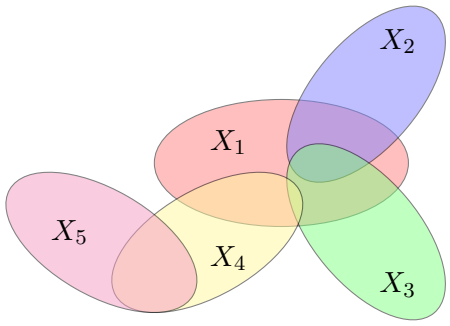


Figure 2.5: A junction tree domain Ω visualized as a graph on $X = \cup_{j=1}^5 X_j$ with the edges given by interpreting the regions X_j as cliques. The tree $\mathcal{T}$ is given by $\{(1,2), (2,1), (1,3), (3,1), (1,4), (4,1), (4,5), (5,4)\}$.

Notice that if we apply the completion formula (2.18) to X_i and X_j for two adjacent vertices i, j of $\mathcal{T}$, then we get a partially specified reproducing kernel over a larger domain with a simpler junction tree representation of $n-1$ vertices since X_i and X_j get replaced by $X_i \cup X_j$. Iterating the procedure, results in a completion of K_Ω.

Theorem 2.23. *Suppose that* Ω *admits a juction tree representation for some* $n \geq 1$, *a tree* $\mathcal{T}$ *on* $\{1, \ldots, n\}$ *and subsets* $\{X_j\}_{j=1}^n$ *of* X. *Then the following statements apply.*

1. *K_Ω admits a canonical completion.*

2. *The norm $\|\cdot\|_\star$ associated with the canonical completion $K_\star$ of K_Ω can be expressed as*
$$\|f\|_\star^2 = \sum_j \|f_{X_j}\|^2 - \sum_{(i,j)\in\mathcal{T}} \|f_{X_i\cap X_j}\|^2$$
where $\|f_U\|$ for $U \subset X$ denotes the norm of f_U in $\mathcal{H}(K_U)$.

3. *If $x \in X_i$ and $y \in X_j$ for some $1 \le i, j \le n$, then*
$$K_\star(x,y) = \langle [\Phi_{ji_k}\Phi_{i_k i_{k-1}}\cdots\Phi_{i_1 i}]\, k_{x,X_i}, k_{y,X_j}\rangle$$
where $(i,i_1),(i_1,i_2),\ldots,(i_k,j) \in \mathcal{T}$ is the unique path from i to j and for two adjacent vertices p and q, the mapping $\Phi_{pq} : \mathcal{H}_q \to \mathcal{H}_p$ is given by $\Phi_{pq} = \Phi_{X_p, X_p\cap X_q}\Phi_{X_p\cap X_q, X_q}$.

The proof of Theorem 2.23 is very similar to that of Theorem 2.20 and is hence omitted. Almost all of the following results in this chapter which apply to serrated domains can be easily generalized to junction tree domains, although we shall refrain from doing so for the sake of simplicity.

2.5.5 Dual and Variational Characterization

Theorem 2.20 provides an iterative procedure for calculating $K_\star$ for a partially reproducing kernel on K_Ω on a serrated domain K_Ω. We now provide a more direct and elegant characterization of the canonical completion which relies on the simple form of its associated norm. The key idea is that the quadratic form associated with a reproducing kernel and the square of its associated norm are, in a certain sense, convex conjugates of each other.

Theorem 2.24 (Duality Relations)**.** *Let K be a reproducing kernel on X and with the associated Hilbert space $\mathcal{H} = \mathcal{H}(K)$ equipped with the norm $\|\cdot\|$. For every $n \ge 1$, $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$ and $\{x_i\}_{i=1}^n \subset X$,*
$$\frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j K(x_i,x_j) = \max_f \left[\sum_{i=1}^n \alpha_i f(x_i) - \frac{1}{2}\|f\|^2\right] \tag{2.33}$$
where the maximum is taken over functions $f : X \to \mathbb{R}$. Furthermore, for every $f \in \mathcal{H}$,
$$\frac{1}{2}\|f\|^2 = \sup_{\boldsymbol{\alpha},\mathbf{x}} \left[\sum_{i=1}^n \alpha_i f(x_i) - \frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j K(x_i,x_j)\right] \tag{2.34}$$
where the supremum is taken over $n \ge 1$, $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^n \subset \mathbb{R}$ and $\mathbf{x} = \{x_i\}_{i=1}^n \subset X$.

Proof. Let $g = \sum_{i=1}^{n} \alpha_i k_{x_i}$. Then $\sum_{i=1}^{n} \alpha_i f(x_i) = \langle f, g \rangle$ and we can write the right-hand side of Equation 2.33 as

$$\begin{aligned}\max_f \left[\langle f, g\rangle - \tfrac{1}{2}\langle f, f\rangle\right] &= \tfrac{1}{2} \max_f \left[\|g\|^2 - \|f - g\|^2\right] \\ &= \tfrac{1}{2}\|g\|^2 - \tfrac{1}{2} \min_f \left[\|f - g\|^2\right] \\ &= \tfrac{1}{2}\|g\|^2\end{aligned}$$

which is equal to $\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j)$. Similarly, we can simplify the right-hand side of Equation 2.34 as

$$\sup_{\boldsymbol{\alpha},\mathbf{x}} \left[\langle f, g\rangle - \tfrac{1}{2}\|g\|^2\right] = \tfrac{1}{2}\|f\|^2 - \tfrac{1}{2} \inf_{\boldsymbol{\alpha},\mathbf{x}} \left[\|f - g\|^2\right] = \tfrac{1}{2}\|f\|^2$$

since the set of linear combinations $g = \sum_{i=1}^{n} \alpha_i k_{x_i}$ is dense in $\mathcal{H}$. Hence proved. □

Given a space of functions and an inner product which under which point evaluations are continuous, Equation (2.33) reduces the problem of calculating the corresponding reproducing kernel to a calculus of variations problem, thus providing a direct method for calculating the kernel. Indeed, once we calculate the quadratic form $q(\alpha, \beta) = \frac{1}{2}\left[\alpha^2 K(x, x) + 2\alpha\beta K(x, y) + \beta^2 K(y, y)\right]$ using

$$q(\alpha, \beta) = \max_f \left[\alpha f(x) + \beta f(y) - \tfrac{1}{2}\|f\|^2\right]$$

we have $K(x, y) = \frac{1}{4}[q(1, 1) - q(1, -1)]$. Using this observation, the duality relation (2.33) can be used to derive a very elegant formula for calculating the canonical completion.

Corollary 2.4. *Let $K_\star$ be the canonical completion of a partially reproducing kernel K_Ω on a serrated (or junction tree) domain Ω. Then,*

$$K_\star(x, y) = -\tfrac{1}{2}\left[K_\Omega(x, x) + K_\Omega(y, y)\right] + \max_f \left[f(x) + f(y) - \tfrac{1}{2}\|f\|_\Omega^2\right] \tag{2.35}$$

for $x, y \in X$.

The connection to convex analysis also makes obvious many fundamental results in the theory of reproducing kernels. Take for example, the fact that the norm associated with the sum $K_1 + K_2$ of two reproducing kernels K_1 and K_2, can be expressed as

$$\|f\|^2 = \inf_h \left[\|f - h\|_1^2 + \|h\|_2^2\right]$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ are the norms associated with K_1 and K_2. In light of the previous observation, this can be seen as an corollary to the fact that for two convex functionals F and G, the sum of their convex conjugates $F^* + G^*$ is equal to the convex conjugate of the infimal convolution $F \square G$ of F and G given by $(F \square G)(x) = \inf_y [F(x - y) + G(y)]$.

Indeed, let $g_i = \sum_{j=1}^n \alpha_j k_{x_j,i}$ for $n \geq 1$, $\{\alpha_j\}_{j=1}^n \subset \mathbb{R}$ and $\{x_j\}_{j=1}^n \subset X$, where $k_{x,i}(y) = K_i(x,y)$ for $x, y \in X$. The convex conjugate of $\frac{1}{2}\|f\|^2$ can then be written as

$$\begin{aligned}
\max_f \left[\sum_{j=1}^n \alpha_j f(x_j) - \tfrac{1}{2}\|f\|^2\right] &= \sup_{f,h}\left[\langle f-h, g_1\rangle_1 + \langle h, g_2\rangle_2 - \frac{1}{2}\left[\|f-h\|_1^2 + \|h\|_2^2\right]\right] \\
&= \tfrac{1}{2}\sup_{f,h}\left[\|g_1\|_1^2 + \|g_2\|_2^2 - \|f-h-g_1\|_1^2 - \|h-g_2\|_2^2\right] \\
&= \tfrac{1}{2}\left[\|g_1\|_1^2 + \|g_2\|_2^2\right] \\
&= \tfrac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j[K_1(x_i,x_j) + K_2(x_i,x_j)].
\end{aligned}$$

which implies that the reproducing kernel corresponding to the norm $\|\cdot\|$ is indeed K_1+K_2. Similarly, we can show that the norm corresponding to the subkernel $K_A = K|_{A\times A}$ for some kernel K on X and $A \subset X$ is $\|f\| = \inf_h\{\|h\| : h \in \mathcal{H}(K) \text{ and } h|_A = f\}$.

Theorem 2.25 (Variational Characterization)**.** *Let K be a reproducing kernel on X and with the associated Hilbert space $\mathcal{H} = \mathcal{H}(K)$ equipped with the norm $\|\cdot\|$. Then*

$$k_x = \arg\min_f \left[\|f\|^2 - 2f(x)\right]$$

where the minimum is taken over all functions $f : X \to \mathbb{R}$. Additionally,

$$k_x = \arg\min\left\{\|f\| \;\middle|\; f : X \to \mathbb{R} \text{ such that } f(x) = K(x,x)\right\}.$$

Proof. Clearly, $\|f\|^2 - 2f(x) = \|f - k_x\|^2 - K(x,x)$ is minimized for $f = k_x$. Hence proved. □

Corollary 2.5. *Let $K_\star$ be the canonical completion of a partially reproducing kernel K_Ω on a serrated (or junction tree) domain Ω. Then,*

$$\begin{aligned}
k_x^\star &= \arg\min_f \left[\|f\|_\Omega^2 - 2f(x)\right], \text{ and} \\
k_x^\star &= \arg\min\left\{\|f\|_\Omega \;\middle|\; f : X \to \mathbb{R} \text{ such that } f(x) = K_\Omega(x,x)\right\}
\end{aligned}$$

for $x \in X$ where $k_x^\star(y) = K_\star(x,y)$ for $y \in X$.

Proof. The first characterization is an immediate corollary of Theorem 2.25, while the second one follows from the observation that for a reproducing kernel K on X and $x \in X$, the generator k_x minimizes the associated norm among all functions $f \in \mathcal{H}(K)$ satifying $f(x) = K(x,x)$. □

We conclude this section by demonstrating how Theorem 2.25 can be used to recover the reproducing kernel of a Hilbert space of functions from the norm.

Example 2.2 (Brownian Motion)**.** *Let $\mathcal{H}$ be the Sobolev space of functions $f : [0,1] \to \mathbb{R}$ such that $f(0) = 0$, f is absolutely continuous and $f' \in L^2[0,1]$ equipped with the inner product $\langle f, g \rangle = \int_0^1 f(u)g(u)\, du$. We would like to calculate the corresponding reproducing kernel K. Pick $x \in [0,1]$ and by* Theorem 2.25*, we can write*

$$k_x = \arg\min_f \left[\int_0^1 [f'(u)]^2 \, du - 2f(x) \right].$$

Fix $f(x) = z$ for some $z \in \mathbb{R}$. Then the problem reduces to minimizing the integral $\int_0^1 [f'(u)]^2 \, du$ and the minimum occurs when f is given by the linear interpolation

$$f(u) = \begin{cases} [z/x]u & u \in [0, x] \\ z & u \in [x, 1] \end{cases}$$

when $x \neq 0$ and $f(u) = 0$ for $u \in [0,1]$ when $x = 0$. The problem reduces to finding the value of z which minimizes

$$\int_0^1 [f'(u)]^2 \, du - 2f(x) = \left[\tfrac{z}{x}\right]^2 x - 2z$$

which is when $z = x$! Thus, $k_x = f$ with $z = x$, or in other words, $K(x, y) = x \wedge y$, which is the Brownian motion covariance.

2.5.6 Vanishing Trace and Determinant Maximization

We are now in position to establish that the canonical completion is the reproducing kernel counterpart of the determinant maximizing completion whose inverse vanishes outside the specified region which appears in the classical theory of completions of partially specified matrices. The role of the matrix inverse and matrix determinant is played by the trace and the Fredholm determinant of an operator in the reproducing kernel Hilbert space.

Theorem 2.26. *Let $K_\star$ be the canonical completion of a partially reproducing kernel K_Ω on a serrated or junction tree domain Ω on X and $\Psi : \mathcal{H}(K_\star) \to \mathcal{H}(K_\star)$ be a trace-class operator. If $\langle \Psi k_x^\star, k_y^\star \rangle = 0$ for $(x, y) \in \Omega$, then*

1. $\operatorname{tr} \Psi = 0$*, and*

2. $\det(\mathbf{I} + \Psi) \leq 0$ *with equality if and only if $\Psi = \mathbf{0}$.*

Proof. By Theorem 2.21, we can write $\operatorname{tr}\Psi$ for a serrated domain Ω as

$$\begin{aligned}\operatorname{tr}(\mathbf{I}\Psi) &= \textstyle\sum_{j=1}^{n}\operatorname{tr}(\Pi_{X_i}\Psi) - \sum_{j=1}^{n-1}\operatorname{tr}(\Pi_{X_i\cap X_{i+1}}\Psi)\\ &= \textstyle\sum_{j=1}^{n}\operatorname{tr}(\Pi_{X_i}\Psi\Pi_{X_i}) - \sum_{j=1}^{n-1}\operatorname{tr}(\Pi_{X_i\cap X_{i+1}}\Psi\Pi_{X_i\cap X_{i+1}}).\end{aligned}$$

Clearly, the operators $\Pi_{X_i}\Psi\Pi_{X_i}$ for $1 \leq i \leq n$ and $\Pi_{X_i\cap X_{i+1}}\Psi\Pi_{X_i\cap X_{i+1}}$ for $1 \leq i \leq n-1$ are both zero, since $\langle \Psi k^\star_x, k^\star_y\rangle = 0$ for $x, y \in X_i$. Therefore, $\operatorname{tr}\Psi = 0$.

The second conclusion follows from noticing that $\Psi \mapsto -\log\det(\mathbf{I}+\Psi)$ is a strictly convex function whose Gateux derivative vanishes at $\mathbf{0}$ because

$$\frac{d}{dt}\left[\log\det(\mathbf{I}+t\Psi)\right]\Big|_{t=0} = \operatorname{tr}\Psi$$

and that the operators Ψ form a linear subspace. The proof for junction tree domains is analogous. □

To see the analogy with the matrix setting, it helps to write down the trace $\operatorname{tr}\Psi$ in terms of matrix trace assuming that X is finite. For $X = \{x_i\}_{i=1}^n$, the statement $\operatorname{tr}\Psi = 0$ (Theorem 2.26 (1)) can be expressed as

$$\operatorname{tr}\mathbf{K}^{-1}\mathbf{P} = \textstyle\sum_{i,j}(\mathbf{K}^{-1})_{ij}\mathbf{P}_{ij} = 0$$

if $\mathbf{P}_{ij} = 0$ for $(x_i, x_j) \in \Omega$, where $\mathbf{K} = [K_\star(x_i, x_j)]_{i,j=1}^n$ and $\mathbf{P} = [\mathbf{P}_{ij}]_{i,j=1}^n$ with $\mathbf{P}_{ij} = \langle \Psi k_{x_i}, k_{x_j}\rangle$. This implies that $(\mathbf{K}^{-1})_{ij} = 0$ for $(x_i, x_j) \notin \Omega$. To make sense of Theorem 2.26 (2) in the same way, we need the following lemma.

Lemma 2.3. *Let K be a strictly positive reproducing kernel and $\Psi : \mathcal{H}(K) \to \mathcal{H}(K)$ be a trace-class operator. Consider the nets $\{\mathbf{K}_\mathcal{F}\}_\mathcal{F}$ and $\{\mathbf{H}_\mathcal{F}\}_\mathcal{F}$ of matrices indexed by finite subsets $\mathcal{F}$ of X ordered by inclusion where*

$$\mathbf{K}_\mathcal{F} = [K(x,y)]_{x,y\in\mathcal{F}} \quad \text{and} \quad \mathbf{H}_\mathcal{F} = [\langle \Psi k_x, k_y\rangle]_{x,y\in\mathcal{F}} \quad \text{for } j \geq 1.$$

Then $\lim_\mathcal{F}\left[\log\det(\mathbf{K}_\mathcal{F}+\mathbf{H}_\mathcal{F}) - \log\det(\mathbf{K}_\mathcal{F})\right] = \log\det(\mathbf{I}+\Psi)$.

Proof. We have by Grümm's Convergence Theorem that $\Pi_\mathcal{F}\Psi\Pi_\mathcal{F} \to \Psi$ in trace norm (see Theorem 2.19 of Simon (2005) and also, Theorem 3.8 in Simon (1977)) because $\Pi_\mathcal{F}$ strongly converges to $\mathbf{I}$ (Proposition 3.9 of Paulsen and Raghupathi (2016)). Therefore, $\det(\mathbf{I}+\Pi_\mathcal{F}\Psi\Pi_\mathcal{F}) \to \det(\mathbf{I}+\Psi)$. It suffices to show that $\log\det(\mathbf{I}+\Pi_\mathcal{F}\Psi\Pi_\mathcal{F}) = \log\det(\mathbf{K}_\mathcal{F}+\mathbf{H}_\mathcal{F}) - \log\det(\mathbf{K}_\mathcal{F})$.

For finite rank Ψ, we can write $\Psi = \sum_{i=1}^r f_i \otimes g_i$ for $r \geq 1$ and $\{f_i\}_{i=1}^r, \{g_i\}_{i=1}^r \subset \mathcal{H}(K)$.

So,

$$\log\det(\mathbf{I}+\Pi_{\mathcal{F}}\Psi\Pi_{\mathcal{F}}) = \log\det\left(\mathbf{I}+\sum_{i=1}^{r}\Pi_{\mathcal{F}}f_i\otimes\Pi_{\mathcal{F}}g_i\right).$$

By Plemelj's formula,

$$\log\det\left(\mathbf{I}+\sum_{i=1}^{r}\Pi_{\mathcal{F}}f_i\otimes\Pi_{\mathcal{F}}g_i\right) = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\operatorname{tr}\left\{\left[\sum_{i=1}^{r}\Pi_{\mathcal{F}}f_i\otimes\Pi_{\mathcal{F}}g_i\right]^n\right\} \tag{2.36}$$

where the trace terms can be written as sums of products of the inner products $\langle\Pi_{\mathcal{F}}f_i,\Pi_{\mathcal{F}}g_k\rangle$ which can be simplified as

$$\langle\Pi_{\mathcal{F}}f_i,\Pi_{\mathcal{F}}g_k\rangle = \langle f_i|_{\mathcal{F}},g_k|_{\mathcal{F}}\rangle = \mathbf{f}_i^{\top}\mathbf{K}_{\mathcal{F}}^{-1}\mathbf{g}_k = (\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{f}_i)^{\top}\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{g}_k,$$

where $\mathbf{f}_i=[f_i(x)]_{x\in\mathcal{F}}$ and $\mathbf{g}_j=[g_k(x)]_{x\in\mathcal{F}}$ are to be thought of as column vectors. By working our way backwards with the matrix form, we can rewrite (2.36) as

$$\begin{aligned}
\log\det\left(\mathbf{I}+\sum_{i=1}^{r}\Pi_{\mathcal{F}}f_i\otimes\Pi_{\mathcal{F}}g_i\right) &= \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\operatorname{tr}\left\{\left[\sum_{i=1}^{r}(\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{f}_i)\left(\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{g}_i\right)^{\top}\right]^n\right\}\\
&= \log\det\left(\mathbf{I}+\sum_{i=1}^{r}\left(\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{f}_i\right)\left(\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{g}_i\right)^{\top}\right)\\
&= \log\det\left(\mathbf{I}+\mathbf{K}_{\mathcal{F}}^{-1/2}\left[\sum_{i=1}^{r}\mathbf{f}_i\mathbf{g}_i^{\top}\right]\mathbf{K}_{\mathcal{F}}^{-1/2}\right)\\
&= \log\det\left(\mathbf{I}+\mathbf{K}_{\mathcal{F}}^{-1/2}\mathbf{H}_{\mathcal{F}}\mathbf{K}_{\mathcal{F}}^{-1/2}\right)\\
&= \log\det(\mathbf{K}_{\mathcal{F}}+\mathbf{H}_{\mathcal{F}})-\log\det(\mathbf{K}_{\mathcal{F}}).
\end{aligned}$$

Even if Ψ is not finite rank, we can approximate by finite rank operators in trace norm and the conclusion follows from the continuity of the Fredholm determinant in trace norm. □

Roughly speaking, Theorem 2.26 (2) seems to say that trace-class self-adjoint perturbations of the canonical solution tend to decrease the determinant.

2.5.7 Canonical Completion for Regular Domains

In this section, we shall study the problem of canonical completion for a different class of domains which can be thought of as the limit of a sequence of serrated domain.

Definition 2.4 (Regular Domain)**.** *Let $X=[0,1]\subset\mathbb{R}$. We say that a domain Ω on X is a* regular domain *if we can write*

$$\Omega = \cup_{t\in T}(I_t\times I_t)$$

where $T = [0, t_0]$ *for some* $t_0 \in (0, 1)$ *and* $I_t = [t, b(t)]$ *for a strictly increasing function* $b : [0, w] \to \mathbb{R}$ *satisfying* $b(t) > t$ *for* $t \in T$ *and* $b(w) = 1$.

Regular domains are particularly nice in that I_t for $t \in (0, t_0)$ are all minimal separators of Ω. Note that by appropriately rescaling X, we can make any regular domain Ω in to the *band* $\{(x, y) : |x - y| \leq a\}$ for some $a > 0$. Thus regular domains are domains which are equivalent to the band.

We shall prove the existence of a canonical completion $K_\star$ of every partially reproducing kernel K_Ω on a regular domain Ω. Roughly speaking, our proof relies on approximating Ω with a sequence of serrated domains $\Omega_j \subset \Omega$ and a canonical completion $K_\star$ as a limit of the canonical completions K_j of the partially reproducing kernels $K_{\Omega_j} = K_\Omega|_{\Omega_j}$ on the serrated domains Ω_j. In spite of their simple definition, the norms $\|\cdot\|_{\Omega_j}$ we do not have much insight into their limiting behaviour as $j \to \infty$. We circumvent this problem by relying on sequential compactness properties of the sequence $\|\cdot\|_{\Omega_j}$ under a special notion of convergence known as Γ-convergence or epiconvergence (Dal Maso, 1993; Braides, 2002).

The opaque nature of our construction makes it difficult to show that the canonical completion is unique, although our experience with serrated domains is a compelling reason to believe that this is certainly the case. Regardless, we are still able to derive an interesting algebraic characterization of canonical completion in terms of semigroupoids of contraction maps. In the next section, we establish the uniqueness of canonical completion for some stationary partially reproducing kernels.

Γ-Convergence in Separable Hilbert Spaces

As a result, we are forced to rely on general properties of the norm, such as the fact that $\|f\|^2$ is a quadratic form on the space of functions $f : X \to \mathbb{R}$. Let $\bar{\mathbb{R}}$ denote $\mathbb{R} \cup \{\infty\}$.

Definition 2.5 (Γ-Convergence in Hilbert Space)**.** *Let* $\mathcal{X}$ *be a Hilbert space. We say that a sequence* $\{\Lambda_j\}_{j=1}^\infty$ *of functionals* $\Lambda_j : \mathcal{X} \to \bar{\mathbb{R}}$ *converges in the* Γ *sense or simply,* Γ*-converges to* $\Lambda : \mathcal{X} \to \bar{\mathbb{R}}$ *if for every* $f \in \mathcal{X}$ *we have:*

1. *for every* $\{f_j\}_{j=1}^\infty \subset \mathcal{X}$, $\lim_{j\to\infty} f_j = f$ *implies* $\Lambda(f) \leq \liminf_j \Lambda_j(f_j)$, *and*
2. *there exists* $\{f_j\}_{j=1}^\infty \subset \mathcal{X}$ *such that* $\lim_{j\to\infty} f_j = f$ *and* $\Lambda(f) \geq \limsup_j \Lambda_j(f_j)$.

If Λ_j Γ*-converges to* Λ, *we write* $\Gamma\text{-}\lim_j \Lambda_j = \Lambda$.

Γ-convergence is quite different from other modes of convergence with which the reader may be familiar. Perhaps most strikingly, even the limit $\Gamma\text{-}\lim_j \Lambda_j$ of a constant sequence

$\Lambda_j = \Lambda$ is not necessarily equal to Λ unless Λ is *lower semicontinuous*. A functional $\Lambda : \mathcal{X} \to \mathbb{R}$ is said to be lower semicontinuous if for every $f \in \mathcal{X}$ and $\{f_j\}_{j=1}^{\infty} \subset \mathcal{X}$ such that $f_j \to f$, we have $\liminf_j \Lambda(f_j) \geq \Lambda(f)$ (Remark 1.8 in Braides (2002)). Note that Γ-limits themselves are always lower semicontinuous (Proposition 6.8 of Dal Maso (1993)).

Somewhat unsurprisingly, the limit $\Gamma\text{-}\lim_j \Lambda_j$ is not necessarily equal to the pointwise limit $f \mapsto \lim_j \Lambda_j(f)$. Thankfully, many intuitive properties continue to hold, if only under certain conditions. Let $\{\Lambda_j\}_{j=1}^{\infty}$ be sequence of functionals on $\mathcal{X}$ such that $\Gamma\text{-}\lim_j \Lambda_j = \Lambda$. For a continuous and increasing function $\varphi : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ and then $\Gamma\text{-}\lim_j \varphi \circ \Lambda_j = \varphi \circ \Lambda$ (Proposition 6.16 of Dal Maso (1993)). We also have *monotonicity*. Let $\{\tilde{\Lambda}_j\}_{j=1}^{\infty}$ be another sequence of functionals $\tilde{\Lambda}_j : \mathcal{X} \to \bar{\mathbb{R}}$ such that $\Gamma\text{-}\lim_j \tilde{\Lambda}_j = \tilde{\Lambda}$. If $\Lambda_j(f) \leq \tilde{\Lambda}_j(f)$ for $f \in \mathcal{X}$ and $j \geq 1$, then $\Lambda \leq \tilde{\Lambda}$ (Proposition 6.7 of Dal Maso (1993)). Naturally, this means that if $\bar{\Lambda}(f) \leq \tilde{\Lambda}_j(f)$ for $f \in \mathcal{X}$ and $j \geq 1$ for some lower semicontinuous functional $\bar{\Lambda} : \mathcal{X} \to \bar{\mathbb{R}}$, then $\bar{\Lambda} \leq \tilde{\Lambda}$. Furthermore, Γ-limits are *superadditive*, in that

$$\Lambda + \tilde{\Lambda} \leq \Gamma\text{-}\lim_j \left[\Lambda_j + \tilde{\Lambda}_j\right]$$

so long as the limit $\Gamma\text{-}\lim_j \left[\Lambda_j + \tilde{\Lambda}_j\right]$ exists and the sums $\Lambda_j + \tilde{\Lambda}_j$ and $\Lambda + \tilde{\Lambda}$ are well-defined, in the sense that for no point in $\mathcal{X}$ is one of the functionals in the sum equal to ∞ when the other is $-\infty$.

Interestingly, Γ-convergence is sequentially compact on second-countable spaces that is, every sequence $\{\Lambda_j\}_{j=1}^{\infty}$ of functionals $\Lambda_j : \mathcal{X} \to \bar{\mathbb{R}}$ has a Γ-convergent subsequence (Theorem 8.5 of Dal Maso (1993)). Note that because separability and second-countability are equivalent for metric spaces, this also holds true for our setting of separable Hilbert spaces. We shall use this property to construct our canonical completion from a sequence of canonical completions on serrated domains. To this end, we shall need another importantly property of Γ-convergence, which is that the Γ-limit of non-negative quadratic forms is a non-negative quadratic form (Theorem 11.10 of Dal Maso (1993)).

Existence of Canonical Completion on Regular Domains

In this section, we shall establish the existence of the canonical completion for partially reproducing kernel on regular domains.

Theorem 2.27. *Let K_Ω be a partially reproducing kernel on a regular domain Ω. Then K_Ω admits a canonical completion.*

Let K_Ω be a continuous partially reproducing kernel on a regular domain Ω on $[0, a]$ with $\Omega = \cup_{t \in T}(I_t \times I_t)$ where $T = [0, t_0]$ as in Definition 2.4. Define an increasing sequence $\{\Omega_j\}_{j=1}^{\infty}$ of serrated domains on $[0, a]$ by $\Omega_j = \cup_{t \in T_j}(I_t \times I_t)$ where $T_j = \{\frac{i}{2^j} t_0 : 0 \leq i \leq 2^j\}$. To retain the serrated domain notation, we denote $X_{ij} = I_{t_{ij}}$ where $t_{ij} = \frac{i-1}{2^j} t_0$, and now

we can write $\Omega_j = \cup_{i=1}^{m_j}(X_{ij} \times X_{ij})$ where $m_j = 2^j + 1$.

For $j \geq 1$, let K_j denote the canonical completion of the restriction $K_{\Omega_j} = K_\Omega|_{\Omega_j}$ with the associated Hilbert space $\mathcal{H}_j = \mathcal{H}(K_j)$. Define the quadratic forms Λ_j as the squares of the associated norms $\|f\|_j$ of $\mathcal{H}_j$:

$$\Lambda_j(f) = \|f\|_j^2 = \sum_{i=1}^{m_j} \|f_{X_{ij}}\|^2 - \sum_{i=1}^{m_j-1} \|f_{X_{ij}\cap X_{i+1,j}}\|^2$$

where $\|f\|_j$ is the norm associated with $\mathcal{H}_j$.

Let $\mathcal{X} = \mathcal{H}_1$ and $\|\cdot\| = \|\cdot\|_1$. We shall treat Λ_j as functionals on $\mathcal{X}$. Notice that $\{\Lambda_j\}_{j=1}^\infty$ is *equicoercive* in that $\frac{1}{m_1}\|f\|^2 \leq \Lambda_j(f)$ for every $f \in \mathcal{X}$ and $j \geq 1$, since

$$\tfrac{1}{m_1}\|f\|^2 \leq \tfrac{1}{m_1}\sum_{i=1}^{m_1} \|f_{X_{i1}}\|^2 \leq \tfrac{1}{m_1}\sum_{i=1}^{m_1} \|f\|_j^2 = \|f\|_j^2 = \Lambda_j(f) \tag{2.37}$$

as $X_{i1} \times X_{i1} \subset \Omega_j$ for every $1 \leq i \leq m_1$. Moreover, note that K_1 is continuous because K_{Ω_1} is continuous. This implies that $\mathcal{X} = \mathcal{H}(K_1)$ is a second-countable space, since continuous kernels induce separable Hilbert spaces and for Hilbert spaces, separability and second-countability are equivalent.

By the sequential compactness property of Γ-convergence (Dal Maso, 1993, Theorem 8.5), there exists a subsequence $\{j_k\}_{k=1}^\infty \subset \{j\}_{j=1}^\infty$ such that $\Gamma\text{-}\lim_k \Lambda_{j_k} = \Lambda$ for some lower-semicontinuous functional Λ. Because every Λ_{j_k} is a non-negative quadratic form so is Λ (by Theorem 11.10 of Dal Maso (1993)). Define $\mathcal{H}_\Lambda = \{f \in \mathcal{X} : \Lambda(f) < \infty\}$. By virtue of being a non-negative quadratic form, $\Lambda(f)$ defines an inner product $\langle\cdot,\cdot\rangle_\Lambda$ on $\mathcal{H}_\Lambda$ given by

$$\langle f, g\rangle_\Lambda = \tfrac{1}{4}[\Lambda(f+g) - \Lambda(f-g)].$$

The inner product induces the norm $\|f\|_\Lambda = \sqrt{\Lambda(f)}$ on $\mathcal{H}_\Lambda$.

Lemma 2.4. *The space $\mathcal{H}_\Lambda$ equipped with the inner product $\langle\cdot,\cdot\rangle_\Lambda$ is a reproducing kernel Hilbert space.*

Proof. $\mathcal{H}_\Lambda$ is clearly an inner product space. We need to show that $\mathcal{H}_\Lambda$ is complete with respect to the norm $\|\cdot\|_\Lambda$. Let $\{f_j\}_{j=1}^\infty \subset \mathcal{X}$ be a $\|\cdot\|_\Lambda$-Cauchy sequence. This implies that $\{\|f_j\|_\Lambda\}_{j=1}^\infty$ is bounded. Furthermore, because (2.37) we can conclude using Proposition 6.7 of Dal Maso (1993) that

$$\frac{1}{m_1}\|f\|^2 \leq \|f\|_\Lambda^2 = \Lambda(f)$$

implying that $\{f_j\}_{j=1}^\infty$ is also Cauchy with respect to $\|\cdot\|$ and therefore it must converge to some $f \in \mathcal{X}$. We conclude from the lower semicontinuity of Λ that $\Lambda(f) \leq \lim_{j\to\infty}\|f_j\|_\Lambda < \infty$ and thus, $f \in \mathcal{H}_\Lambda$. Notice that $f_j - f_i \to f_j - f$ in $\mathcal{X}$ as $i \to \infty$, so we can write again using the lower semicontinuity of Λ that $\|f_j - f\|_\Lambda \leq \lim_{i\to\infty}\|f_j - f_i\|_\Lambda$ for every $j \geq 1$.

Taking the limit as $j \to \infty$ gives

$$\lim_{j\to\infty} \|f_j - f\|_\Lambda \le \lim_{j,i\to\infty} \|f_j - f_i\|_\Lambda = 0.$$

Thus $f_j \to f$ in the norm $\|\cdot\|_\Lambda$ and $\mathcal{H}_\Lambda$ is a Hilbert space. To show that $\mathcal{H}_\Lambda$ is a reproducing kernel Hilbert space we need only observe that for every $f \in \mathcal{H}_{j_k} \subset \mathcal{X}$ and $x \in X$,

$$|f(x)| = |\langle f, k_{x,j_k}\rangle| \le \|f\|_{j_k}\|k_{x,j_k}\|_{j_k} = \sqrt{K_\Omega(x,x)} \cdot \|f\|_{j_k} = \sqrt{K_\Omega(x,x)\cdot\Lambda_{j_k}(f)}$$

by Cauchy-Schwarz inequality, where $k_{x,j_k}(y) = K_{j_k}(x,y)$ for $y \in X$. Proposition 6.7 of Dal Maso (1993) allows us to conclude that the same is true for Λ, and thus

$$|f(x)| \le \sqrt{K_\Omega(x,x)} \cdot \|f\|_\Lambda = \sqrt{K_\Omega(x,x)\cdot\Lambda(f)}$$

which implies that point evaluations are continuous in $\mathcal{H}_\Lambda$. Hence proved. □

Let K_Λ denote the reproducing kernel of $\mathcal{H}_\Lambda$.

Lemma 2.5. *K_Λ is a completion of K_Ω and $\lim_{k\to\infty} K_{j_k}(x,y) = K_\Lambda(x,y)$ for every $x, y \in X$.*

Proof. Pick $x, y \in X$. Proposition 6.21 of Dal Maso (1993) tells us that if G is a continuous functional on $\mathcal{X}$, then $\Gamma\text{-}\lim_{j\to\infty}\Lambda_j = \Lambda$ implies $\Gamma\text{-}\lim_{j\to\infty}\Lambda_j + G = \Lambda + G$. Notice that the point evaluations $f \mapsto f(u)$ for $u \in X$ are continuous linear functionals on $\mathcal{X}$. We deduce from $\Gamma\text{-}\lim_{k\to\infty}\Lambda_{j_k} = \Lambda$ that $\Gamma\text{-}\lim_{k\to\infty}(-\tilde{\Lambda}_{j_k}) = (-\tilde{\Lambda})$, where

$$\tilde{\Lambda}_{j_k}(f) = f(x) + f(y) - \frac{1}{2}\|f\|_{j_k}^2 \quad \text{and} \quad \tilde{\Lambda}(f) = f(x) + f(y) - \frac{1}{2}\|f\|_\Lambda^2.$$

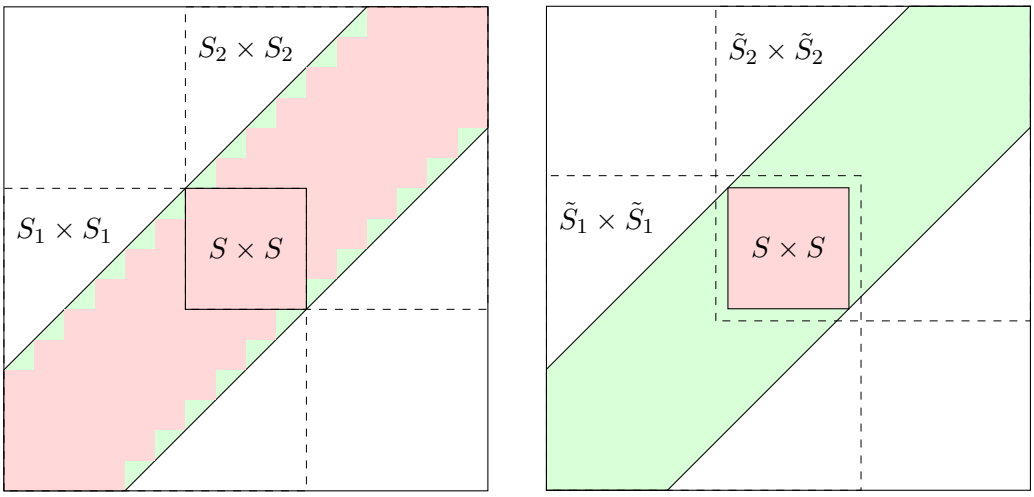
Because $\{\Lambda_{j_k}\}_{k=1}^\infty$ is equicoercive (2.37), we can conclude using Theorem 7.8 of Dal Maso (1993), that $\min(-\tilde{\Lambda}_{j_k}) \to \min(-\tilde{\Lambda})$ as $k \to \infty$ or alternatively, $\max\tilde{\Lambda}_{j_k} \to \max\tilde{\Lambda}$ as $k \to \infty$. By Theorem 2.24, we have

$$\tfrac{1}{2}\left[K_\Omega(x,x) + K_\Omega(y,y) + 2K_{j_k}(x,y)\right] \to \tfrac{1}{2}\left[K_\Omega(x,x) + K_\Omega(y,y) + 2K_\Lambda(x,y)\right]$$

or $K_{j_k}(x,y) \to K_\Lambda(x,y)$ as $k \to \infty$. Thus the kernel K_Λ of $\mathcal{H}_\Lambda$ is actually the pointwise limit of the kernels K_{j_k} of $\mathcal{H}_{j_k}$. Notice that K_Λ is a completion of K_Ω. Indeed, for (x,y) in the interior of Ω, $\lim_{k\to\infty} K_{j_k}(x,y) = K_\Omega(x,y)$ since $K_{j_k}(x,y) = K_\Omega(x,y)$ for large enough k. For $(x,y) \in \Omega$ which lie on the boundary of Ω, the same conclusion follows from the continuity of the kernels. □

Lemma 2.6. *K_Λ is a canonical completion of K_Ω.*

Proof. It suffices to show that the separation property is satisfied for minimal separators, for otherwise we can argue as in Theorem 2.20. Let $S = X_{pq}$. Then S is a minimal separator of Ω which separates Ω into two connected components $Y_1, Y_2 \subset X$. Define $S_1 = S \cup Y_1$ and $S_2 = S \cup Y_2$ (see Figure 2.6a). Note that $S \times S \subset \Omega_{j_k}$ for $j_k \geq q$. Consider the sequence $\{\Lambda_{j_k}\}$ for $j_k \geq q$.



(a) The red region is Ω_{j_k} for $j_k \geq q$. (b) $\tilde{S} \times \tilde{S}$ is the dashed square around $S \times S$.

Figure 2.6: Canonical Completion. The colored regions represent Ω.

The norm of the function f_{S_1} in $\mathcal{H}(K_{j_k}|_{S_1 \times S_1})$ can be expressed as

$$\|f_{S_1}\|^2_{j_k,S_1} = \sum_{i=1}^{p_{j_k}} \|f_{X_{ij_k}}\|^2 - \sum_{i=1}^{p_{j_k}-1} \|f_{X_{ij_k} \cap X_{i+1,j_k}}\|^2$$

where p_{j_k} is given by $X_{p_{j_k},j_j} = X_{pq}$, that is $p_{j_k} = 1 + (p-1)2^{j_k - q}$. Similarly, the norm of the function f_{S_2} in $\mathcal{H}(K_{j_k}|_{S_2 \times S_2})$ is

$$\|f_{S_2}\|^2_{j_k,S_2} = \sum_{i=p_{j_k}}^{m_{j_k}} \|f_{X_{ij_k}}\|^2 - \sum_{i=p_{j_k}}^{m_{j_k}-1} \|f_{X_{ij_k} \cap X_{i+1,j_k}}\|^2.$$

Because $\tilde{\Omega} = (S_1 \times S_1) \cup (S_2 \times S_2) \supset \Omega_{j_k}$, we can write using the inheritance property of canonical completion that

$$\|f\|^2_{j_k} = \|f_{S_1}\|^2_{j_k,S_1} + \|f_{S_2}\|^2_{j_k,S_2} - \|f_S\|^2_{j_k,S}$$

where $f_{S_1} = f|_{S_1}$, $f_{S_2} = f|_{S_2}$ and of course, $f_S = f|_S$. Notice that $S \times S \subset \Omega_{j_k}$, so $\|f_S\|_{j_k,S} = \|f_S\|_S$. Since $f \mapsto \|f_S\|^2_S$ is a continuous function on $\mathcal{X}$, we get by taking the

Γ-limit of both sides that

$$\begin{aligned}\|f\|_\Lambda^2 &= \Gamma\text{-lim}_{k\to\infty}\left[\|f_{S_1}\|_{j_k,S_1}^2 + \|f_{S_2}\|_{j_k,S_2}^2\right] - \|f_S\|_S^2 \\ &\geq \Gamma\text{-lim}_{k\to\infty}\|f_{S_1}\|_{j_k,S_1}^2 + \Gamma\text{-lim}_{k\to\infty}\|f_{S_2}\|_{j_k,S_2}^2 - \|f_S\|_S^2\end{aligned}$$

because we taking the limit of a sum with a continuous functional $-\|f_S\|_S^2$ (Dal Maso, 1993, Proposition 6.21) and superadditivity of the Γ-limits of $\|f_{S_1}\|_{j_k,S_1}^2$ and $\|f_{S_2}\|_{j_k,S_2}^2$ (Dal Maso, 1993, Proposition 6.7). Furthermore,

$$\Gamma\text{-lim}_{k\to\infty}\|f_{S_1}\|_{j_k,S_1}^2 = \|f_{S_1}\|_{\Lambda,S_1}^2 \quad \text{and} \quad \Gamma\text{-lim}_{k\to\infty}\|f_{S_2}\|_{j_k,S_2}^2 = \|f_{S_2}\|_{\Lambda,S_2}^2$$

where $\|f_{S_1}\|_{\Lambda,S_1}^2$ and $\|f_{S_2}\|_{\Lambda,S_2}^2$ denote the norms of f_{S_1} in $\mathcal{H}(K_\Lambda|_{S_1\times S_1})$ and f_{S_2} in $\mathcal{H}(K_\Lambda|_{S_2\times S_2})$ respectively. To see why, note that using the same arguments as before, we can show that every subsequence of $f_{S_1} \to \|f_{S_1}\|_{j_k,S_1}^2$ will admit a Γ-convergent subsequence which converges to the square of the norm of a completion of $K_\Omega|_{S_1\times S_1}$ to $S_1 \times S_1$ which is the pointwise limit of the corresponding subsequence of the completions $K_{j_k}|_{S_1\times S_1}$ corresponding to the norms $\|\cdot\|_{j_k,S_1}^2$. But $K_{j_k}|_{S_1\times S_1}$ converges pointwise to $K_\Lambda|_{S_1\times S_1}$ by Lemma 2.5, so it follows that all such subsequences of norms squared converge to $\|f_{S_1}\|_{\Lambda,S_1}^2$ implying that $\Gamma\text{-lim}_{k\to\infty}\|f_{S_1}\|_{j_k,S_1}^2 = \|f_{S_1}\|_{\Lambda,S_1}^2$ by the Urysohn property (Dal Maso, 1993, Proposition 8.3) of Γ-convergence which states that if every subsequence of a sequence Γ-converges to the same limit then the sequence Γ-converges to that limit. Similarly, we can show that same for $\|f_{S_2}\|_{j_k,S_2}^2$. It follows that

$$\|f\|_\Lambda^2 \geq \|f_{S_1}\|_{\Lambda,S_1}^2 + \|f_{S_2}\|_{\Lambda,S_2}^2 - \|f_S\|_{\Lambda,S}^2.$$

The expression of the right hand side is actually the squared norm of the canonical completion $\tilde{K}$ of the restriction of K_Λ to $\tilde{\Omega} = S_1^2 \cup S_2^2$. Taking the convex conjugates of the two sides gives for every $n \geq 1$, $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$ and $\{x_i\}_{i=1}^n \subset X$ that

$$\sum_{i,j=1}^n \alpha_i\alpha_j K_\Lambda(x_i,x_j) \leq \sum_{i,j=1}^n \alpha_i\alpha_j \tilde{K}(x_i,x_j)$$

which implies that the difference $\tilde{K} - K_\Lambda$ is a reproducing kernel. But $(\tilde{K} - K_\Lambda)(x,x) = 0$ for every $x \in X$, which implies that $(\tilde{K} - K_\Lambda)(x,y) = 0$ for every $x, y \in X$. Since $\tilde{K} = K_\Lambda$, we have

$$\|f\|_\Lambda^2 = \|f_1\|_{\Lambda,S_1}^2 + \|f_2\|_{\Lambda,S_2}^2 - \|f_S\|_{\Lambda,S}^2.$$

This implies the separation property for the minimal separators of the form $S = X_{pq}$ and by inheritance, the separators which contains these minimal separators. To extend the result to all minimal separators, let S be a minimal separator. Then it is contained inside a separator $\tilde{S}$ which separates Ω into two connected components $\tilde{Y}_1, \tilde{Y}_2 \subset X$ which contains a minimal separator of the form X_{pq} (see Figure 2.6b). Let $\tilde{S}_1 = \tilde{S} \cup Y_1$ and $\tilde{S}_2 = \tilde{S} \cup Y_2$. We can write the separation property for $\tilde{S}$ in terms of the projections $\Pi_{\tilde{S}_1}$,

$\Pi_{\tilde{S}_2}$ and $\Pi_{\tilde{S}}$ in $\mathcal{H}_\Lambda = \mathcal{H}(K_\Lambda)$ as:

$$\mathbf{I} = \Pi_{\tilde{S}_1} + \Pi_{\tilde{S}_2} - \Pi_{\tilde{S}}.$$

Let $\tilde{S} \downarrow S$ in the sense of sets. The conclusion now follows from the continuity of K_Λ and the strong convergence of the projection $\Pi_{\tilde{S}}$ to the projection on to the intersection of the closed subspaces generated by $\{k_{x,\Lambda} : x \in \tilde{S}\}$ in $\mathcal{H}_\Lambda$ where $k_{x,\Lambda}(y) = K_\Lambda(x, y)$ for $y \in X$(and likewise for $\Pi_{\tilde{S}_1}$, $\Pi_{\tilde{S}_2}$). □

We have thus shown the following:

Theorem 2.28. *Let K_Ω be a partially reproducing kernel on a regular domain. Then K_Ω admits a canonical completion $K_\star$. Furthermore, there exists an increasing sequence $\{\Omega_j\}_{j=1}^\infty$ of serrated domains with $\Omega_j \subset \Omega$ and $\cup_{j=1}^\infty \Omega_j = \Omega$ such that the canonical completions K_j of $K_{\Omega_j} = K_\Omega|_{\Omega_j}$ with the associated norms $\|\cdot\|_j$ such that K_j converges pointwise to $K_\star$ as $j \to \infty$ and $\Gamma\text{-}\lim_{j\to\infty} \|\cdot\|_j = \|\cdot\|_\star$, where $\|\cdot\|_\star$ is norm associated with $K_\star$.*

Notice that for every serrated domain $\tilde{\Omega} \supset \Omega$, the canonical completion of $K_{\tilde{\Omega}} = K_\star|_{\tilde{\Omega}}$ is still $K_\star$ by inheritance. And as a result, the determinant maximization and trace zero properties hold for $K_\star$ as the completion of $K_{\tilde{\Omega}}$.

Although, we are unable to draw any conclusion about the uniqueness of canonical completion, using arguments similar to those in the proof of Lemma 2.5, we can show the following result.

Theorem 2.29. *Let K_Ω be a partially reproducing kernel on domain Ω such that there exists an increasing sequence of serrated domains $\{\Omega_j\}_{j=1}^\infty$ with $\Omega_j \subset \Omega$ and $\cup_{j=1}^\infty \Omega_j = \Omega$. Let K_j denote the canonical completions of $K_{\Omega_j} = K_\Omega|_{\Omega_j}$ with the associated norms $\|\cdot\|_j$.*

If (a) Ω is serrated or (b) Ω is regular and K_Ω admits a unique canonical completion, then K_j converges pointwise to $K_\star$ as $j \to \infty$ and $\Gamma\text{-}\lim_{j\to\infty} \|\cdot\|_j = \|\cdot\|_\star$, where $K_\star$ denotes the canonical completion and $\|\cdot\|_\star$ denotes its associated norm.

2.5.8 Canonical Semigroupoids

The contraction maps of a canonical completion $K_\star$ of K_Ω are remarkable in that they mimic the structure of the underlying graph Ω. Suppose that A, S and B be subsets of X. If S separates A and B, then every path from A to B can be decomposed into two paths: one from A to S, followed by another from S to B. Then the contraction $\Phi_{BA} : \mathcal{H}_A \to \mathcal{H}_B$ can be written as product of the contractions $\Phi_{SA} : \mathcal{H}_A \to \mathcal{H}_S$ and $\Phi_{BS} : \mathcal{H}_B \to \mathcal{H}_S$ as $\Phi_{BA} = \Phi_{BS}\Phi_{SA}$. Indeed, let $x \in A$ and $y \in B$. Naturally, $K_\star(x, y) = \langle \Phi_{BA} k_{x,A}, k_{y,B} \rangle$. The separation property gives us another way of writing

$K_\star(x,y)$ which is

$$\langle k_{x,S}, k_{y,S}\rangle = \langle \Phi_{SA}k_{x,A}, \Phi_{SB}k_{y,B}\rangle = \langle \Phi_{BS}\Phi_{SA}k_{x,A}, k_{y,B}\rangle.$$

Thus $\langle \Phi_{BA}k_{x,A}, k_{y,B}\rangle = \langle \Phi_{BS}\Phi_{SA}k_{x,A}, k_{y,B}\rangle$. Since x and y can be chosen arbitrarily, it follows that $\Phi_{BA} = \Phi_{BS}\Phi_{SA}$. It is not difficult to see that the converse is also true.

Lemma 2.7. *The contraction maps corresponding to a completion K of K_Ω satisfy for the subsets A, B and S of X*

$$\Phi_{BA} = \Phi_{BS}\Phi_{SA} \quad \text{if } S \text{ separates } A \text{ and } B,$$

if and only if K is a canonical completion of K_Ω.

Recall that the spaces $\mathcal{H}_A$, $\mathcal{H}_B$ and $\mathcal{H}_S$ can be thought of as subspaces in $\mathcal{H}$. The subspace $\mathcal{H}_S$ can thus be said to "separate" $\mathcal{H}_A$ and $\mathcal{H}_B$ in $\mathcal{H}$ in a way similar to how S separates A and B in Ω.

Consider a regular domain $\Omega = \cup_{t\in T}(I_t \times I_t)$. Define for $x, y \in T$ such that $x \geq y$, $\Phi_{xy} : \mathcal{H}_y \to \mathcal{H}_x$ as $\Phi_{xy}k_{z,I_y} = k_{z,I_x}$. Thus, $\Phi_{xy} = \Phi_{I_xI_y}$ and $\Phi_{xx} = \mathbf{I}$. Because of the separation property, we have $\Phi_{xz} = \Phi_{xy}\Phi_{yz}$ for $x \geq y \geq z$. Hence, we can write

$$\Phi_{wz} = \Phi_{wx}(\Phi_{xy}\Phi_{yz}) = (\Phi_{wx}\Phi_{xy})\Phi_{yz}$$

for $w \geq x \geq y \geq z$. Thus multiplication in $\{\Phi_{xy}\}_{x\geq y}$ is associative when defined. The maps $\{\Phi_{xy}\}_{x\geq y}$ form a group-like algebraic structure which is called a *semigroupoid*. Much like a group, a semigroupoid consists of a set of elements along with an associative binary operation, but unlike a group, the operation need not be defined for all pairs of elements and moreover, there need not be an inverse. In our case, the operation is operator multiplication.

We say that a set $\{\Phi_{xy}\}_{x\geq y}$ of contractions $\Phi_{xy} : \mathcal{H}_y \to \mathcal{H}_x$ is a *canonical semigroupoid* of K_Ω if $\Phi_{xz} = \Phi_{xy}\Phi_{yz}$ for $x \geq y \geq z$ and $\Phi_{xy}k_{z,I_y} = k_{z,I_x}$ for $z \in I_x \cap I_y$. Notice that the second condition ensures that $\Phi_{xx} = \mathbf{I}$. Moreover, for $x < y$, we can define $\Phi_{xy} = \Phi_{yx}^*$.

Theorem 2.30. *Let K_Ω be a partially reproducing kernel on a regular domain Ω. Then there is a bijective correspondence between canonical completions $\tilde{K}$ and canonical semigroupoids $\{\Phi_{xy}\}_{x\geq y}$ of K_Ω given by*

$$\tilde{K}(x,y) = \langle \Phi_{ts}k_{x,I_s}, k_{y,I_t}\rangle \tag{2.38}$$

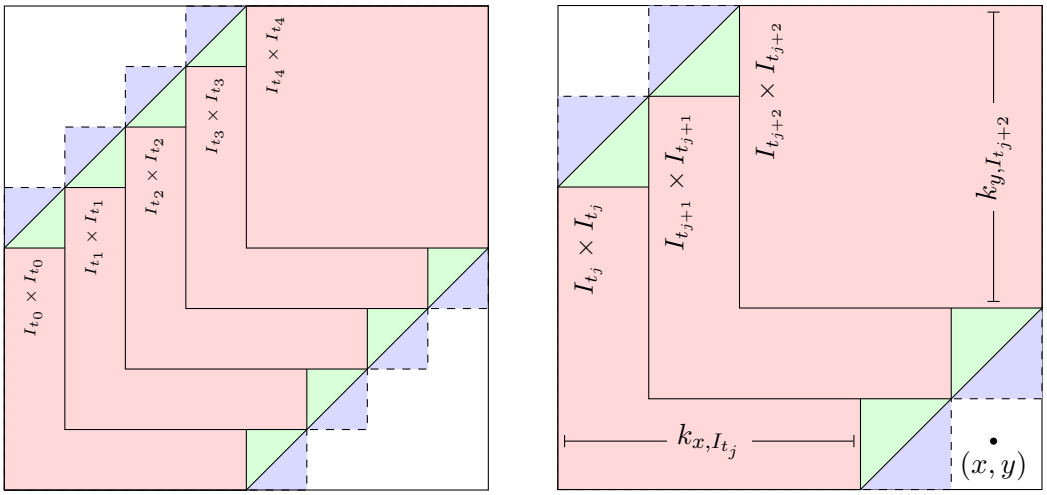
for $x \in I_s$ and $y \in I_t$, where $k_{u,I_v}(w) = K_\Omega(u,w)$ for $v \in T$ and $u, w \in I_v$.

Proof. Given a canonical semigroupoid $\{\Phi_{xy}\}_{x\geq y}$ of K_Ω we can define $\tilde{K}$ using (2.38). Notice that $\tilde{K}(x,y)$ does not depend on the choice of $s, t \in T$ so long as $x \in I_s$ and $y \in I_t$.

Indeed, for $x \in I_s, I_{s'}$ and $y \in I_t$ we have without loss of generality that $s < s' < t$ and

$$\langle \Phi_{ts} k_{x,I_s}, k_{y,I_t} \rangle = \langle \Phi_{ts'} \Phi_{s's} k_{x,I_s}, k_{y,I_t} \rangle = \langle \Phi_{ts'} k_{x,I_{s'}}, k_{y,I_t} \rangle.$$

Thus $\tilde{K}$ is well-defined. To see why $\tilde{K}$ is a reproducing kernel, pick an increasing



(a) $I \times I$. (b) $(I_{t_j} \cup I_{t_{j+2}}) \times (I_{t_j} \cup I_{t_{j+2}})$.

Figure 2.7: Semigroupoid characterization of Canonical Completion. The red and green regions represent Ω while the red, green and blue regions together is $\tilde{\Omega}$.

sequence $\{t_j\}_{j=1}^n \subset T$ such that $t_{j+1} \in I_{t_j}$ for $j \geq 1$ and $\cup_j I_{t_j} = I$. By Theorem 2.7, the restriction of $\tilde{K}$ to $(I_{t_j} \cup I_{t_{j+1}}) \times (I_{t_j} \cup I_{t_{j+1}})$ is a reproducing kernel for $j \geq 1$, since the contraction $\Phi_{t_j t_{j+1}}$ satisfies the necessary conditions. Thus the restriction $\tilde{K}|_{\tilde{\Omega}}$ is a partially reproducing kernel on a serrated domain where $\tilde{\Omega} = \cup_j [(I_{t_j} \cup I_{t_{j+1}}) \times (I_{t_j} \cup I_{t_{j+1}})]$ (see Figure 2.7a). It is now becomes clear from Figure 2.7b that $\tilde{K}$ is merely the canonical completion of $\tilde{K}|_{\tilde{\Omega}}$. For every $x \in I_{t_j} \setminus I_{t_{j+1}}$ and $y \in I_{t_{j+2}} \setminus I_{t_{j+1}}$ are separated by $I_{t_{j+1}}$ in $\tilde{\Omega}$ and we have

$$\begin{aligned}
\tilde{K}(x,y) &= \langle \Phi_{t_{j+2} t_j} k_{x,I_{t_j}}, k_{y,I_{t_{j+2}}} \rangle \\
&= \langle \Phi_{t_{j+2} t_{j+1}} \Phi_{t_{j+1} t_j} k_{x,I_{t_j}}, k_{y,I_{t_{j+2}}} \rangle \\
&= \langle \Phi_{t_{j+1} t_j} k_{x,I_{t_j}}, \Phi_{t_{j+1} t_{j+2}} k_{y,I_{t_{j+2}}} \rangle \\
&= \langle \tilde{k}_{x,I_{t_{j+1}}}, \tilde{k}_{y,I_{t_{j+1}}} \rangle
\end{aligned}$$

where

$$\tilde{k}_{u,I_{t_{j+1}}}(v) = \langle \Phi_{t_{j+1} t_j} k_{u,I_{t_j}}, k_{v,I_{t_{j+1}}} \rangle = \tilde{K}(u,v)$$

for $u \in I_{t_j}$ and $v \in I_{t_{j+1}}$ and we can argue similarly for $\tilde{k}_{y,I_{t_{j+1}}}$. One can now show that $\tilde{K}$ is a canonical completion by proving the separation propoerty for minimal separators I_t for $t \in T$ using arguments similar to the ones above. The converse follows from Lemma 2.7. □

Theorem 2.30 can be thought of as a generalization of Theorem 2.7. It is a more algebraic view which allows us to see canonical completion as simply a consistent way of extending the generators k_{x,I_y} to I. In Section 2.6, we shall see that this semigroupoid can actually be reduced to a nicer algebraic structure called *semigroup* if K_Ω is stationary. The semigroupoids (semigroups) that we are dealing with are equipped with identities (identity) and strictly speaking, should thus be described as *small categories* (*monoids*). Regardless, we shall stick to our chosen terminology for the sake of simplicity.

2.6 Canonical Extensions of Positive-Definite Functions

In this section, we show that every continuous positive-definite function F on an interval $[-a, a]$ $(a > 0)$ admits a canonical extension to the entire real line corresponding to a canonical completion of the partial kernel $K_\Omega(x, y) = f(x-y)$ for $|x-y| \leq a$. Furthermore, this extension admits a representation in terms of a strongly continuous one-parameter semigroup on the RKHS of the kernel $K(x, y) = F(x - y)$ for $x, y \in [0, a]$. The canonical extension can be described in terms of the generator of this semigroup using many classical formulas such as the post-Widder inversion formula.

In addition to proving the existence, we shall also establish the uniqueness of canonical extension under certain technical conditions, although experience with serrated domains suggests that this is generally the case.

2.6.1 The Canonical Extension

The extension problem for positive-definite functions can be framed in terms of completion of a partially reproducing kernel. Let $a > 0$ and F be a positive-definition function on $[-a, a]$. Define $X = [0, b]$ for some $b > a$ and $\Omega = \{(x, y) : |x - y| \leq a\} \subset X \times X$ and $K_\Omega : \Omega \to \mathbb{R}$ as $K_\Omega(x, y) = F(x - y)$. The stationary completions $\tilde{K}$ of K_Ω correspond precisely to the extensions $\tilde{F}$ of F.

A celebrated result of Krein (1940) states that every such function F admits an extension $\tilde{F}$ to $\mathbb{R}$. We shall now prove a stronger statement by showing that there exists a canonical extension which correponds to a stationary canonical completion of K_Ω. Let $K : [0, a] \times [0, a] \to \mathbb{R}$ be the reproducing kernel $K(x, y) = F(x - y)$ and $\mathcal{H} = \mathcal{H}(K)$. Let $I_t = [t, t + a] \cap X$ for $t \geq 0$ and notice that Ω is a regular domain and we can write $\Omega = \cup_{t \in T}(I_t \times I_t)$ for $T = [0, b - a]$.

Theorem 2.31. *Let F be a continuous positive-definite function on $(-a, a)$. Define*

$$K_\Omega(x, y) = F(x - y) \text{ for } (x, y) \in \Omega = \{(u, v) : |u - v| < a\}.$$

Then K_Ω admits a stationary canonical completion.

Proof. Recall that $I_t = [t, t+a]$. Let $\Omega_j = \cup_{k\geq 0}(I_{k/2^j} \times I_{k/2^j})$ for $j \geq 1$. For $j \geq 1$, let K_j denote the canonical completion of $K_{\Omega_j} = K_\Omega|_{\Omega_j}$. By the arguments of Section 2.5.7, we can show that there exists a subsequence $\{K_{j_k}\}$ which converges pointwise to a canonical completion $\tilde{K}$ of K_Ω. Note that for $1 \leq m \leq j_k$,

$$\tilde{K}_{j_k}(x, y) = \tilde{K}_{j_k}(x + j/2^m, y + j/2^m)$$

for $x, y \in \mathbb{R}$ because of the construction (see Figure 2.8). It follows that $\tilde{K}(x, y) = \tilde{K}(x+j/2^n, y+j/2^n)$ for every $j \in \mathbb{Z}$ and $n \geq 1$. By continuity, $\tilde{K}(x, y) = \tilde{K}(x+h, y+h)$ for $h \in \mathbb{R}$. So $\tilde{K}$ is indeed stationary. Hence proved. □

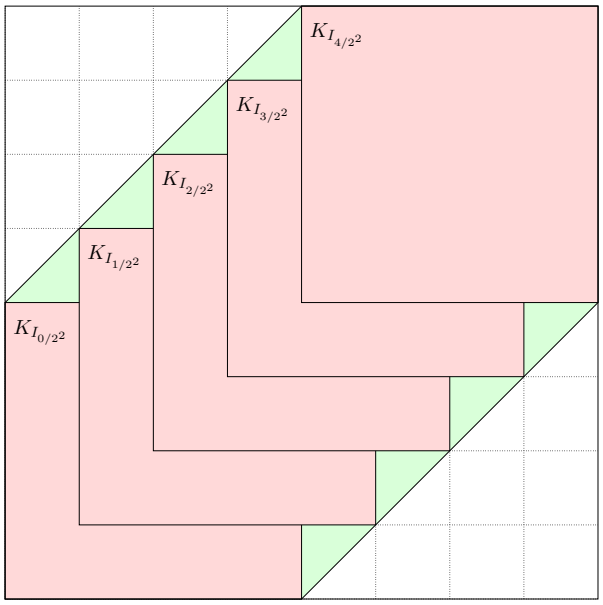


Figure 2.8: Stationary Canonical Completion. The red region represents Ω_j for $j = 2$, while the coloured region represents Ω.

For a stationary canonical completion $\tilde{K}$ of K we can define an extension $\tilde{F} : \mathbb{R} \to \mathbb{R}$ of F as $\tilde{F}(x) = \tilde{K}(x, 0)$. We shall refer to $\tilde{F}$ as a *canonical extension* of F. Conversely, $\tilde{F}$ is a canonical extension of f if $\tilde{K}(x, y) = \tilde{F}(x - y)$ a canonical completion of $K_\Omega(x, y) = F(x - y)$.

Corollary 2.6. *Every continuous positive-definite function $F : [-a, a] \to \mathbb{R}$ admits a canonical extension $\tilde{F}$ which satisfies*

$$\tilde{F}(x - y) = \langle \tilde{f}_x, \rho \tilde{f}_y \rangle$$

where $\tilde{f}_x : (0, a) \to \mathbb{R}$ is given by $\tilde{f}_x(y) = \tilde{f}(x + y)$ and $\rho : \mathcal{H} \to \mathcal{H}$ with $\rho g(y) = g(a - y)$.

2.6.2 The Canonical Semigroup

We shall now describe how the canonical semigroup can be constructed from the canonical semigroupoid corresponding to a canonical completion of K_Ω.

Construction

The canonical semigroup naturally arises when one attempts to describe the semigroupoid picture in terms of a single Hilbert space. Notice that the kernels $K_{I_t} = K_\Omega|_{I_t \times I_t}$ are essentially identical up to a translation in that $K_{I_s}(s+u, s+v) = K_{I_t}(t+u, t+v)$ for every $s, t \in \mathbb{R}$ and every $u, v \in I_0$. Let $\mathcal{H}_t = \mathcal{H}(K_{I_t})$ for $t \in T$.

By Theorem 2.7, there exists a semigroupoid $\{\Phi_{ts} : t, s \in T \text{ and } t \geq s\}$ of contractions maps $\Phi : \mathcal{H}_s \to \mathcal{H}_t$ satisfying $\Phi_{ts} k_{u,I_s} = k_{u,I_t}$ for $t \geq s$. Moreover, define $\tilde{k}_{u,I_v} : I_v \to \mathbb{R}$ $\tilde{k}_{u,I_v}(w) = \tilde{K}(u, w)$ for $w \in I_v$ and note that $\Phi_{st} \tilde{k}_{u,I_s} = \tilde{k}_{u,I_t}$. Define $T_s : \mathcal{H} \to \mathcal{H}_s$ as $T_s g(u) = g(u - s)$. It follows that the adjoint of T_s is given by $T_s^* = T_{-s}$. Notice that $T_s k_{u,I_0} = k_{u+s,I_s}$ for $s, t \in \mathbb{R}$. We shall now reduce the canonical semigroupoid to a one-parameter semigroup.

Theorem 2.32 (Canonical Semigroup)**.** *The operators $T_s^* \Phi_{st} T_t : \mathcal{H} \to \mathcal{H}$ for $s \geq t$ depend only on the difference $(s - t)$, that is*

$$T_s^* \Phi_{st} T_t = T_{s+h}^* \Phi_{s+h,t+h} T_{t+h} \quad \text{for } h \in \mathbb{R}.$$

Define $\Phi_s = T_s^* \Phi_{s0}$ for $s \geq 0$. Then $\{\Phi_t\}_{t \geq 0}$ forms a strongly continuous semigroup on $\mathcal{H}$:

1. $\Phi_0 = \mathbf{I}$,
2. $\Phi_s \Phi_t = \Phi_{s+t}$ *for* $s, t > 0$, *and*
3. $\lim_{h \to 0^+} \|\Phi_h g - g\| = 0$ *for every* $g \in \mathcal{H}$.

Proof. Let $s, t \in \mathbb{R}$. Notice that for $u \in I_0$, we can write

$$\begin{aligned} T_{s+h}^* \Phi_{s+h,t+h} T_{t+h} k_{u,I_0} &= T_{s+h}^* \Phi_{s+h,t+h} \tilde{k}_{u+t+h,I_{t+h}} \\ &= T_{s+h}^* k_{u+t+h,I_{s+h}} \\ &= \tilde{k}_{u+t+h-(s+h),I_0}, \end{aligned}$$

while $T_s^* \Phi_{s,t} T_t k_{u,I_0} = T_s^* \Phi_{s,t} k_{u+t,I_t} = T_s^* \tilde{k}_{u+t,I_s} = \tilde{k}_{u+t-s,I_0}$ implying $T_{s+h}^* \Phi_{s+h,t+h} T_{t+h} k_{u,I_0} = T_s^* \Phi_{s,t} T_t k_{u,I_0}$ for $u \in I_0$ and therefore, $T_{s+h}^* \Phi_{s+h,t+h} T_{t+h} = T_s^* \Phi_{s,t} T_t$. It is obvious that

$\Phi_0 = \mathbf{I}$ and we can argue as before that $\Phi_s\Phi_t = \Phi_{s+t}$ because

$$\begin{aligned}\Phi_s\Phi_t k_{u,I_0} &= T_s^*\Phi_{s0}T_t^*\Phi_{t0}k_{u,I_0}\\ &= T_s^*\Phi_{s0}T_t^*\tilde{k}_{u,I_t}\\ &= T_s^*\Phi_{s0}\tilde{k}_{u-t,I_0}\\ &= \tilde{k}_{u-t-s,I_0}\\ &= \Phi_{s+t}k_{u,I_0}\end{aligned}$$

for $u \in I_0$. We need only verify that $\lim_{h\to 0^+}\Phi_h g = g$ for $g \in \mathcal{H}$. Let $g = \sum_{i=1}^n \alpha_i k_{u_i,I_0}$ for some $n \geq 1$, $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$ and $\{u_i\}_{i=1}^n \subset I_0$. Then

$$\|\Phi_h g - g\|^2 = \sum_{i,j=1}^n \alpha_i\alpha_j\left[\tilde{K}(u_i+h,u_j+h) - \tilde{K}(u_i,u_j+h) - \tilde{K}(u_i+h,u_j) + \tilde{K}(u_i,u_j)\right] \to 0$$

as $h \to 0^+$. Because $\mathrm{Span}\{k_{u,I_0} : u \in I_0\}$ is dense in $\mathcal{H}$ and $\|\Phi_h\| \leq 1$, it follows that Φ_h converges strongly to $\mathbf{I}$ as $h \to 0^+$ (see (Eidelman et al., 2004, Lemma 9.4.7)) and we are done. Alternatively, we could have used the equivalence of strong and weak continuity (see (Engel and Nagel, 2000, Theorem 5.8)). □

The following corollary is now immediate from Theorem 2.30.

Theorem 2.33. *Let $a > 0$. There is a bijective correspondence between the canonical extensions $\tilde{F}$ of a continuous positive-definite function F on $[-a,a]$ and strongly continuous semigroups $\{\Phi_t\}_{t\geq 0}$ of contractions on $\mathcal{H} = \mathcal{H}(K)$ where $K(x,y) = F(x-y)$ for $x,y \in [0,a]$ satisfying*

$$\Phi_t k_u = k_{u-t} \quad \text{for } 0 \leq t \leq u \leq a$$

given by

$$\tilde{F}(t) = \begin{cases}\langle k_0, \Phi_t k_0\rangle & \text{for } t \geq 0\\ \langle k_0, \Phi_{-t}k_0\rangle & \text{for } t < 0\end{cases} \tag{2.39}$$

where $k_0(u) = F(u)$ for $0 \leq u \leq a$.

Although, all positive-definite functions admit a unitary representation resembling (2.39) (Stewart, 1976), canonical extensions $F_\star$ admit a very concrete representation of that kind. As a consequence of Theorem 2.6, we have

Corollary 2.7. *Let $a > 0$. Every continuous positive-definite function F on $[-a,a]$ admits a canonical extension $\tilde{F}$ with the representation* (2.39) *for some strongly continuous semigroup of contractions $\{\Phi_t\}_{t\geq 0}$ on $\mathcal{H}(K)$ where $K(x,y) = F(x-y)$ for $x,y \in [0,a]$.*

Generators of Canonical Semigroups

The canonical semigroup $\{\Phi_t\}_{t\geq 0}$, like all strongly continuous one-parameter semigroups, admits a generator which is defined as the linear operator $\partial_\star : \mathcal{D}(\partial_\star) \to \mathcal{H}$ given by

$$\partial_\star f = \lim_{h\to 0^+} \frac{1}{h}\left[\Phi_h f - f\right]$$

for $f \in \mathcal{D}(\partial_\star) = \{f \in \mathcal{H} : \lim_{h\to 0^+} \frac{1}{h}\left[\Phi_h f - f\right] \text{ exists}\}$. In general, the operator $\partial_\star$ is not bounded and its domain $\mathcal{D}(\partial_\star)$ is not equal to $\mathcal{H}$. However, according to the Hille-Yosida Theorem for contraction semigroups (see Theorem 3.5 of Engel and Nagel (2000)), the operator $\partial_\star$ is closed, its domain $\mathcal{D}(\partial_\star)$ is dense in $\mathcal{H}$, and the operator $\lambda\mathbf{I} - \partial_\star$ has a bounded inverse satisfying $\|(\lambda\mathbf{I} - \partial_\star)^{-1}\| \leq 1/\lambda$ for $\lambda > 0$. Most importantly, the generator $\partial_\star$ uniquely determines the semigroup $\{\Phi_t\}_{t\geq 0}$.

The connection with semigroups furnishes many interesting representations for a canonical extension $F_\star$ in terms of the generator $\partial_\star$.

Theorem 2.34. *Let $F_\star$ be a canonical extension of a continuous positive-definite function F on $[-a, a] \subset \mathbb{R}$ for some $a > 0$ and let $\partial_\star$ be the generator of the corresponding canonical semigroup. For $x > 0$, we have*

$$F_\star(t) = \lim_{\mu\to 0} \left\langle e^{t\partial_\star[\mathbf{I}-\mu\partial_\star]^{-1}} k_0, k_0 \right\rangle \qquad \text{(Yosida Approximation Formula)}$$

$$F_\star(t) = \lim_{n\to\infty} \left\langle \left[\mathbf{I} - \tfrac{t}{n}\partial_\star\right]^{-n} k_0, k_0 \right\rangle \qquad \text{(Post-Widder Inversion Formula)}$$

$$F_\star(t) = \frac{1}{2\pi \mathrm{i}} \lim_{n\to\infty} \int_{\epsilon-in}^{\epsilon+in} \langle e^{zt}[z\mathbf{I} - \partial_\star]^{-1} k_0, k_0\rangle \, dz \qquad \text{(Cauchy Integral Formula)}$$

where $\mathrm{i} = \sqrt{-1}$ *and the convergence is uniform over compact intervals of* $\mathbb{R}_+$*. The integral is the last equation is to be understood as the usual contour integral from complex analysis.*

Proof. These can be readily seen as straightforward consequences of Theorem 3.5, Corollary 5.5 and Theorem 5.14 from Engel and Nagel (2000) in our setting. □

Of course, the beautiful expressions in Theorem 2.34 don't mean much to us if we can't calculate $\partial_\star$ independently of $F_\star$. In the following section, we consider a plausible situation in which the canonical extension can be recovered as the closure $\bar{\partial}$ of an explicitly defined operator ∂. In addition to bringing the expressions in Theorem 2.34 to life, this proves that under the considered scenario the canonical extension is unique.

2.6.3 Differential Equations in Hilbert Space

The semigroup connection also allows us to think of the canonical extension as the solution of an abstract Cauchy problem in $\mathcal{H} = \mathcal{H}(K)$. Consider a function $f : \mathbb{R}_+ \to \mathcal{H}$. We shall denote the value of f at $t \in \mathbb{R}_+$ by f_t. We say that f is *Fréchet differentiable* at $t \in \mathbb{R}_+$ if there exists $\partial_t f_t \in \mathcal{H}$ such that $\lim_{h\to 0} \|\frac{1}{h}(u_{t+h} - u_t) - \partial_t u_t\| = 0$. According to Proposition 6.2 of Engel and Nagel (2000), if $k_0 \in \mathcal{D}(\partial_\star)$, then $f_t = \Phi_t k_0 (= \tilde{k}_{0,I_t})$ is the unique solution of the abstract Cauchy problem: for $t \geq 0$

$$\begin{cases} \partial_t f_t = \partial_\star f_t, \\ f_0 = k_0. \end{cases} \tag{2.40}$$

When $k_0 \notin \mathcal{D}(\partial_\star)$, then Proposition 6.4 of Engel and Nagel (2000) tells us that the function $f : t \mapsto \Phi_t k_0$ can still be understood as the unique *mild solution* to (2.40) in the sense that for $t \geq 0$, $\int_0^t f_u \, du \in \mathcal{D}(\partial_\star)$ and

$$f_t = k_0 + \partial_\star \int_0^t f_u \, du. \tag{2.41}$$

In essence, the problem of canonical positive-definite extension is equivalent to solving an abstract differential equation in a certain Hilbert space.

Recovery of the Generator

Computing the operator $\partial_\star$ explicitly or even identifying its domain $\mathcal{D}(\partial_\star)$ precisely is usually very difficult even when $\{\Phi_t\}_{t\geq 0}$ is known. Ours is a more complicated situation since we only know certain images of $\{\Phi_t\}_{t\geq 0}$ as given by

$$\Phi_s k_t = k_{t-s} \text{ for } 0 \leq s \leq t \leq a.$$

Fortunately, it is often possible to evaluate $\partial_\star$ over a subset $\mathcal{D} \subset \mathcal{D}(\partial_\star)$. Let $\mathcal{D}$ denote the set of integrals $\int_0^a \alpha(u) k_u \, du \in \mathcal{H}$ where α is an infinitely differentiable real-valued function on $(0, a)$ with compact support. Note that $\mathcal{D}$ is a dense linear subspace of $\mathcal{H}$. The elements of $\mathcal{D}$ serve essentially the same purpose as that of test functions in the theory of distributions. By definition,

$$\begin{aligned} \partial_\star \left[\int_0^a \alpha(u) k_u \, du \right] &= \lim_{h\to 0^+} \frac{1}{h} \left[\int_0^a \alpha(u) k_{u-h} \, du - \int_0^a \alpha(u) k_u \, du \right] \\ &= \lim_{h\to 0^+} \frac{1}{h} \left[\int_0^a \alpha(u+h) k_u \, du - \int_0^a \alpha(u) k_u \, du \right] \\ &= \int_0^a \left[\lim_{h\to 0^+} \frac{\alpha(u+h) - \alpha(u)}{h} \right] k_u \, du \\ &= \int_0^a \alpha'(u) k_u \, du. \end{aligned}$$

Thus $\partial_\star\left[\int_0^a \alpha(u)k_u\ du\right] = \int_0^a \alpha'(u)k_u\ du \in \mathcal{D}$ and $\mathcal{D}$ is invariant under $\partial_\star$. In the same way, one can also work out $\partial_\star\left[\int_0^a \alpha(u)k_u\ du\right]$ for piecewise once-differentiable α with compact support as in $\partial_\star \int_s^t k_u\ du = k_s - k_t$ for $0 < s < t < a$, although expanding $\mathcal{D}$ to include such elements is probably not of much consequence.

Define the operator $\partial : \mathcal{D} \to \mathcal{H}$ as

$$\partial\left[\int_0^a \alpha(u)k_u\ du\right] = \int_0^a \alpha'(u)k_u\ du.$$

Thus $\partial = \partial_\star|_{\mathcal{D}}$ but here we have defined it exclusively in terms of K without referring to $\partial_\star$ or $F_\star$. We would like to recover $\partial_\star$ from its restriction ∂ to a dense subset $\mathcal{D} \subset \mathcal{H}$. If $\partial_\star$ is continuous on $\mathcal{H}$, then this is possible using an ordinary extension by continuity argument. However, F is analytic if $\partial_\star$ is continuous (Remark 2.2). Because analyticity implies unique extension anyway, the special case of bounded generators is unintersting in that it does not offer us any insight into the problem of canonical extension.

In general, it is not possible to recover an unbounded operator from its restriction to a dense subspace. Fortunately, $\partial_\star$ is a closed operator, which means that the graph $\mathcal{G}_\star = \{(f, \partial_\star f) : f \in \mathcal{D}(\partial_\star)\}$ is a closed subset of $\mathcal{H} \times \mathcal{H}$. This makes a different kind of extension by continuity possible. If the closure $\bar{\mathcal{G}}$ in $\mathcal{H}\times\mathcal{H}$ of a graph $\mathcal{G} = \{(f, \partial f) : f \in \mathcal{D}\}$ for some operator ∂, is equal to $\mathcal{G}_\star$ then that would mean that we can recover $\partial_\star$ as the *closure* $\bar{\partial}$ of ∂, given by:

$$\bar{\partial} f = \lim_{j\to\infty} \partial f_j$$

where $\{f_j\}_{j=1}^\infty \subset \mathcal{D}$ such that $\lim_{j\to\infty} f_j = f \in \mathcal{D}(\partial_\star)$ and $\lim_{j\to\infty} \partial f_j$ exists. An alternative way of stating this is to say that $\mathcal{D}$ is a *core* of $\partial_\star$, which is to say that $\mathcal{D}$ is dense in $\mathcal{D}(\partial_\star)$ with respect to the norm $\|f\|_\partial = \|f\| + \|\partial f\|$ where $f \in \mathcal{D}$. We now present certain criteria for $\mathcal{D}$ to be a core of $\partial_\star$.

Theorem 2.35. *Suppose that $e^{-\lambda x} \notin \mathcal{H}$ for some $\lambda > 0$. Then $(\lambda\mathbf{I} - \partial)\mathcal{D}$ is dense in $\mathcal{H}$ and*

1. *F admits a unique canonical extension $F_\star$,*
2. *$\partial_\star = \bar{\partial}$ and $\mathcal{D}$ is a core of $\mathcal{D}(\partial_\star)$,*

Proof. Let f be in the orthogonal complement of $(\lambda\mathbf{I} - \partial)\mathcal{D}$. Then

$$\int_0^a [\lambda\alpha(u) - \alpha'(u)]f(u)\ du\rangle = \langle f, \int_0^a [\lambda\alpha(u) - \alpha'(u)]k_u\ du = 0$$

for every infinitely differentiable α with a compact support in $(0, a)$. Using some elementary distribution theory, it can be shown that this can only be true if $\lambda f + f' = 0$ or $f(u) = ce^{-\lambda u}$ for $0 \le x < a$ for some $c \in \mathbb{R}$. Since, $e^{-\lambda x} \notin \mathcal{H}$, it follows that $c = 0$ and

$f = 0$, thus implying that $(\lambda \mathbf{I} - \partial)\mathcal{D}$ is dense in $\mathcal{H}$. The conclusion now follows from Theorem 5.2 of Engel and Nagel (2000). □

The condition $e^{-\lambda x} \notin \mathcal{H}$ is equivalent to saying that $K_\lambda(x, y) = e^{\lambda(x+y)}F(x - y) - c$ is not a reproducing kernel for any $c > 0$. It is unclear how stringent this requirement is, but in light of the consequence, we need only worry about the case of positive-definite functions F for which $e^{-\lambda x} \in \mathcal{H}$ for every $\lambda > 0$, which remains unsolved.

Theorem 2.36. *If for every infinitely differentiable $\alpha : (0, a) \to \mathbb{R}$ with a compact support there exists $r > 0$ such that*

$$\sum_{j=0}^{\infty} \frac{r^j}{j!} \sqrt{\int_0^a \int_0^a D^j\alpha(u)D^j\alpha(v)F(u - v)\ dudv} < \infty. \tag{2.42}$$

then, F admits a unique canonical extension $F_\star$, $\partial_\star = \bar{\partial}$ and $\mathcal{D}$ is a core of $\mathcal{D}(\partial_\star)$.

Proof. This follows from Theorem 1.51 of Davies (1980) by noticing that the expression (2.42) is equivalent to $\sum_{j=0}^{\infty} \frac{r^j}{j!}\|\partial^j f\| < \infty$ for $f = \int_0^a \alpha(u)k_u\ du$. □

Both conditions $e^{-\lambda x} \notin \mathcal{H}$ and (2.42) are very difficult to verify in general and as a result, we are unable to construct examples of positive-definite function F for which they apply. We conclude this section by pointing out the stringency of assuming that the generator $\partial_\star$ is bounded.

Remark 2.2. *The generator $\partial_\star$ is bounded precisely when the semigroup is uniformly continuous, that is* $\lim_{h\to 0^+} \|\Phi_t - \mathbf{I}\| = 0$. *In this case, $\Phi_t = \exp(t\partial_\star)$ and we can write*

$$\tilde{F}(t) = \langle k_0, \Phi_t k_0\rangle = \langle k_0, \exp(t\partial_\star)k_0\rangle = \sum_{j=0}^{\infty} \langle k_0, \partial_\star^j k_0\rangle \frac{t^j}{j!}$$

which implies that $F_\star$ and hence, F is analytic! The series has an infinite radius of convergence, which makes $\tilde{F}$ an entire function.

3 Continuously Indexed Graphical Models

> The greatest value of a picture is when it forces us to notice what we never expected to see.
>
> John W. Tukey, *Exploratory Data Analysis* (1977)

Abstract

Let $X = \{X_u\}_{u \in U}$ be a real-valued Gaussian process indexed by a set U. It can be thought of as an undirected graphical model with every random variable X_u serving as a vertex. We characterize this graph in terms of the covariance of X through its reproducing kernel property. Unlike other characterizations in the literature, our characterization does not restrict the index set U to be finite or countable, and hence can be used to model the intrinsic dependence structure of stochastic processes in continuous time/space. Consequently, the said characterization is not (and apparently cannot be) of the inverse-zero type. This poses novel challenges for the problem of recovery of the dependence structure from a sample of independent realizations of X, also known as structure estimation. We propose a methodology that circumvents these issues, by targeting the recovery of the underlying graph up to a finite resolution, which can be arbitrarily fine and is limited only by the available sample size. The recovery is shown to be consistent so long as the graph is sufficiently regular in an appropriate sense, and convergence rates are provided. Our methodology is illustrated by simulation and two data analyses.

3.1 Introduction

We consider the problem of defining undirected graphical models with *uncountable* vertex sets with the purpose of describing conditional independence relationships inherent in stochastic process over continuous time/space – in the same way as ordinary (finite) undirected graphical models do for random vectors in Euclidean spaces. Furthermore, we consider the statistical problem of *recovering* the graph from a finite number of

independent realizations of the process up to a degree of resolution commensurate with the amount of data available.

Consider a zero-mean Gaussian process $X = \{X_u\}_{u \in U}$ on a (possibly uncountably infinite) set U. We would like to think of X as a Gaussian graphical model with every random variable X_u corresponding to a vertex of a graph Ω_X on the index set U. The conditional independence structure of X should likewise correspond to the edge structure of Ω, in that, for $u, v \in U$ separated by $W \subset U$ we have

$$X_u \perp\!\!\!\perp X_v \mid X_W$$

where $X_W = \{X_w : w \in W\}$. To this end, we will characterize the covariance of processes admitting a given graphical structure in terms of the reproducing kernel property. And, going in the other direction, we will use this characterization to define the *graph of a process* in terms of its covariance.

Although the stated characterization is always valid, it is somewhat unwieldy for the purpose of parsing the graph of a given process from its covariance. In the finite vertex set case, we have a particularly handy result, sometimes called the *inverse zero characterization*, which says that if the covariance matrix is invertible, then the ij-th entry of the inverse of the covariance matrix is zero precisely when there is no edge between the ith and jth vertices. For (uncountably) infinite U, a direct analogous characterization for covariance kernels is unavailable to us, if indeed it exists at all. In order to derive an analogous result, we develop a notion of *resolution* of a graph. This allows for an alternative inverse zero characterization, yielding a *pixelated* version of the graph of a continuously indexed process, from the zero entries of a certain *correlation operator matrix* related to its covariance. The choice of resolution can be arbitrary large, and under appropriate conditions yields an exact characterization of the graph in the limit.

This framework also allows us to meaningfully pose the problem of recovering the graph from n independent realizations of the process, with the resolution being dictated by the available sample size. Because arbitrarily small changes in the covariance kernel can greatly alter the graph of the associated process, targeting the graph at a sample-dependent finite resolution can also be seen as quantifying how finely the graphical structure can be resolved with a given amount of finite information. In this framework, we propose a graph estimator that relies on thresholding (in the operator norm) of the entries of the inverse empirical correlation operator matrix. Under standard regularity assumptions on the correlation operator matrix, we show that the underlying graph can be recovered with high probability as the number of samples increases. Also, we give a lower bound for the sample size to recover the graph at a given familywise error rate.

Although we restrict focus on Gaussian processes, our analysis can be easily extended to sub-Gaussian processes by interpreting the graph in terms of "conditional uncorrelatedness"

instead of conditional independence. The resulting structure corresponds to a correlational graphoid (Pearl and Paz, 1985) and a basic strong separoid (Dawid, 2001), and therefore serves as a reasonable alternative to conditional independence.

The main contributions of this chapter are the notion of graph of a Gaussian process, its finite resolution inverse zero characterization and the idea of regularization by pixelation. Furthermore, under standard regularity conditions, we derive better and simpler convergence rates for the estimation of the correlation operator matrix of a second order random element. For Gaussian random elements, we derive concentration bounds for the estimator.

3.1.1 Background and Related Work

Undirected graphical models allow us to distinguish direct and indirect associations in data, and thus have a long history in statistics. They have been investigated as models (Dempster, 1972; Darroch et al., 1980), and as targets of inference (Lauritzen, 1996), with a particular emphasis on high-dimensional settings more recently (Meinshausen and Bühlmann, 2006; Ravikumar et al., 2011; Rothman et al., 2008). Infinite dimensional graphical models have been investigated by (Montague and Rajaratnam, 2018) from an axiomatic and probabilistic point of view.

Graphical models with uncountably infinite number of vertices have not received much attention in the literature but they are implicit in the study of Markov processes, which can be regarded as infinite graphical models with infinitesimally small graphs. The generalization of the Markov property to Euclidean spaces by McKean (1963) using the concept of splitting fields and to locally compact metric spaces by Rozanov (1982), along with the generalization of the Markov property itself to the quasi-Markov property by Chay (1972), can be thought of as important steps in this direction.

Graphical models are frequently used to model continuous time (or space) stochastic processes under the label of "Gaussian Markov random fields (GMRFs)" (Rue and Held, 2005). Although this is usually done for computational benefits, there are important cases in which there is an explicit link between the underlying process and the GMRFs used to model them (Lindgren et al., 2011). Roughly speaking, this amounts to modelling the graphical structure of the process itself.

Our own work complements some of the recent developments in functional data analysis concerning graphical models. In the context of functional data, a graphical model can refer to several distinct possibilities. To explain the nuances involved, we introduce some

notation. Consider an $\mathbb{R}^p$-valued stochastic process X on an interval $I \subset \mathbb{R}$ given by

$$t \mapsto \begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_p(t) \end{bmatrix} \in \mathbb{R}^p.$$

Viewing this as a vector-valued function, Qiao et al. (2020) deal with recovering the graphical structure between $\{X_j(t) : 1 \leq j \leq p\}$ as a function of t. This can be thought of as a pointwise finite graphical model: for every t, one has a graphical model on p vertices. This perspective is related to Mogensen and Hansen (2022), who consider finitely indexed graphical models on diffusions in $\mathbb{R}^p$. On the other hand, viewing each function globally $t \mapsto X_j(t)$ as a random element in a Hilbert space $\mathbb{H}$, one has a single p-vector with Hilbertian entries,

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \in \mathbb{H}^p.$$

In this context, Qiao et al. (2019), Li and Solea (2018a) and Lee et al. (2023) address the problem of recovering the graphical structure between the p vector coordinates X_j for $1 \leq j \leq p$. Thus they address the problem of recovering the structure *between* a finite number of related random functions. This can be seen as a global, rather than pointwise approach.

In either case, the problem can be seen as recovering the dependence structure *between* a finite collection of p random functions. In contrast, we wish study the structure *within* a single random function. That is, the graphical structure of the collection $\{X_j(t) : t \in I\}$ for a given fixed j. Thus, we are interested in an *intrinsic* graphical model. Importantly, this means that we are concerned with the problem of recovering the dependence structure between an *uncountably infinite* number of jointly distributed random variables, unlike the above mentioned literature, which deals with a *finite* number of real random variables or Hilbertian random elements. Indeed, we will see that our setting subsumes existing notions of functional graphical models as special cases.

3.1.2 Outline of the Chapter

In Section 3.2, we introduce some notation and review certain basic concepts concerning the theory of graphs, reproducing kernels, and linear operators. In Section 3.3, we present our characterization of the conditional independence structure of a Gaussian process in terms of its covariance function. Furthermore, we make concrete the notion of *the graph of a process* and derive the graphs of some familiar classes of Gaussian

processes explicitly. In Section 3.4, we explain the concept of resolution in greater detail, and derive an analogue of the finite-dimensional inverse zero characterization (3.5), which we use to come up with a sufficient criterion for the approximate and exact identifiability of the graph of a process. Additionally, we discuss parallels and differences in our setting/approach and existing approaches to functional graphical models. Finally, in Section 3.5, we describe our algorithm for graph recovery. In Section 3.6 and 3.6.2, we provide asymptotic theory and recovery guarantees. In Section 3.7, we present a simulation study to gauge the performance recovery procedure, covering a variety of covariances at different resolutions and samples sizes. In Section 3.8, we illustrate our method by applying it to spectroscopy and intraday stock price data.

3.2 Preliminaries and Notation

An *undirected graph* on a set U is defined as a pair (U, Ω) where $\Omega \subset U \times U$ such that for any $(u, v) \in U \times U$ we have $(u, u) \in \Omega$ and $(u, v) \in \Omega \iff (v, u) \in \Omega$.

The set U is called the vertex set and the set Ω is called the edge set. All graphs in this chapter are undirected. Since the vertex set will always be fixed, we shall refer to a graph by its edge set Ω. We shall say $u, v \in U$ are *adjacent* if $(u, v) \in \Omega$, that is, if they have an edge between them. By convention, we shall assume that every vertex has an edge with itself. To visualize the graph Ω, notice that the *adjacency function* $\mathbf{1}_\Omega : U \times U \to \mathbb{R}$ given by

$$\mathbf{1}_\Omega(u, v) = \begin{cases} 1 & (u, v) \in \Omega \\ 0 & (u, v) \notin \Omega \end{cases}$$

describes the structure of the graph in a way analogous to how the adjacency matrix does the same when U restricted to be finite. A graph is called *complete* if all vertices are adjacent to each other. The unique complete graph on U is given by $\Omega = U \times U$.

For $u, v \in U$, a *path* on Ω from u to v is a finite sequence $\{w_k\}_{k=0}^{n+1}$ of vertices such that $w_0 = u$, $(w_k, w_{k+1}) \in \Omega$ (they are adjacent) for $0 \leq k \leq n$, and $w_{n+1} = v$. The vertices u and v are called *connected* if there is a path between them and *disconnected* otherwise. A subset W of U is said to separate $u, v \in U$ if every path between u and v passes through W. If u and v are disconnected, then they can be said to be separated by the empty set $\varnothing$.

A graphical model consists of a set of random variables $X = \{X_u : u \in U\}$ indexed by a set U, and a graph $\Omega \subset U \times U$, such for every $u, v \in U$ separated by $W \subset U$ in Ω, X satisfies

$$X_u \perp\!\!\!\perp X_v \mid X_W. \tag{3.1}$$

Here $X_W := \{X_u : u \in W\}$ represents the restriction of X to $W \subset U$. It is implicit in the definition that if u and v are disconnected, then X_u and X_v are independent.

The *separation condition* (3.1) brings together X and Ω by making the conditional independence structure of X conform with the edge structure of the graph Ω. Note that, for notational convenience, we have defined our graphical models slightly differently than the standard nomenclature: the vertex set of our graph is the domain U instead the set of random variables $\{X_u : u \in U\}$.

Let K be the covariance of the process X. Define the functions $K(u,\cdot), K(\cdot,u) : U \to \mathbb{R}$ as $v \mapsto K(u,v)$ for $u, v \in U$. The reproducing kernel Hilbert space $\mathcal{H}(K)$ of K is defined as the closure of the linear span of $\{K(u,\cdot) : \text{ for } u \in U\}$ under the norm induced by the inner product $\langle K(u,\cdot), K(\cdot,v)\rangle = K(u,v)$ for $u, v \in U$. We shall denote the inner product of $f, g \in \mathcal{H}(K)$ as $\langle f, g\rangle_{\mathcal{H}(K)}$.

We shall work with operators on Hilbert spaces. Boldface alphabet such as $\mathbf{A}$ will be used to denote an operator or an operator matrix. Note that an operator matrix can also be thought of as an operator on an appropriate product Hilbert space. For an operator matrix $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1}^{p}$, we shall use $\operatorname{dg}\mathbf{A}$ to denote the diagonal part $[\delta_{ij}\mathbf{A}_{ij}]_{i,j=1}^{p}$ where δ_{ij} is the Kronecker delta and $\mathbf{A}_0$ to denote the off-diagonal part $(\mathbf{A} - \operatorname{dg}\mathbf{A})$. The spectrum of a self-adjoint operator $\mathbf{A}$ will be denoted by $\sigma(\mathbf{A})$. If $\mathbf{A}$ is compact, then its kth eigenvalue shall be denoted by $\lambda_k(\mathbf{A})$.

3.3 Graphical Representation of Gaussian Processes

In this section, we characterize the relationship between the conditional independence structure of a Gaussian process X and its covariance kernel K. We then use this characterization to define the graph of a Gaussian process and discuss certain conceptual differences with respect to the finite index setting.

3.3.1 The Separation Equation

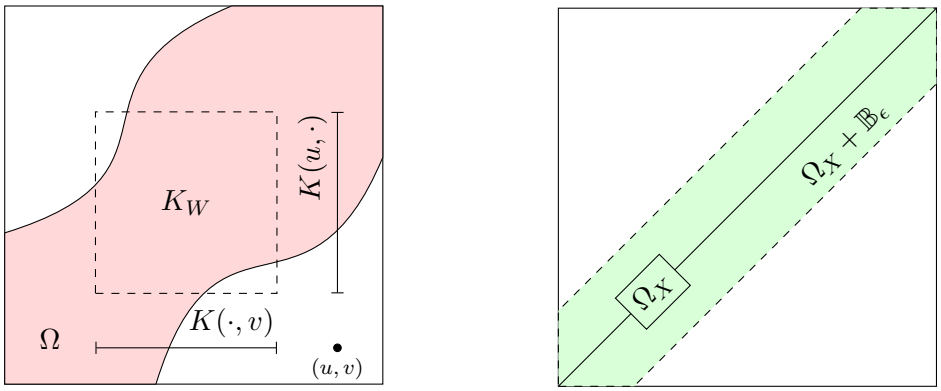
Let $X = \{X_u : u \in U\}$ be a Gaussian process on a set U satisfying the separation condition (3.1) for some graph $\Omega \subset U \times U$. Because X is Gaussian, this is equivalent to requiring that for every $u, v \in U$ separated by $W \subset U$ (see Figure 3.1 (a)), the conditional covariance given by

$$\operatorname{Cov}(X_u, X_v | X_W) = \mathbb{E}[X_u X_v | X_W] - \mathbb{E}[X_u | X_W] \cdot \mathbb{E}[X_v | X_W]$$

must vanish almost surely. Taking the expectation and using the law of iterated expectation, this implies that

$$\mathbb{E}[X_u X_v] = \mathbb{E}\Big[\mathbb{E}[X_u | X_W] \cdot \mathbb{E}[X_v | X_W]\Big] \tag{3.2}$$

almost surely. We shall now express this statement in terms of the kernel K.



(a) (b)

Figure 3.1: (a) An example of $u, v \in U$ separated by $W \subset U$ indicated by $W \times W$ (dashed square) along with the restrictions $K(u,\cdot)|_W$ and $K(\cdot,v)|_W$, and (b) The graph Ω_X of Brownian motion (the diagonal) and the ϵ-envelope $\Omega_X + \mathbb{B}_\epsilon$ (in green).

The closed linear span $\mathcal{L}(X)$ of $X = \{X_u : u \in U\}$ under the norm $Y \mapsto \mathbb{E}[Y^2]$ forms a Hilbert space under the inner product $(Y_1, Y_2) \mapsto \mathbb{E}[Y_1 Y_2]$ which is induced by the norm. By Loève's isometry (Loeve, 2017), $\mathcal{L}(X)$ is isometrically isomorphic to $\mathcal{H}(K)$. When $W \subset U$ separates $(u, v) \in U \times U$, this enables us to rewrite Equation (3.2) as

$$\langle K(u,\cdot), K(\cdot,v)\rangle_{\mathcal{H}(K)} = \langle \Pi_W K(u,\cdot), \Pi_W K(\cdot,v)\rangle_{\mathcal{H}(K)} \tag{3.3}$$

where Π_W denotes the projection in $\mathcal{H}(K)$ to the closed linear subspace generated by $\{K(w,\cdot) : w \in W\}$. As before, we shall consider it implicit that if u and v are disconnected then they are separated by $W = \varnothing$ and $K(u, v) = 0$.

By the reproducing property, $\langle K(u,\cdot), K(\cdot,v)\rangle_{\mathcal{H}(K)} = K(u,v)$. By the subspace isometry (Paulsen and Raghupathi, 2016), the inner product $\langle \Pi_W K(u,\cdot), \Pi_W K(\cdot,v)\rangle_{\mathcal{H}(K)}$ can be evaluated by taking the inner product of the restrictions $K(u,\cdot)|_W$ and $K(\cdot,v)|_W$ in the reproducing kernel Hilbert space of the restriction $K_W = K|_{W\times W}$ of the kernel K. Thus,

$$K(u,v) = \langle K(u,\cdot), K(\cdot,v)\rangle_{\mathcal{H}(K_W)}. \tag{3.4}$$

We shall refer to (3.4) as the *separation equation*. Going in the opposite direction, notice that the above equation implies

$$\langle K(u,\cdot) - \Pi_W K(u,\cdot), K(\cdot,v) - \Pi_W K(\cdot,v)\rangle_{\mathcal{H}(K)} = 0.$$

Because of Gaussianity and the Loève isometry, this means that $X_u - \mathbb{E}[X_u|X_W]$ and $X_v - \mathbb{E}[X_v|X_W]$ are independent. Additionally, they are both independent of X_W. It

follows that

$$\mathrm{Cov}(X_u, X_v|X_W) = \mathbb{E}\Big[(X_u - \mathbb{E}[X_u|X_W])(X_v - \mathbb{E}[X_v|X_W])|X_W\Big] = 0.$$

To summarize, we have established the following theorem.

Theorem 3.1. *Given a Gaussian process $X = \{X_u : u \in U\}$ and a graph $\Omega \subset U \times U$, the following two statements are equivalent:*

(A) For every $u, v \in U$ separated by $W \subset U$ in Ω,

$$X_u \perp\!\!\!\perp X_v \mid X_W.$$

(B) For every $u, v \in U$ separated by $W \subset U$ in Ω,

$$K(u, v) = \langle K(u, \cdot), K(\cdot, v)\rangle_{\mathcal{H}(K_W)}.$$

Simply stated, the conditional independence statement (3.1) can be exchanged with the equation (3.4) in the definition of a graphical model.

One of the properties which force a Gaussian process to obey the separation equation with respect to a "memory" graph is the analyticity of the covariance kernel, as illustrated by the following example.

Example 3.1. *Let $X = \{X_t\}_{t\in I}$ be a Gaussian process on the unit interval I with an analytic covariance K. Then K satisfies the separation equation for every Ω which contains the strip $\{(u, v) : |u - v| \le w\}$ for some $w > 0$. Indeed, for any two points $u, v \in I$ separated by $W \subset I$, W must contain an interval of finite length. This implies that the function $f = K(u, \cdot) - \Pi_W K(u, \cdot)$ is zero on an interval of finite length because $f(w) = \langle K(u, \cdot) - \Pi_W K(u, \cdot), K(\cdot, w)\rangle = 0$ for $w \in W$ by the projection theorem. But f is analytic and hence,*

$$f = K(u, \cdot) - \Pi_W K(u, \cdot) \equiv 0.$$

By repeating the same argument for v, we can show that

$$K(u, v) = \langle K(u, \cdot), K(\cdot, v)\rangle = \langle \Pi_W K(u, \cdot), \Pi_W K(\cdot, v)\rangle = \langle K(u, \cdot), K(\cdot, v)\rangle_{\mathcal{H}(K_W)}$$

and the conclusion follows. This argument can be easily extended to Gaussian processes on connected domains in a Euclidean space which have analytic covariances.

It is natural to ask why the relationship between the conditional independence structure of X and its covariance K has to be expressed by such tortuous means. After all, if $\mathbf{X} = \{X_j\}_{j=1}^p$ is a Gaussian random vector with a non-singular covariance matrix $\mathbf{C}$, satisfying the separation condition (3.1) for some graph $\Omega \subset \{1, \ldots, p\}^2$, then the relation

between Ω and $\mathbf{C}$ is described very elegantly by the following well-known result:

$$\mathbf{P}_{ij} = 0 \text{ if and only if } i \text{ and } j \text{ are not adjacent in } \Omega \tag{3.5}$$

where $\mathbf{P}$ is the inverse of the covariance matrix $\mathbf{C}$. In other words, the zero entries of the matrix $\mathbf{P}$ correspond precisely to missing edges of the graph Ω.

Having an elegant *inverse zero characterization* as (3.5) for kernels is impeded by technical difficulties, however. Namely, the "inverse" of a kernel on an uncountable domain $U \times U$ is not a well-defined notion in general. If we attempt to make the space of kernels into a ring by defining the product of two kernels K_1 and K_2 in a natural way by

$$K_1 \odot K_2(u, v) = \int_U K_1(u, w) K_2(w, v) \, d\mu(u)$$

where μ is a Borel measure on U, then the resulting space ends up being a non-unital ring. This because no kernel can serve as a multiplicative identity the way the identity matrix does for matrices. Even if we admit the Dirac delta $\delta(u - v)$ as the identity, no kernel would admit an inverse. On the other hand, we can directly consider the inverse of the integral operator $\mathbf{K}$ induced by K as

$$\mathbf{K}f(u) = \int_U K(u, v) f(v) \, d\mu(v)$$

and define its support indirectly as follows: $U_1 \times U_2 \subset \operatorname{supp}(\mathbf{K}^{-1})^c$ if for every pair f, g in the range of $\mathbf{K}$ such that $\operatorname{supp} f = U_1$ and $\operatorname{supp} g = U_2$, we have $\langle f, \mathbf{K}^{-1} g \rangle_{L^2(\mu)} = 0$. This parallels the matrix case, which can also be interpreted via quadratic forms $\mathbf{x}^\top \mathbf{P} \mathbf{y}$ involving sparse vectors $\mathbf{x}, \mathbf{y}$. But this too is inconvenient given that $\mathbf{K}^{-1}$ is unbounded in general, leading to delicate conditions on suitable test functions f, g – this, particularly in a statistical context, where $\mathbf{K}$ is to be estimated from finitely many observations, and hence the true RKHS is not identifiable.

Unlike the inverse zero characterization (3.5), the separation equation (3.4) has the virtue of holding true *regardless* of whether U is finite or whether the covariance is boundedly invertible. But this comes at the expense of the condition being tedious to verify since one needs to exhaust all admissible combinations of u, v and W.

In Section 3.4, however, we will show that this shortcoming can be circumvented, by appealing to a notion of *resolution*. Namely, we will show that an analogue of the inverse zero characterization (3.5) holds even for infinite domains U, as long as we are willing to specify the graph Ω up to some finite resolution, and that the characterization behaves coherently under refinement of the resolution.

3.3.2 The Graph of a Stochastic Process

Theorem 3.1 allows us to verify whether the conditional independence structure of a given Gaussian process is compatible with a given graph, in the sense of the separation condition (3.1). But it does not specify the graph, nor does it guarantee the uniqueness of a graph compatible with a Gaussian process X. In the finite-dimensional setting, these questions are answered unequivocally: the zero pattern of the inverse covariance (3.5) defines an adjacency matrix, so the question boils down to invertibility of the covariance.

To address this question, we note that compatibility with the separation equation is inherited with respect to graph inclusion: it is not hard to show that when K satisfies the separation equation (3.4) for a graph Ω then it also does so for every graph $\tilde{\Omega}$ which contains Ω (see Waghmare and Panaretos (2022)). Assume that the index set U of X is a compact subset of $\mathbb{R}^n$ with the natural topology. The previous observation suggests intersecting all compatible graphs to define *the* graph of a process.

Definition 3.1. *We define the graph of X, denoted by Ω_X, to be the intersection of all closed graphs Ω for which the separation equation (3.4) is satisfied by the covariance K of X.*

Unlike the finite-dimensional setting, there is no guarantee that X will satisfy the separation condition (3.1) for $\Omega = \Omega_X$. This may seem dissatisfying given that we would have hoped Ω_X to be interpretable as the "minimal" graph satisfying the separation equation. But it does point to an interesting aspect special to the infinite-dimensional case, namely that satisfaction of the separation equation is not closed under infinite intersections. This means that for certain processes there is simply no "minimal" graph for which the process satisfies the separation condition. The following example illustrates this peculiar feature of infinite dimensions. It also demonstrates how being Markov forces a Gaussian process to satisfy the separation equation.

Example 3.2. *Let $W = \{W_t\}_{t\in I}$ be the Brownian motion process on the unit interval I. Its covariance $K(u,v) = u \wedge v$ satisfies the separation equation for every strip for $\Omega_w = \{(u,v) : |u-v| \leq w\}$, for every $w > 0$. Indeed, if $u, v \in I$ are separated by some subinterval $J \subset I$, then we can assume without loss of generaity that $u > v$ and by the Markov property $W_u = \mathbb{E}[W_u|W_J]$. Then $K(u,\cdot) = \Pi_W K(u,\cdot)$ and the conclusion follows by taking the inner product with $K(\cdot, v)$.*

Consequently, $\Omega_W = \cap_{w>0}\, \Omega_w$ is the empty graph on I given by the diagonal $\{(u,v) : u = v\}$, in which no two vertices are adjacent. If K were to satisfy the separation equation for Ω_W, it would mean that $K(u,v) = 0$, which is contradictory. The same argument can be made for Gaussian processes which are Markov, multiple Markov (Hida and Hitsuda, 1993) or possess analytic covariances covered in Example 3.1.

Determining the conditions under which X satisfies the separation equation for $\Omega = \Omega_X$

seems to be a challenging technical problem interfacing the theory of infinite graphs and the analytical properties of covariances, and is beyond the scope of this chapter. However, if $U \times U$ is equipped with a metric, then one can make up for the gap in intuition resulting from this anomaly by thinking of the conditional independence structure of a process X as being represented by $\Omega_X + \mathbb{B}_\epsilon$ instead of Ω_X where $\Omega_X + \mathbb{B}_\epsilon$ is the ϵ-envelope of Ω_X (see Figure 3.1 (b)). That is, the set of points within ϵ distance from Ω_X where ϵ can be taken to be arbitrarily small. For Gaussian processes on the unit interval which are Markov, multiple Markov or have analytic covariances, the conditional independence structure is then given by an ϵ-strip centered along the diagonal. This "open" formulation rescues the intuition sought in situations like Example 3.2.

The graph Ω_X (or its ϵ-envelope), presents an interesting target for estimation given n independent and identically distributed realizations of X. In Section 3.4, we shall present an analogue of the inverse zero characterization (3.5) for kernels up to a finite resolution, and we shall present sufficient conditions on K for Ω_X to be identifiable exactly or up to such a finite resolution.

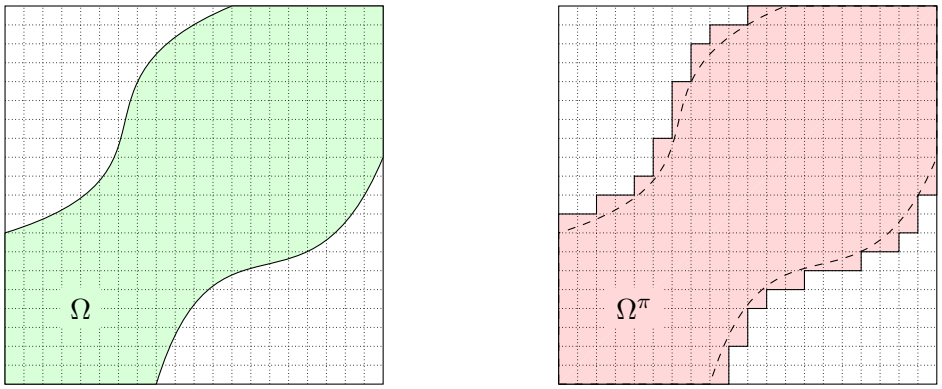
3.4 Resolving Uncountably Infinite Graphs

In this section, we shall recover an analogue of the inverse zero characterization (3.5) for kernels. This will enable us to verify the separation condition in a practically feasible manner, and will also makes allow us to deploy the well-established thresholding approach of graph recovery (developed in Section 3.5).

As previously argued, an *exact* inverse zero characterization is unavailable for our setting and likely infeasible, in light of the distinctly different algebraic properties of kernels in comparison with, matrices. Our approach will thus consist in introducing an appropriate notion of *resolution*, and contenting ourselves with a characterisation valid for any given finite resolution. That being said, we will also require that our characterisation be compatible across refinements of the resolution, and that it identify the true graph as resolution diverges.

From a mathematical point of view, resolving a graph consists in specifying a sequence of constructible approximations thereof. From a statistical point of view, focussing on a finite resolution is arguably natural, or even necessary, since the number of potential graphs is uncountably infinite, and we need to infer the graph from finitely many realizations. Our estimation theory will reflect how the resolution can increase as a function of sample size, thus informing us on how finely we can hope to discern the conditional independence structure of the process from a given amount of finite data.

Our results thus far applied to any covariance kernel K on any set U. From this point onward, we shall additionally assume K to be continuous and U to be a compact subset of



(a) (b)

Figure 3.2: (a) A graph Ω on an interval and (b) its π-resolution approximation Ω^π. Each cell of the grid represents a pixel $U_i \times U_j$ where $U_i, U_j \in \pi$.

$\mathbb{R}^d$ equipped with a Borel measure μ supported on U. The results can be extended without much difficulty to more general sets with topological structure enabling a generalization of Mercer's theorem to apply, however we shall stick to the compact Euclidean setting for simplicity.

3.4.1 Resolution

Let U to be a compact subset of $\mathbb{R}^d$ equipped with a Borel measure μ supported on U. Let $K : U \times U \to \mathbb{R}$ be a continuous covariance kernel. We shall now make precise what we understand by the term *resolution* in this context.

A *partition* π of U is a finite collection $\{U_j\}_{j=1}^p$ such that (a) U_j are Borel subsets of U such that $\mu(\tilde{U}) > 0$ for every nonempty subset $\tilde{U} \subset U_j$ which is relatively open in U_j, (b) which are *exhaustive* in that $\cup_{j=1}^p U_j = U$ and (c) *disjoint* in that $U_i \cap U_j = 0$ for $i \neq j$. The additional technical conditions in (a) simply ensure that Mercer's theorem applies to U_j individually as it does to U as a whole. In common mathematical parlance, a partition need not be finite nor contain only Borel sets but using the above definition lends brevity to our presentation.

We shall refer to sets of the form $U_i \times U_j$ for $1 \leq i, j \leq p$ as *pixels*. A *π-resolution graph* $\Omega \subset U \times U$ is a union of pixels which includes the pixels on the diagonal, that is, $\cup_{j=1}^p U_j \times U_j \subset \Omega$. Every graph Ω on U admits what we shall call the *best π-resolution approximation* Ω^π which we define as the intersection of all π-resolution graphs on U which contain Ω. Thus, Ω^π is the smallest π-resolution graph which contains Ω. Alternatively, we can express Ω^π as the union of all $U_i \times U_j$ which intersect with Ω. As above, we shall denote the best π-resolution approximation of a graph Ω on U by Ω^π. Figure 3.2

illustrates the difference between Ω and Ω^π.

We shall denote by $\tilde{\Omega}^\pi$ the intersection $\cap_{\epsilon>0}\,(\Omega+\mathbb{B}_\epsilon)^\pi$ where $\mathbb{B}_\epsilon$ denotes the intersection of the Euclidean ball of radius ϵ in $\mathbb{R}^{2d}$ with $U\times U$ and the sum $A+B$ denotes the set $\{a+b : a\in A \text{ and } b\in B\}\cap U\times U$. Because the sets "decrease" as $\epsilon\to 0$ in that $\Omega_X+\mathbb{B}_{\epsilon_1}\subset\Omega_X+\mathbb{B}_{\epsilon_2}$ for $\epsilon_1<\epsilon_2$, we can also write $\tilde{\Omega}^\pi$ as $\lim_{\epsilon\to 0}\,(\Omega_X+\mathbb{B}_\epsilon)^\pi$. The distinction between Ω^π and $\tilde{\Omega}^\pi$ is mainly technical and is a consequence of the fact observed in Example 3.2 that for certain processes there is no minimal graph Ω for which the covariance satisfies the separation equation. For this reason and for lack of a better alternative, we shall refer to both Ω^π and $\tilde{\Omega}^\pi$ as the best π-resolution approximation of Ω while indicating which of the two we mean by their respective symbols.

Example 3.3. *A simple instance of how $\tilde{\Omega}_X$ can differ from Ω_X^π is given by the processes considered in Examples 3.1 and 3.2, where $\Omega_X=\{(u,v) : u=v\}$. Thus, $\Omega_X^\pi=\cup\{U_i\times U_j : |i-j|=0\}$ but $\tilde{\Omega}_X^\pi=\cup\{U_i\times U_j : |i-j|\le 1\}$ since the strip $\{(u,v) : |u-v|<\epsilon\}$ always intersects the pixels $U_i\times U_j$ for which $|i-j|=1$.*

3.4.2 Approximate Inverse Zero Characterization

We shall now show how one can recover the best π-resolution approximation of Ω from the covariance kernel $K(s,t)=\mathbb{E}[X_sX_t]$ of X. Let $K_{ij}=K|_{U_i\times U_j}$. For $1\le i,j\le p$, let $\mathbf{K}_{ij}:L^2(U_j,\mu)\to L^2(U_i,\mu)$ be the integral operator induced by the integral kernel K_{ij} given by

$$\mathbf{K}_{ij}f(u)=\int_{U_j}K_{ij}(u,v)f(v)\;d\mu(v)$$

Define the *covariance operator matrix* $\mathbf{K}_\pi$ induced by the partition π as $\mathbf{K}_\pi=[\mathbf{K}_{ij}]_{i,j=1}^p$. Furthermore, we define the *correlation operator matrix* $\mathbf{R}_\pi$ induced by the partition π as $\mathbf{R}_\pi=[\mathbf{R}_{ij}]_{i,j=1}^p$ specified entrywise by $\mathbf{R}_{ij}=\mathbf{K}_{ii}^{-1/2}\mathbf{K}_{ij}\mathbf{K}_{jj}^{-1/2}$. Alternatively, we can write $\mathbf{R}_\pi$ as $\mathbf{R}_\pi=[\mathrm{dg}\,\mathbf{K}_\pi]^{-1/2}\mathbf{K}_\pi[\mathrm{dg}\,\mathbf{K}_\pi]^{-1/2}$. If $\mathbf{R}_\pi$ is invertible and then we can define the *precision operator matrix* $\mathbf{P}_\pi=[\mathbf{P}_{ij}]_{i,j=1}^p$ as the inverse of $\mathbf{R}_\pi$, that is $\mathbf{P}_\pi=\mathbf{R}_\pi^{-1}$.

The key result is now stated as follows:

Theorem 3.2. *If $\mathbf{R}_\pi$ is invertible, then the graph Ω_X and the precision operator matrix $\mathbf{P}_\pi$ induced by the partition π are related as:*

$$\tilde{\Omega}_X^\pi\equiv\lim_{\epsilon\to 0}\,(\Omega_X+\mathbb{B}_\epsilon)^\pi\subset\cup\,\{U_i\times U_j : \|\mathbf{P}_{ij}\|\neq 0\}. \tag{3.6}$$

If, in addition, for every $\epsilon>0$ there exists a partition π_ϵ of U such that every pixel is contained within a ball of radius ϵ and $\mathbf{R}_{\pi_\epsilon}$ is invertible, then the above relation is an equality. In other words, $\tilde{\Omega}_X^\pi$ is same as the union of $U_i\times U_j$ for (i,j) such that $\mathbf{P}_{ij}\neq\mathbf{0}$.

Thus by discerning which entries of the partition-induced correlation operator matrix are zero, one can work out a finite resolution approximation $\tilde{\Omega}_X^\pi$ of Ω_X. It follows immediately

that $\tilde{\Omega}_X^\pi$ is identifiable if $\mathbf{P}_\pi$ is invertible. We expect that the technical condition for equality is an artifact of our proof technique, and not an essential feature of the problem.

3.4.3 Refinement and Identifiability

If we know Ω^{π_1} and Ω^{π_2} then we can get a finer approximation of Ω by simply taking their intersection. The resulting graph $\Omega^{\pi_1 \wedge \pi_2} = \Omega^{\pi_1} \cap \Omega^{\pi_2}$ is the best $(\pi_1 \wedge \pi_2)$-approximation where the partition $\pi_1 \wedge \pi_2$ is the *refinement* of the partitions π_1 and π_2 given by $\{U_1 \cap U_2 : U_1 \in \pi_1 \text{ and } U_2 \in \pi_2\}$ which is in other words composed of the intersections of the sets in the original partitions. We shall say that π_2 is *finer* than π_1 if $\pi_2 = \pi_1 \wedge \pi_2$. We can define the refinement of a countable number of partitions $\{\pi_j\}_{j=1}^\infty$ as

$$\wedge_{j=1}^\infty \pi_j = \{\cap_{j=1}^\infty U_j : U_j \in \pi_j \text{ for } j \geq 1\}$$

and thus if we know Ω^{π_j} for $j \geq 1$ then the best π-resolution approximation for $\pi = \wedge_{j=1}^\infty \pi_j$ is given by $\Omega^\pi = \cap_{j=1}^\infty \Omega^{\pi_j}$. Additionally, we shall say that the partitions $\{\pi_j\}_{j=1}^\infty$ *separate points on* U if $\wedge_{j=1}^\infty \pi_j = \{\{u\} : u \in U\}$.

We shall say that Ω_X is *identifiable up to* π*-resolution* if its best π-resolution approximation $\tilde{\Omega}_X^\pi$ is identifiable. Moreover, we shall say that Ω_X is identifiable *exactly* if its closure in U is identifiable. In essence, the distinction between Ω_X and its closure does not concern us here, nor is it amenable to our method. The following corollary is now almost immediate from Theorem 3.2 and gives sufficient conditions for identifiability of Ω_X.

Corollary 3.1. *Let X be a Gaussian process on U with a continuous covariance. If π is a partition of U such that the correlation operator $\mathbf{R}_\pi$ is invertible, then Ω_X is identifiable up to π-resolution.*

Furthermore, if there exists a sequence $\{\pi_j\}_{j=1}^\infty$ of partitions on U such that (a) the correlation operators $\mathbf{R}_{\pi_j}$ are invertible and (b) the partitions separate points on U, then Ω_X is identifiable exactly.

The criteria for exact identifiability may appear to be too demanding but they are required only for an *infinite* resolution or *exact* identifiability of Ω. For applications, we can always content ourselves with identifiability up to π-resolution for a reasonably fine partition π which would only require that the correlation operator $\mathbf{R}_\pi$ induced by π be invertible.

3.4.4 Relation to Functional Graphical Models

Consider the *functional graphical model* introduced in Qiao et al. (2019) in which the set of vertices consists of $\mathbf{X} = (X_1, \ldots, X_p)$ where every X_k is a random real-valued function

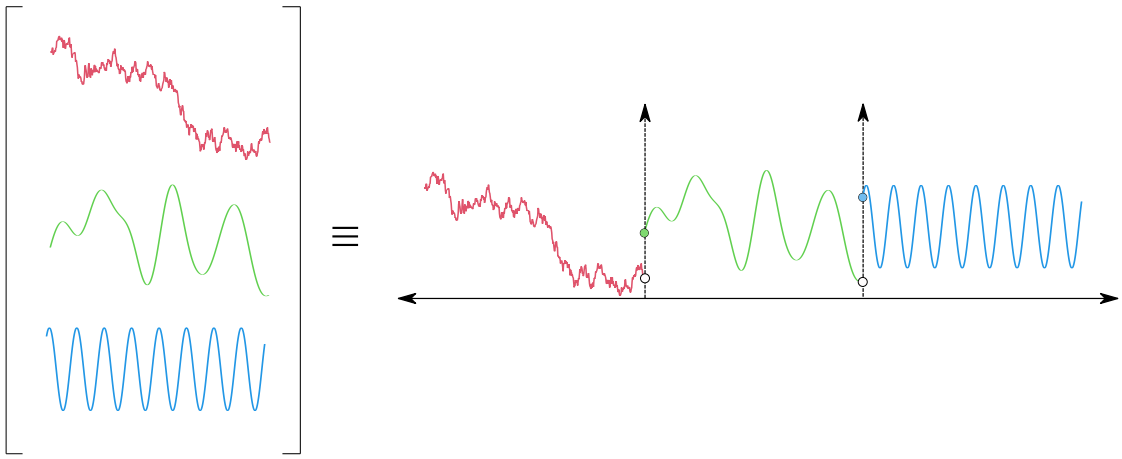


Figure 3.3: A functional graphical model can be seen as a single stochastic process by concatenating successive vector components.

on an interval I_k and there is an edge between X_i and X_j unless

$$\mathrm{Cov}[X_i(u), X_j(v)|X_k(w) \text{ for } k \neq i, j \text{ and } w \in I_k] = 0 \text{ for } u \in I_i \text{ and } v \in I_j.$$

If we define

$$U = \bigsqcup_{j=1}^{p} I_j = \cup_{j=1}^{p} \{j\} \times I_j$$

to be the *disjoint union* of $\{I_1, \ldots, I_p\}$, the vector-valued function $\mathbf{X} = (X_1, \ldots, X_p)$ can be thought of as a single real-valued stochastic process $X = \{X_u : u \in U\} = \{X_j(t) : 1 \leq j \leq p\,,\, t \in I_j\}$ indexed by both j and t. This can be visualized by serially concatenating successive vector components (see Figure 3.3) and the set U can thus be thought of as a compact subset of $\mathbb{R}$. Recovering the graph of $\mathbf{X}$ in the functional sense reduces to recovering the graph of Ω_X in the uncountably indexed sense, but only up to a specific π-resolution, namely where the partition π consists of the sets $\{(j, I_j)\}_{j=1}^{p}$. Thus, $\Omega_{\mathbf{X}} \equiv \tilde{\Omega}_X^{\pi}$. This restriction highlights the fact that functional graphical models concern interactions *between* the random functions $\{X_j\}_{j=1}^{p}$ and not with interactions *within* a random function X_j – the latter requires the notion of coherently resolving an uncountable graph. Furthermore, in the same vein, it shows that functional graphical models can be cast as special cases of our more general uncountably indexed graphical models.

3.5 Graph Recovery from Sample Paths

Given a partition π of the index set $U \subset \mathbb{R}^d$, we now present our approach to the problem of recovering the graph Ω_X of a process X given n independent realizations $\{X^k\}_{k=1}^n$, up to resolution π. Equivalently, this amounts to determining which of the entries of the π-induced precision operator matrix $\mathbf{P}_\pi = \mathbf{R}_\pi^{-1}$ are zero based on $\{X^k\}_{k=1}^n$.

Evidently, for the last statement to make sense at all, we must assume that $\mathbf{R}_\pi$ is indeed invertible. Consequently, any consistent estimator of $\mathbf{R}_\pi$ based on a sample of size n will also be eventually invertible w.r.t. n, almost surely. Whenever the inverse of such an estimator appears, it is implicit that n is sufficiently large.

Since the partition π that induces the operators $\mathbf{K}_\pi$, $\mathbf{R}_\pi = \mathbf{P}_\pi^{-1}$ is the same, we shall denote these operators simply as $\mathbf{K}$, $\mathbf{R}$, and $\mathbf{P} = \mathbf{R}^{-1}$ whenever there is no danger of confusion. By writing $\mathbf{K} = \operatorname{dg} \mathbf{K} + \mathbf{K}_0$, the correlation operator matrix can be expressed as

$$\mathbf{R} = \mathbf{I} + [\operatorname{dg} \mathbf{K}]^{-1/2} \mathbf{K}_0 [\operatorname{dg} \mathbf{K}]^{-1/2}.$$

Thus the diagonal entries $\mathbf{R}_{ii}$ of $\mathbf{R}$ are all equal to identity and we need not burden ourselves with their estimation. Furthermore, since we are effectively trying to invert the compact operator $\operatorname{dg} \mathbf{K}$, regularization is necessary, which we do by adding a ridge of size κ. Once an estimator of the precision operator matrix is formed, we threshold it entrywise in operator norm to estimate Ω_X.

In summary, the estimation procedure consists of the following two steps:

Step 1. Estimation. We estimate the mean vector $\mathbf{m} = [m_j]_{j=1}^p$, the covariance operator matrix $\hat{\mathbf{K}} = [\hat{\mathbf{K}}_{ij}]_{i,j=1}^p$, and the correlation matrix $\hat{\mathbf{R}} = [\hat{\mathbf{R}}_{ij}]_{i,j=1}^p$ corresponding to the partition $\pi = \{U_j\}_{j=1}^p$ by

$$\begin{aligned} \hat{m}_j &:= \sum_{k=1}^n X_{U_j}^k \\ \hat{\mathbf{K}}_{ij} &:= \frac{1}{n} \sum_{k=1}^n \left[X_{U_i}^k - \hat{m}_i \right] \otimes \left[X_{U_j}^k - \hat{m}_j \right] \\ \hat{\mathbf{R}} &:= \mathbf{I} + [\kappa \mathbf{I} + \operatorname{dg} \hat{\mathbf{K}}]^{-1/2} \hat{\mathbf{K}}_0 [\kappa \mathbf{I} + \operatorname{dg} \hat{\mathbf{K}}]^{-1/2}, \end{aligned}$$

for a ridge parameter $\kappa > 0$.

Step 2. Thresholding. The estimate $\hat{\Omega}_\pi$ of the best π-resolution approximation $\tilde{\Omega}_X^\pi$ is calculated as

$$\hat{\Omega}_\pi = \cup \{U_i \times U_j : \|(\hat{\mathbf{R}}^{-1})_{ij}\| > \rho\}$$

for a thresholding parameter $\rho > 0$.

There are two tuning parameters involved in the procedure: the ridge κ, and the threshold ρ. Their choice is guided via our asymptotic theory (see the next Section), in relation to the sample size n and the partition size p (the partition π will typically be a regular partition into p intervals of equal length). Practical rules for their choice are discussed in Section 3.7.

We remark that the ridge estimator of the correlation operator matrix in Step 2(b) is essentially the same as the estimator introduced by Li and Solea (2018a) in the context of graphical models for random vectors with Hilbertian entries, adapted to our setting. Though the context is somewhat different, there are direct parallels to be drawn, and we hence compare to their asymptotic analysis in the next section.

3.6 Large Sample Theory

Developing asymptotic guarantees for our procedure will rely on controlling the estimation error for the entries of the precision operator matrix in operator norm. As remarked in the previous section, the ridge estimator is of the same form as in Li and Solea (2018a), and thus we opt to work with the same regularity conditions. We improve upon their results in two ways, however. Firstly, we derive both improved and simplified rates of convergence for the estimation of the correlation operator. Secondly, under the assumption that X is a sub-Gaussian random element in some Hilbert space, we derive concentration bounds for the estimated correlation and precision operator matrices, along with a tail bound on the precision operator matrix. Taken in combination, these results allow us to then establish consistency and rates for our graph recovery method, quantifying what resolutions can be attained at given sample sizes.

3.6.1 Rates and Bounds

Recall that we defined our estimator of the correlation operator matrix as

$$\hat{\mathbf{R}} = \mathbf{I} + [\kappa_n \mathbf{I} + \operatorname{dg} \hat{\mathbf{K}}]^{-1/2} \hat{\mathbf{K}}_0 [\kappa_n \mathbf{I} + \operatorname{dg} \hat{\mathbf{K}}]^{-1/2}. \tag{3.7}$$

for $\hat{\mathbf{K}}$ our estimator of the covariance operator matrix, and κ_n the regularization parameter. The error $\hat{\mathbf{R}} - \mathbf{R}$ of estimating $\mathbf{R}$ using $\hat{\mathbf{R}}$ can be split into estimation error $\mathcal{E} = \hat{\mathbf{R}} - \mathbf{R}_e$ (related to variance) and approximation error $\mathcal{A} = \mathbf{R}_e - \mathbf{R}$ (related to bias).

To control the approximation error, we will require the following regularity condition on $\mathbf{R}$:

Assumption 1. *For some bounded operator matrix Φ_0 with all the diagonal entries zero and $\beta > 0$, we have*

$$\mathbf{R}_0 = [\operatorname{dg} \mathbf{K}]^{\beta} \Phi_0 [\operatorname{dg} \mathbf{K}]^{\beta}. \tag{3.8}$$

Note that this implies that $\mathbf{R}_0$ is compact. From an inverse problems perspective, the assumption simply ensures that $\mathbf{K}_0 = [\operatorname{dg}\mathbf{K}]^{1/2+\beta}\Phi_0[\operatorname{dg}\mathbf{K}]^{1/2+\beta}$ is linearly well-conditioned for inversion by $[\operatorname{dg}\mathbf{K}]^{1/2}$.

Our first result now relates $\|\hat{\mathbf{R}} - \mathbf{R}\|$ to $\|\hat{\mathbf{K}} - \mathbf{K}\|$, $\mathbf{K}$ and $\|\mathbf{R}\|$:

Theorem 3.3 (Bounding $\|\hat{\mathbf{R}} - \mathbf{R}\|$)**.** *Under Assumption 1, given any sequences $\kappa_n > 0$ and $\delta_n \geq \|\hat{\mathbf{K}} - \mathbf{K}\|$, we have*

$$\|\hat{\mathbf{R}} - \mathbf{R}\| \leq \|\mathcal{A}\| + \|\mathcal{E}\| \leq 5 \cdot \|\mathbf{R}\| \cdot \Big[(\delta_n/\kappa_n)^2 + (\delta_n/\kappa_n)\Big] + 2 \cdot \kappa_n^{\beta\wedge 1} \cdot \|\Phi_0\| \cdot \|\mathbf{K}\|^{2\beta-\beta\wedge 1}.$$

The estimator $\hat{\mathbf{R}}$ is consistent so long as the regularization parameter κ_n is chosen such that $\kappa_n \to 0$ and $\delta_n/\kappa_n \to 0$ as $n \to \infty$. The optimal rate is given by

$$10 \cdot (\|\mathbf{R}\| \vee \|\Phi_0\|\|\mathbf{K}\|^{2\beta-\beta\wedge 1}) \cdot \delta_n^{\frac{\beta\wedge 1}{1+\beta\wedge 1}}$$

and it is achieved for the choice $\kappa_n \asymp \delta_n^{\frac{1}{1+\beta\wedge 1}}$.

In fact, the theorem is valid for any choice of estimator $\hat{\mathbf{K}}$, provided that it is non-negative definite (for a suitable δ_n, of course). Under our specific choice of $\hat{\mathbf{K}}$ as an empirical covariance, the central limit theorem yields $\|\hat{\mathbf{K}} - \mathbf{K}\| = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$. So we can substitute $\mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ for δ_n and obtain the following rate of convergence for the estimator of the correlation operator matrix:

Corollary 3.2 (Rate of Convergence for $\hat{\mathbf{R}}$)**.** *Under Assumption 1, the optimal choice of the regularization parameter is given by $\kappa \asymp n^{-\frac{1}{2}\cdot\frac{1}{1+\beta\wedge 1}}$ and we have*

$$\|\hat{\mathbf{R}} - \mathbf{R}\| = \mathcal{O}_{\mathbb{P}}\Big(n^{-\frac{1}{2}\cdot\frac{\beta\wedge 1}{1+\beta\wedge 1}}\Big).$$

Note that when β ranges in $(0, 1/2]$, the above rate is strictly better than the rate $n^{-\frac{2\beta}{5+2\beta}}$. derived in Li and Solea (2018a), and the two rates coincide when $\beta > 1/2$. In addition to slightly improving the rate of convergence for poorly conditioned $\mathbf{R}_0$ corresponding to $\beta < 1/2$, this implies that the apparent phase transition at $\beta = 1/2$ observed in the rates of Li and Solea (2018a) is an artefact of their analysis. The only transition we observe in the convergence is at $\beta = 1$ as for $\beta > 1$, the rate is same as that for $\beta = 1$ which is $n^{-1/4}$. However, the dependence on $\|\mathbf{K}\|$ does change, as observed in Theorem 3.3.

Turning to the precision operator matrix, recall that for $\mathbf{P} := \mathbf{R}^{-1}$ to be well defined at all, we need $\mathbf{R}$ to be strictly positive definite. The following assumption is only slightly stronger, and represents the non-compact counterpart of the familiar assumption that eigenvalues are separated from 0:

Assumption 2. *The spectrum of $\mathbf{R}_0$ satisfies $r = 1 + \inf\sigma(\mathbf{R}_0) > 0$.*

Under assumption 2, $\mathbf{R}$ is certainly strictly positive. Consequently, in the context of Corollary 3.2, the operator $\hat{\mathbf{R}}$ is strictly positive for all sufficiently large n, by virtue of being consistent. Hence, for all sufficiently large n, we may write

$$\hat{\mathbf{P}} - \mathbf{P} = \hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{R}^{-1} - \hat{\mathbf{R}}^{-1}\hat{\mathbf{R}}\mathbf{R}^{-1} = \hat{\mathbf{R}}^{-1}\left[\mathbf{R} - \hat{\mathbf{R}}\right]\mathbf{R}^{-1} = \hat{\mathbf{P}}\left[\mathbf{R} - \hat{\mathbf{R}}\right]\mathbf{P}. \tag{3.9}$$

Since $\hat{\mathbf{P}}$ is a random quantity, bounding $\|\hat{\mathbf{P}} - \mathbf{P}\|$ using (3.9) requires us to find a bound for $\|\hat{\mathbf{R}} - \mathbf{R}\|$, as well as $\|\hat{\mathbf{P}}\|$. It was shown in Li and Solea (2018a), that $\|\hat{\mathbf{P}}\|$ is bounded in probability under Assumption 2. As a result, the convergence rates for $\|\hat{\mathbf{R}} - \mathbf{R}\|$ also apply to $\|\hat{\mathbf{P}} - \mathbf{P}\|$.

Corollary 3.3 (Rate of Convergence for $\hat{\mathbf{P}}$)**.** *Under the Assumption 1 and 2, the optimal choice of the regularization parameter is given by* $\kappa \asymp n^{-\frac{1}{2}\cdot\frac{1}{1+\beta\wedge 1}}$ *and we have*

$$\|\hat{\mathbf{P}} - \mathbf{P}\| = \mathcal{O}_{\mathbb{P}}\Big(n^{-\frac{1}{2}\cdot\frac{\beta\wedge 1}{1+\beta\wedge 1}}\Big).$$

We shall now use basically the same principle to derive concentration bounds for $\|\hat{\mathbf{P}} - \mathbf{P}\|$.

Remark 3.1. *It is worth mentioning that our assumptions are rather minimal. It is well known in inverse problems literature that the rate of convergence of the solution of a linear inverse problem can be arbitrarily slow in the absence of any regularity such as that provided by Assumption 1. On the other hand, Assumption 2 is necessary if we are to connect the empirical covariance with the graph of the process via Theorem 3.2. Though it has occasionally been claimed in the literature that* $\mathbf{R}$ *always admits a eigenvalue gap (i.e. that* $\mathbf{R} \geq c\mathbf{I}$ *for some* $c > 0$*), this is not true as the following simple counterexample illustrates: take* $\mathbf{K} = [\mathbf{K}_{ij}]_{i,j=1}^2$ *to be given by* $\mathbf{K}_{11} = \mathbf{K}_{22} = \sum_j \lambda_j e_j \otimes e_j$ *and* $\mathbf{K}_{12} = \mathbf{K}_{21} = -\lambda_1 e_1 \otimes e_1$*. Then* $\mathbf{R} = [\mathbf{R}_{ij}]_{i,j=1}^2$ *given by* $\mathbf{R}_{11} = \mathbf{R}_{22} = \mathbf{I}$ *and* $\mathbf{R}_{12} = \mathbf{R}_{21} = -e_1 \otimes e_1$ *is not invertible since* $\mathbf{R}[e_1\ e_1] = \mathbf{0}$*. The same counterexample shows that intervibility itself of* $\mathbf{R}$ *cannot be secured by requiring* $\operatorname{Ker}\mathbf{K}_{jj} = \{\mathbf{0}\}$*.*

In order to derive concentration bounds on the correlation and precision operator matrices, we exploit a concentration bound in the operator norm which is a consequence of Theorem 9 from Koltchinskii and Lounici (2017). The results can be extended effortlessly to random elements in Banach spaces but in the interest of a simpler presentation we shall refrain form doing so.

Theorem 3.4. *Let* X *be a sub-Gaussian random element in a Hilbert space, with mean zero and covariance operator* $\mathbf{K}$*. Let* $X_1, \ldots, X_n$ *be i.i.d. replications of* X*. Define the empirical covariance operator* $\hat{\mathbf{K}} = \frac{1}{n}\sum_{j=1}^n X_j \otimes X_j$*. For every* $0 < t \leq \|\mathbf{K}\|$*,*

$$\mathbb{P}\{\|\hat{\mathbf{K}} - \mathbf{K}\| \geq t\} \leq e^{-cnt^2/\|\mathbf{K}\|^2}$$

for $n \geq (1 \vee \mathbf{r}(\mathbf{K}))\|\mathbf{K}\|^2/t^2$ *where* $\mathbf{r}(\mathbf{K}) = (\mathbb{E}\|X\|)^2/\|\mathbf{K}\|$ *and* c *is a universal constant.*

Using our earlier results, we can now derive concentration bounds for $\|\hat{\mathbf{R}}-\mathbf{R}\|$ and$\|\hat{\mathbf{P}}-\mathbf{P}\|$ and a tail bound for $\hat{\mathbf{P}}$, which will eventually enable us to prove the consistency of our graph recovery procedure:

Theorem 3.5 (Concentration and Tail Bounds)**.** *Let X be a stochastic process on the set U corresponding to a sub-Gaussian random element in the Hilbert space $L^2(U,\mu)$ with the covariance operator $\mathbf{K}$. Let c_K be the universal constant c appearing in Theorem 3.4, $\rho_K = \|\mathbf{K}\|$, $n_K = [1 \vee \mathbf{r}(\mathbf{K})]\|\mathbf{K}\|^2$,*

$$M_R = 10 \cdot \left[\|\mathbf{R}\| \vee \|\Phi_0\|\|\mathbf{K}\|^{2\beta-\beta\wedge 1}\right] \text{ and } r = \inf_j [1+\lambda_j(\mathbf{R}_0)] = \|\mathbf{P}\|^{-1}.$$

Define, $c_R = c_K M_R^{2+2/\beta\wedge 1}$, $\rho_R = M_R \rho_K^{\beta\wedge 1/(\beta\wedge 1+1)}$, $n_R = n_K M_R^{2+2/\beta\wedge 1}$ and $c_P = c_R(r^2/2)^{2+2/\beta\wedge 1}$.

1. *Under Assumption 1, we have*

$$\mathbb{P}[\|\hat{\mathbf{R}}-\mathbf{R}\| > \rho] \le \exp\left\{-c_R n \rho^{2+2/\beta\wedge 1}\right\} \tag{3.10}$$

   for $0 < \rho < \rho_R$ and $n > n_R/\rho^{2+2/\beta\wedge 1}$.

2. *Under Assumptions 1 and 2, we have*

$$\mathbb{P}[\|\hat{\mathbf{P}}\| > (r-\rho)^{-1}] \le \exp\left\{-c_R n \rho^{2+2/\beta\wedge 1}\right\} \tag{3.11}$$

   for $0 < \rho < r \wedge \rho_R$ and $n > n_R/\rho^{2+2/\beta\wedge 1}$.

3. *Under Assumptions 1 and 2, we have*

$$\mathbb{P}[\|\hat{\mathbf{P}}-\mathbf{P}\| > \rho] \le 2 \cdot \exp\left\{-c_P n \rho^{2+2/\beta\wedge 1}\right\} \tag{3.12}$$

   for $0 < \rho < (r/2) \wedge \rho_R$ and $n > n_R/\rho^{2+2/\beta\wedge 1}$.

Note that the parameters ρ_K and n_K depend only on the covariance kernel K whereas the parameters c_R, c_P, ρ_R, M_R, r and n_R depend only on K and π.

3.6.2 Consistent Graph Recovery

We can now have the tools to establish sufficient conditions for the estimator $\hat{\Omega}_X^\pi$ of $\tilde{\Omega}_X^\pi$ to be consistent.

Theorem 3.6 (Consistency at Given Resolution)**.** *Let X be a Gaussian process on U with continuous covariance kernel K, corresponding to a (Gaussian) random element in the Hilbert space $L^2(U,\mu)$. Let $\{X^k\}_{k=1}^n$ be n independent copies of X and π be a*

partition on U. Under Assmptions 1 and 2, we have for $0 < \rho < \frac{1}{2}r \wedge \rho_R \wedge \rho_P$ and $n > n_R/\rho^{2+2/\beta\wedge 1}$,

$$\mathbb{P}[\hat{\Omega}_X^\pi \neq \tilde{\Omega}_X^\pi] \leq 2p^2 \cdot \exp\left[-c_P n\rho^{2+2/(\beta\wedge 1)}\right] \to 0 \text{ as } n \to \infty$$

where p is the cardinality of π, $\rho_P = \frac{1}{2}\min\{\|\mathbf{P}_{ij}\| : \mathbf{P}_{ij} \neq \mathbf{0}\}$ and the parameters ρ_R, n_R and c_P are as in Theorem 3.5 and depend only on K and π.

Alternatively, for the probability $\mathbb{P}[\hat{\Omega}_X^\pi \neq \tilde{\Omega}_X^\pi]$ to be less than some $\alpha \in (0,1)$, we need the sample size n to satisfy

$$n > \frac{1}{c_P}[\tfrac{1}{2}r \wedge \rho_R \wedge \rho_P]^{-2-2/\beta\wedge 1} \log\left[\frac{2p^2}{\alpha}\right].$$

Notice that even if the thresholding parameter is chosen as a function of the sample size, as in $\rho \equiv \rho(n)$, then the estimator is consistent so long as $n\rho_n^{2+2/\beta\wedge 1} \to \infty$ as $n \to \infty$. Regardless, Theorem 3.6 guarantees exact recovery of $\tilde{\Omega}_X^\pi$ with high probability so long as the thresholding parameter ρ is fixed to be small enough and the sample size n is large enough. It is in contrast to the asymptotic results of Li and Solea (2018a) in which the thresholding parameter needs to decrease as the sample size increases for consistent recovery of the graph and we do not know how quickly $\mathbb{P}[\hat{\Omega}_X^\pi \neq \tilde{\Omega}_X^\pi]$ converges to 0 in terms of the sample size.

A natural question now is: at how fine a resolution p can we estimate the graph Ω_X reliably from a given sample size n? Put differently, how can we refine our partition π as the sample size n increases to construct a consistent estimator for the graph Ω_X itself? Let $\{\pi_j\}_{j=1}^\infty$ be partitions on U which separate points and $\{\alpha_j\}_{j=1}^\infty \subset \mathbb{R}$ be such that $\sum_{j=1}^\infty \alpha_j < \infty$. For every $j \geq 1$, let $\hat{\Omega}_j$ denote the estimator $\hat{\Omega}_X^{\pi_j}$ constructed only using the sample $\{X_k\}_{k=1}^{n_j}$ with an admissible values of the threshold ρ_j according to Theorem 3.6 where the parameter n_j has been chosen to be the smallest n such that

$$n > \frac{1}{c_{P_j}}[\tfrac{1}{2}r_j \wedge \rho_{R_j} \wedge \rho_{P_j}]^{-2-2/\beta_j\wedge 1} \log\left[\frac{2p_j^2}{\alpha_j}\right]. \tag{3.13}$$

Here, p_j is the cardinality of π_j while β_j, r_j, ρ_{R_j}, ρ_{P_j} and c_{P_j} are the parameters β, r, ρ_R, ρ_P and c_P corresponding to the correlation operator $\mathbf{R} = \mathbf{R}_{\pi_j}$. Essentially, we are saying that for larger sample sizes $n > n_j$ we can recover Ω_X to higher resolution p_j with an eventually decreasing probability of failure α_j since $\alpha_j \to 0$ as $j \to \infty$. We now have the following result.

Theorem 3.7 (Consistency under Resolution Refinement)**.** *Let X be a Gaussian process on a compact set $U \subset \mathbb{R}$ with the continuous covariance K corresponding to a (Gaussian) random element in the Hilbert space $L^2(U,\mu)$. Let $\{X^k\}_{k=1}^n$ be independent copies of X and $\{\pi_j\}_{j=1}^\infty$ be partitions on U which separate points such that: (a) π_{j+1} is finer than*

π_j for every $j \geq 1$ and (b) the associated correlation operators $\mathbf{R}_{\pi_j}$ satisfy Assumptions 1 and 2. Then for $\hat{\Omega}_j$ as defined before,

$$\lim_{n\to\infty} \hat{\Omega}_{\max\{j:n_j<n\}} = \Omega_X \text{ almost surely.}$$

In other words, $\Omega_{\max\{j:n_j<n\}}$ is a consistent estimator of Ω_X

3.7 Finite Sample Implementation and Performance

To implement the procedure in practice, one needs to specify the partition π, the ridge κ, and the threshold ρ, and we now discuss this specification in a finite-sample context (as opposed to a large sample context, as in the previous sections).

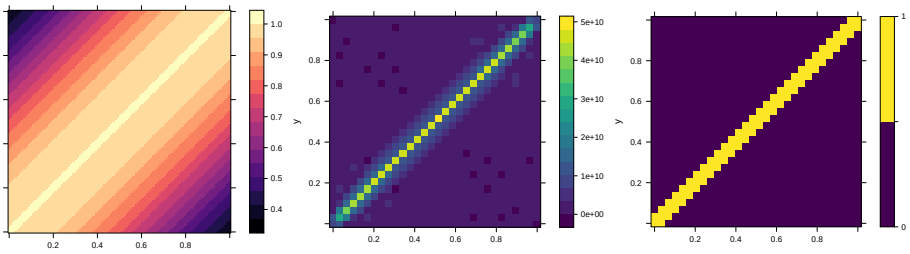
- **Partition**. The choice of partition π is in principle up to the analyst, based on which regions of the domain one is interested to probe for conditional independencies. In most cases, one will work with a regular partition (p contiguous subintervals of $[0, 1]$ of equal length). In any practical setting involving measurement/computation on a grid, it is clear that the finest possible partition is *de facto* that grid. If the paths are sampled very densely (high frequency) relative to the sample size, then it is judicious to not use the finest possible grid as per our large sample theory. In any case, one can also adopt a scale-space approach and consider multiple values of p, searching for persistent zero patterns in the associated correlation operator matrices.

- **Ridge**. The ridge parameter κ ensures that the sample counterpart $[\kappa\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2}$ of $[\operatorname{dg}\mathbf{K}]^{-1/2}$ is stable to sampling variation, in view of the inversion operation. A classical – if computationally intensive– approach is to employ generalized cross validation to make this choice (as in Li and Solea (2018a)). A simpler strategy is to instead choose κ so as to minimize

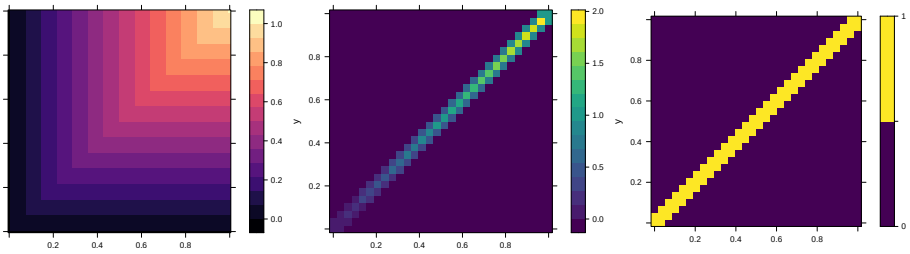
$$\frac{\|(\operatorname{dg}\hat{\mathbf{K}})(\kappa\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}})^{-1}(\operatorname{dg}\hat{\mathbf{K}}) - \operatorname{dg}\hat{\mathbf{K}}\|}{\|\operatorname{dg}\hat{\mathbf{K}}\|}. \tag{3.14}$$

  The justification of this rule is simple: we seek a value of κ which makes $(\kappa\mathbf{I}+\operatorname{dg}\hat{\mathbf{K}})^{-1}$ an approximate generalised inverse of $\operatorname{dg}\hat{\mathbf{K}}$. This selection rule with a search grid of the form $\kappa \in \{10^{-j}\|\hat{\mathbf{K}}\| : 1 \leq j \leq 15\}$ seems to work well in our simulation study, whereas tuning κ more finely does not improve results significantly.

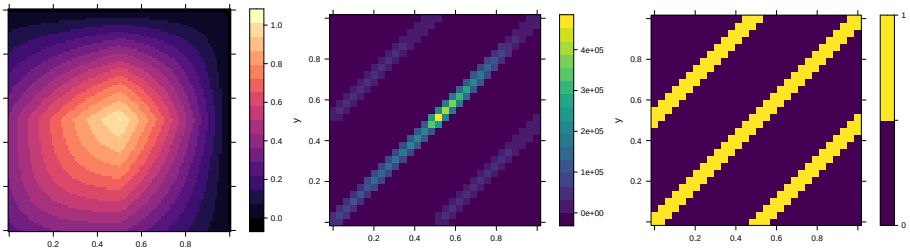
  Notice that the value of (3.14) does not decrease monotonically as κ gets smaller because $\operatorname{dg}\hat{\mathbf{K}}$ is not invertible. Instead, it eventually increases, thus leading to a U-shaped curve with a minimum. This minimum corresponds to the operator $(\kappa\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}})^{-1}$ whose action on $\operatorname{dg}\hat{\mathbf{K}}$ resembles that of a (generalised) inverse the most, over all choices of κ. Roughly speaking, this amounts to choosing κ such



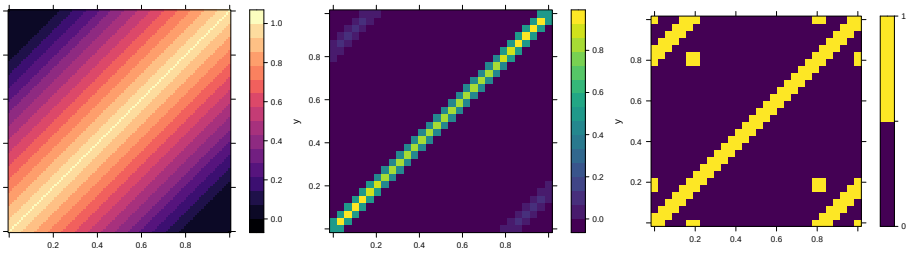
(a) K_1 (Gaussian kernel)



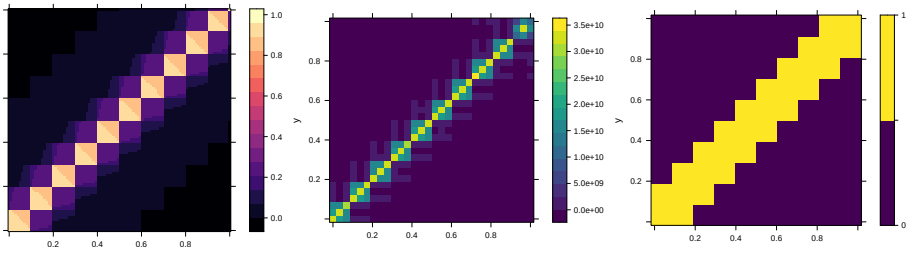
(b) K_2 (Brownian motion)



(c) K_3 (Integrated Brownian motion)



(d) K_4 (Pólya covariance)



(e) K_5 (Linear interpolation of random vector)

Figure 3.4: Plots of the covariance K (left), the matrix of norms $P = [\|\mathbf{P}_{ij}\|]_{i,j=1}^{p}$ (center) and Ω_X^π (right) for $K =$ (a) K_1, (b) K_2, (c) K_3, (d) K_4 and (e) K_5.

that the error in $\|(\operatorname{dg}\hat{\mathbf{K}})(\kappa\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}})^{-1}(\operatorname{dg}\hat{\mathbf{K}}) - \operatorname{dg}\hat{\mathbf{K}}\|$ is around the same as the noise level $\|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|$ of the estimate $\operatorname{dg}\hat{\mathbf{K}}$ as prescribed by the Morozov descrepancy principle (see Kaipio and Somersalo, 2006). Notice furthermore that we have used the same estimator $\operatorname{dg}\hat{\mathbf{K}}$ of $\operatorname{dg}\mathbf{K}$ in all the terms of (3.14) instead of using replicated versions generated using the bootstrap. Although doing things this way would not pose a significant computational burden, it turns out that it also does not significantly improve results. For this reason we have elected to use the simpler method which works and scales well for the purpose of our simulation study.

- **Threshold**. According to our theoretical results, ρ need not decrease with n, but rather any sufficiently small value will suffice. Naturally, as $n \to \infty$, the ij-entries of $\hat{\mathbf{P}}_{ij}$ corresponding to $\mathbf{P}_{ij} = \mathbf{0}$ converge to zero while those for which $\mathbf{P}_{ij} \neq \mathbf{0}$ converge to $\mathbf{P}_{ij}$. In fact, when we plot histograms of the set $\{\log_{10}\|\hat{\mathbf{P}}_{ij}\| : 1 \leq i, j \leq p\}$ for increasing sample sizes, we notice that it tends to separate into roughly two components corresponding to zero and nonzero entries (see Figure 3.5). The separation between the two grows more prominent as the sample size increases and because the scale we have used is logarithmic, the actual difference between the components is that of *an entire order of magnitude*.

  The above observation suggests that the threshold ρ should be chosen so as to divide these two components. Of course, in practice, the two components are rarely as clearly separated as in Figure 3.5 (d) and (e). Realistically, we are more likely to find ourselves in a situation that resembles Figure 3.5 (a). A kernel density estimator can make the components more visible. The local minimums and elbows of the density function can now serve as candidates for the threshold ρ as illustrated in Figure 3.6.

  Intuitively speaking, if ρ is chosen in this manner, then it is ambiguous to which component an entry $\hat{\mathbf{P}}_{ij}$ with $\|\hat{\mathbf{P}}_{ij}\| = \rho$ belongs. In other words, $\|\hat{\mathbf{P}}_{ij}\| = \rho$ represents the decision boundary for the purpose of classifying $\mathbf{P}_{ij}$ into one of the two aforementioned components.

  Alternatively, one can use the stability selection approach of Meinshausen and Bühlmann (2010) which is often used for model selection in LASSO and graphical LASSO. For operator thresholding, the selection probability monotonically decreases with the threshold ρ and we obtain a very simple form for the selection criterion which says that there is an edge between i and j if

$$\frac{1}{N_s}\sum_{k=1}^{N_s} \mathbf{1}_{\{\|\hat{\mathbf{P}}_{ij}^k\| \geq \rho\}} > \pi_{\text{threshold}}$$

  where $\hat{\mathbf{P}}_{ij}^k$ for $1 \leq k \leq N_s$ are bootstrap estimates of $\mathbf{P}_{ij}$ obtained from random subsamples of size $n/2$ and usually, we choose $N_s = 100$. However, this still leaves us with two tuning parameters: ρ and $\pi_{\text{threshold}} \in (1/2, 1)$ and a significant

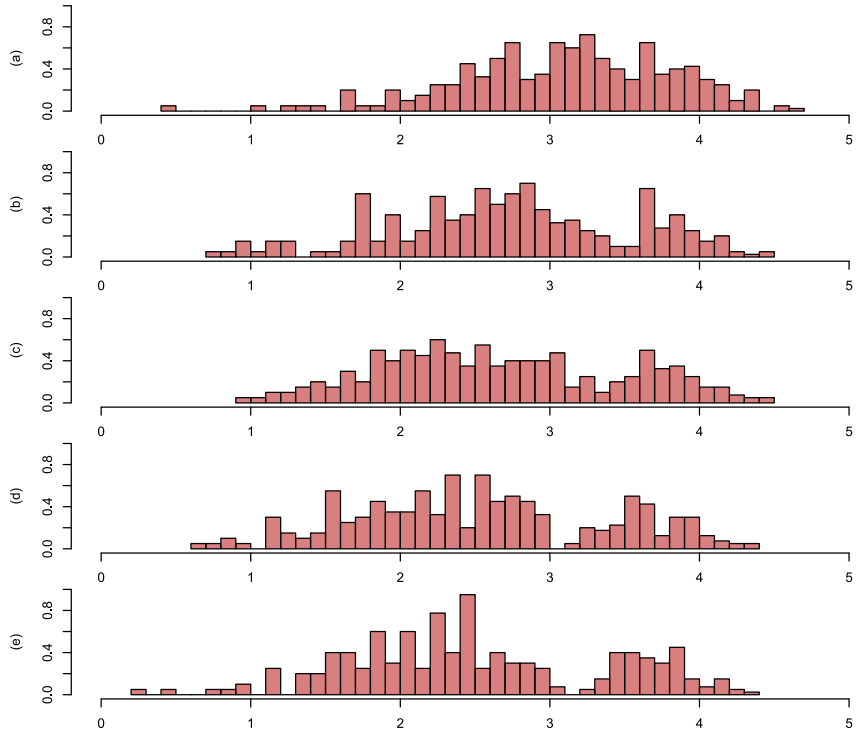
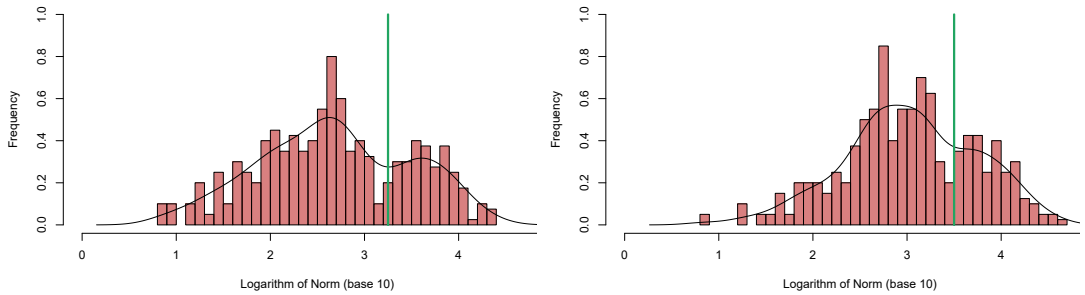


Figure 3.5: Histograms of log-norms $\{\log_{10} \|\hat{\mathbf{P}}_{ij}\| : 1 \leq i, j \leq p\}$ for the integrated Brownian motion covariance K_3 (described in Section 3.7.1) for the sample sizes (a) 200, (b) 400, (c) 600, (d) 800 and (e) 1000. The grid size was 200. The two components increasingly separate with increasing sample size.



(a) (b)

Figure 3.6: The local minima (a) and elbows (b) of the kernel density estimator of $\{\log_{10} \|\hat{\mathbf{P}}_{ij}\| : 1 \leq i, j \leq p\}$ serve as good candidates for the threshold ρ.

computational burden. It thus appears that stability selection is not particularly well adapted to inverse thresholding.

To probe the finite sample performance of our approach as dependent on sample size and the discussion above, we conduct a simulation study considering a variety of Gaussian processes on the unit interval, $U = [0, 1] \subset \mathbb{R}$ and focussing on regular partitions.

3.7.1 Simulations

In this section we shall study the performance of our method for different covariances (K), resolutions (p) and sample sizes (N). We pick U to be the unit interval $[0, 1] \subset \mathbb{R}$ and consider the partitions π given by the collection of subintervals $U_j = [j/p, (j + 1)/p)$ for $0 \leq j \leq p - 1$ and $U_p = [(p - 1)/p, 1]$. This makes it possible to visualise the graphs involved. Furthermore, we consider three values for the sample size N: 50 (low), 100 (moderate) and 200 (high); and three values for the resolution p: 20, 30 and 40, corresponding to p partitions of U. The covariances we study are described in Subsection 3.7.1 and Figure 3.4 displays the level plots of the covariances along with the level plots of the matrix $P = [\mathbf{P}_{ij}]_{i,j=1}^{p}$ (which contains the norms of the entries of the precision matrix $\mathbf{P}$) and the graphs $\tilde{\Omega}_X^\pi$. For some of these covariances, $\tilde{\Omega}_X^\pi$ could not be ascertained from theory and was evaluated numerically instead.

For every covariance K and resolution p, we generate N samples from the Gaussian distribution corresponding to K with mean zero on a regular grid on U of length 600 and calculate $\hat{\Omega}_X^\pi(\rho)$ for various values of ρ using the method described in Section 3.5. We compare $\hat{\Omega}_X^\pi(\rho)$ with the true $\tilde{\Omega}_X^\pi$ and calculate the True Positive Rate (TPR) and the False Positive Rate (FPR) of classifying the pixels $U_i \times U_j$ for every ρ and plot a Receiver Operating Characteristic (ROC) curve as in Figure 3.8. We calculate the Area Under the Curve (AUC) of the ROC curve. We do this 100 times for every combination of K, p and N, and report the median and mean absolute deviation of the AUC rounded to two decimal places. The results are displayed in Table 3.1.

The median AUC naturally increases with the sample size accross covariances and resolutions. Almost perfect results for the covariances $K = K_1$, K_2 and K_3 are most probably due to the structure of their graphs which is, in some sense, simple. For a fixed covariance K and sample size N, the results seem to worsen for the covariances $K = K_2, K_3$ but improve for $K = K_4$ with increasing resolution p. For $K = K_5$, there is an intriguing anomaly for $K = K_5$ and $p = 30$ where the results are noticeably worse than those for the resolutions $p = 20, 40$ which are almost perfect.

We expect that, generally speaking, the results should worsen with increasing resolution eventually. The reason this does not seem to be the case for $K = K_4$ is probably the relatively complicated nature of its graph. Increasing the resolution allows for estimating

the graph at a finer scale, at least for when considering a range of low resolutions. If this is indeed the case then the increase in performance should decrease with the increase in resolution beyond a certain range, which is indeed the case as shown in Table 3.1. The anomaly for $K = K_5$ is probably a result of the corresponding correlation operator $\mathbf{R} = \mathbf{R}_\pi$ being close to noninvertible. Both these observations suggest that a multiresolution approach –one in which one tries to recover the graph of X at several different resolutions so as to detected incidental unfavourable properties of the correlation operator $\mathbf{R}_\pi$ – can be beneficial.

Table 3.1: Medians ± mean absolute deviations (MAD) of Area under the curve (AUC)

Parameters		N		
K	p	50	100	200
K_1	20	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
	30	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
	40	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
K_2	20	0.95 ±0.01	0.97 ±0.01	0.98 ±0.01
	30	0.95 ±0.01	0.96 ±0.01	0.97 ±0.00
	40	0.95 ±0.01	0.96 ±0.06	0.97 ±0.00
K_3	20	0.84 ±0.02	0.88 ±0.02	0.89 ±0.02
	30	0.86 ±0.02	0.87 ±0.02	0.88 ±0.01
	40	0.86 ±0.02	0.87 ±0.01	0.88 ±0.01
K_4	20	0.82 ±0.03	0.85 ±0.00	0.85 ±0.03
	30	0.85 ±0.03	0.87 ±0.02	0.88 ±0.02
	40	0.86 ±0.02	0.89 ±0.02	0.90 ±0.01
K_5	20	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
	30	0.93 ±0.01	0.93 ±0.01	0.94 ±0.01
	40	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00

Construction of Covariances

The five covariances on $U = [0, 1]$ (and corresponding graphs) considered in our simulation study are as follows:

1. **Analytic Covariances.** As we have mentioned before, all analytic covariances have the degenerate graph given by the diagonal $\Omega_X = \{(u, v) : u = v\}$. From this category we shall choose the familiar Gaussian kernel $K_1(u, v) = e^{-(u-v)^2}$.

2. **Covariances of Gaussian Markov Processes.** The most familiar Markov Gaussian process is Brownian motion $X_t = W_t$ which has the graph $\Omega_X = \{(u, v) :$

$u = v\}$ and the covariance $K_2(u, v) = u \wedge v$.

3. **Integrated Brownian Motion.** To see how the effect of applying a linear filter on a process we consider $X_t = \int_{0\vee(t-0.5)}^{t\wedge 1} W_s \, ds$ with the covariance, say K_3. Although we are unable to establish this theoretically, robust numerical evidence suggests that the graph is (approximately) given by $\Omega_X \approx \{(u, v) : |u - v| = 0 \text{ or } 0.5\}$.

4. **Pólya Covariances.** Consider the positive-definite function of the Pólya type Δ_w given by $\Delta_w(t) = (1 - |t/w|)\mathbf{1}_{\{1-|t/w|\geq 0\}}$. We consider $K_4(u, v) = 0.8\Delta_{0.7}(u - v) + 0.2\Delta_{0.8}(u - v)$. This leads to an interesting graph Ω_X approximately given by

$$\Omega_X \approx \{(u, v) : |u - v| = 0 \text{ or } 0.8\} \cup \{0, 0.2, 0.8, 1\}^2.$$

   Once again, this is an approximate result supported by robust numerical evidence and not an exact one justified by theory.

5. **Linear Interpolation of a Random Vector.** To verify that our method for graph recovery in continuous time conforms to our intuition for graph recovery in finite dimensions, we construct a process X_t by linearly interpolating a Gaussian random vector $\mathbf{X} = (X_1, \ldots, X_{q+1}) \in \mathbb{R}^{q+1}$ with mean zero and the covariance given by the Kac-Murdock-Szegö matrix $\mathbf{C} = [\alpha^{|i-j|}]_{i,j=1}^{q+1}$ with the parameters $\alpha = 0.3$ and $q = 10$. Thus $X_t = (1 - t')X_i + t'X_{i+1}$ where $i = 1 + \lfloor tq \rfloor$ and $t' = t - i/q$. Moreover, the covariance is given by

$$K_5(u, v) = (1 - u')(1 - v')\alpha^{|i-j|} + (1 - u')v'\alpha^{|i-j-1|} + u'(1 - v')\alpha^{|i+1-j|} + u'v'\alpha^{|i-j|}$$

   for $i = 1 + \lfloor uq \rfloor$, $j = 1 + \lfloor vq \rfloor$, $u' = u - i/q$ and $v' = v - j/q$. It can be shown that the graph of $\mathbf{X}$ is given by the adjacency matrix $[\mathbf{1}_{|i-j|\leq 1}]_{i,j=1}^{q+1}$ and that $\Omega_X = \{(u, v) : |\lfloor qu \rfloor - \lfloor qv \rfloor| \leq 1\}$.

By numerical evidence above, we mean that this is the structure suggested from computing $\hat{\Omega}_X^\pi$ for the exact covariances K_3 and K_4 with $p = 50$ on a grid size of 1200 for the values of the truncation parameter corresponding to the longest region of stability as explained in Subsection 3.7.1 and illustrated in Figure 3.7.

To understand how well this approach might work, we compute the estimator $\hat{\Omega}_X^\pi(\rho)$ for the covariances K_j for $1 \leq j \leq 5$ discretized on a regular grid of length 600. We then plot histograms of $\log_{10} \|\mathbf{P}_{ij}\|$ and identify regions of stability. We ignore the small number of $\|\mathbf{P}_{ij}\|$ which are computationally zero, so the logarithm does not pose a problem. The results are documented in Figure 3.7.

Notice that the *signal* which constitutes significant entries $\mathbf{P}_{ij}$ of the precision matrix is often comfortably separated from the *noise* which is composed of those entries which are supposed to be zero and often by many orders of magnitude. Although, this is not

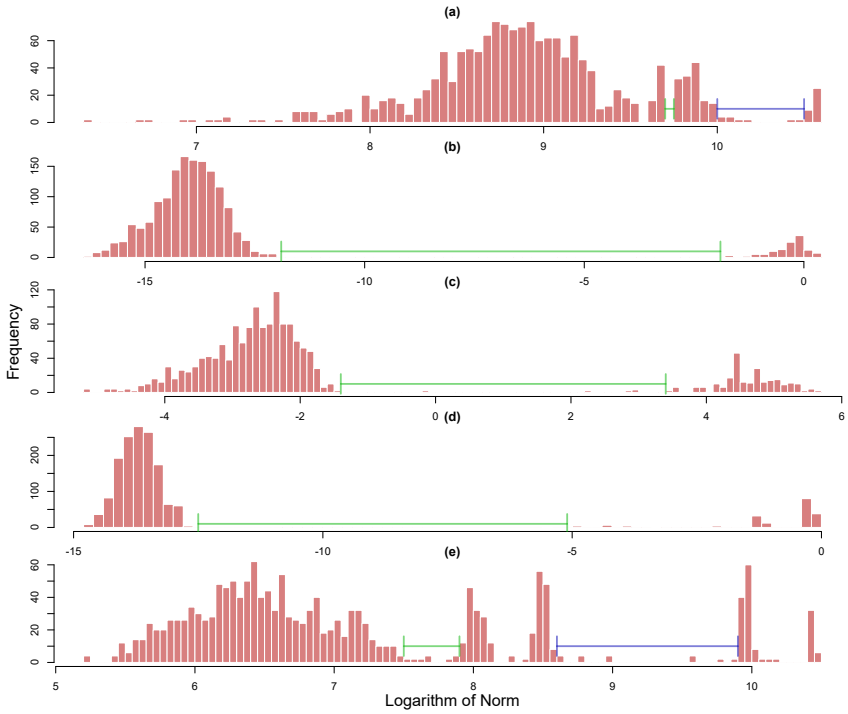


Figure 3.7: Histograms of log-norms $\{\log_{10} \|\mathbf{P}_{ij}\| : 1 \leq i, j \leq p\}$ for the covariances (a) K_1, (b) K_2, (c) K_3, (d) K_4 and (e) K_5. The green bars indicate regions of stability i.e. the values of the thresholding parameter ρ for which $\hat{\Omega}_X^{\pi}(\rho) = \tilde{\Omega}_X^{\pi}$. Additionally, the blue bars in (a) and (e) represent the values of ρ for which $\Omega_X^{\pi}(\rho) = \Omega_X^{\pi}$.

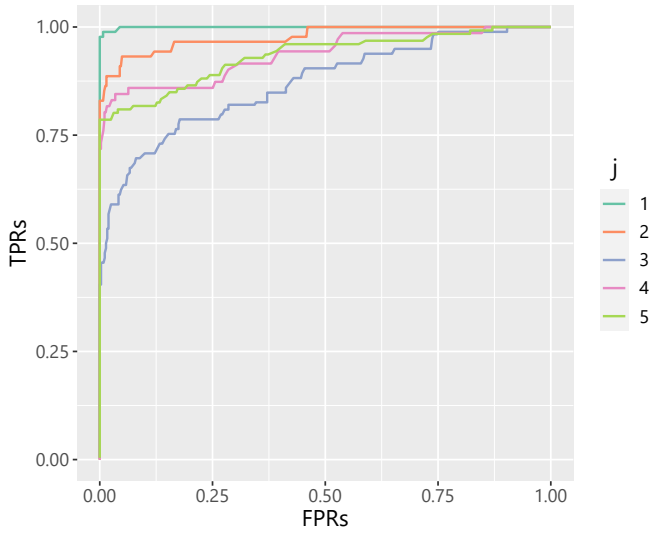


Figure 3.8: Sample ROC curves for simulated instances of the covariances K_j for $1 \leq j \leq 5$.

exactly the case for K_1 (a) (probably due to its smoothness) and K_5 (e), we are still able to identify relatively long regions (blue) over which $\hat{\Omega}_X^{\pi}(\rho) = \Omega_X^{\pi}$ for (a) and (e).

3.8 Illustrative Data Analysis

In this section, we illustrate our method by analysing two data sets. The first concerns infrared absorption spectra obtained from fruit purees where we expect the graph to have significant associations between distant locations. The second involves the intraday price of a certain stock where we expect the graph to resemble that of a Markov process as in Figure 3.4 (a) or (b).

3.8.1 Infrared Absorption Spectroscopy

A very interesting application of graphical modelling to absorption spectrometry appears in Codazzi et al. (2022), in which the absorption spectra obtained from a sample of strawberry purees are modelled as continuous functions and an attempt is made at estimating their conditional dependence structure via a Bayesian inference procedure. The method involves B-spline smoothing of the spectra and uses the conditional dependence between the smoothing coefficients as a substitute for the conditional dependence structure of the spectra.

This structure is of interest to determining the chemical composition of the puree samples. In particular, if different regions of the spectrum are related, then they

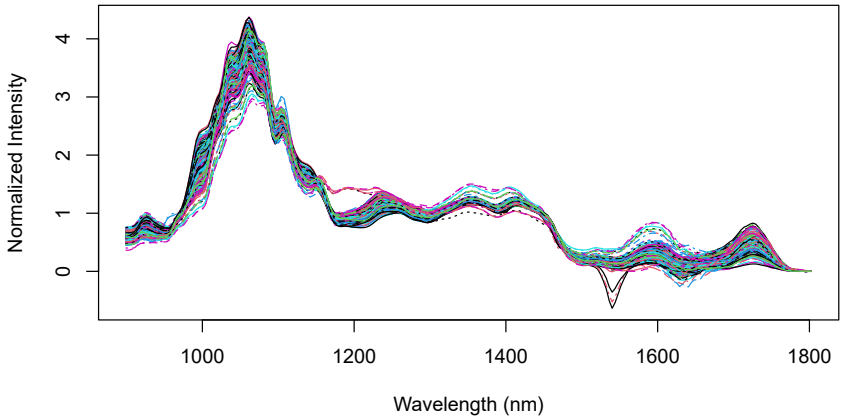
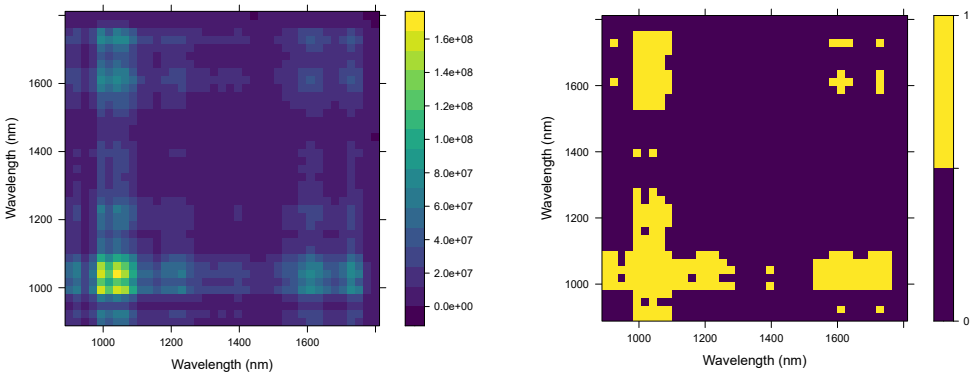


Figure 3.9: Absorption spectra of strawberry purees.



(a) (b)

Figure 3.10: (a) The matrix of operator norms $[\|\mathbf{P}_{ij}\|]_{i,j=1}^{p}$ and (b) the graph $\tilde{\Omega}_X^{\pi}$ obtained for the threshold $\rho = 10^{7.5}$ for the absorption spectra of strawberry purees.

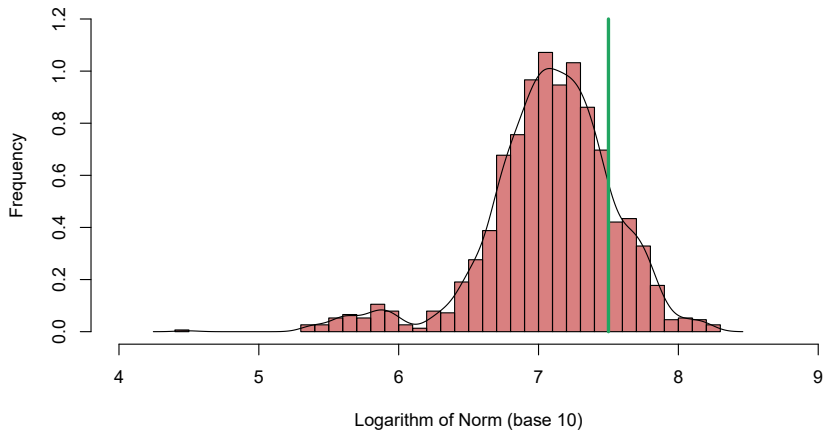


Figure 3.11: Histogram and density of the log-norms $\{\log_{10} \|\hat{\mathbf{P}}_{ij}\| : 1 \leq i, j \leq p\}$ for the strawberry puree data. The green line indicates the threshold ρ chosen for the graph in Figure 3.10 (b). It has been manually chosen to be slightly less than the value corresponding to the elbow of the density curve which corresponds to $\rho = 10^{7.6}$.

probably correspond to the same chemical component. This could be useful for detecting adulteration. Using our method, we approach the problem directly. We calculate the covariance of L^1-normalized absorption spectra readings from the dataset Shu et al. (2019), obtained from $n = 351$ samples of freshly prepared strawberry purees on a uniform grid of 235 wavelengths in the interval $I = [899.327 \text{ nm}, 1802.564 \text{ nm}]$ (see Figure 3.9). We discard the last wavelength so as to make it easier to divide the domain into $p = 39$ partitions and calculate the corresponding precision matrix, which is thresholded at a manually chosen level of $\rho = 10^{7.5}$ using the method described in Section 3.7 (see Figure 3.11). The kernel density estimate was automatically calculated using the density function in the R Base package (R Core Team, 2021) with default parameters. The results are summarized in Figure 3.10. The graph thus obtained is very similar to the one obtained in Codazzi et al. (2022).

3.8.2 Stock Price for Pfizer Limited

We consider the intraday price of Pfizer Limited (NSE: PFIZER) listed on India's National Stock Exchange (NSE) at 1 minute intervals during 988 regular trading sessions (09:15 AM - 15:30 PM IST) from 2nd January 2017 to 1st January 2021 (see Figure 3.12). The prices are considered relative to the opening price of the day. The data has been made freely available on Kaggle by Kumar (2022).

On many days, trading was halted during the session, which lead to missing data. To circumvent this problem, we estimate the covariance in a pairwise manner. The resulting

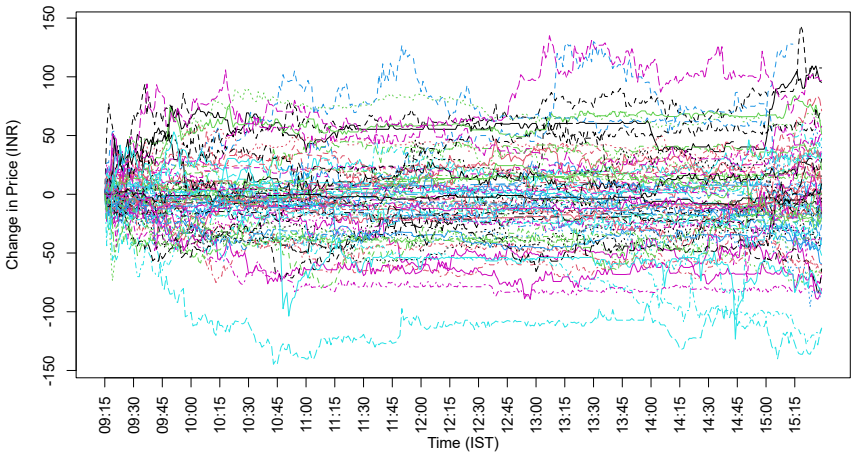


Figure 3.12: Relative price of Pfizer Limited during regular trading sessions from 2nd January 2017 to 1st January 2021.

estimate is almost but not exactly positive semidefinite, so we project it to the cone of positive semidefinite matrices by retaining only the positive part of its eigendecomposition. The resolution of the grid is 375 and we choose $p = 25$. The results are summarized in Figure 3.14. The choice of the threshold using the method described in Section 3.7 is summarized in Figure 3.13 and the kernel density estimate was automatically calculated using the density function in the R Base package R Core Team (2021) with default parameters as before.

The graph almost exactly resembles what one would expect for a Markov process, except for a noticeable clique for times between 12:15 and 13:45. The almost Markov nature of the graph is to be expected since it is widely believed in the academic literature in finance that stocks are mostly efficiently priced. The apparent existence of a clique may or may not be an interesting feature open to financial interpretation.

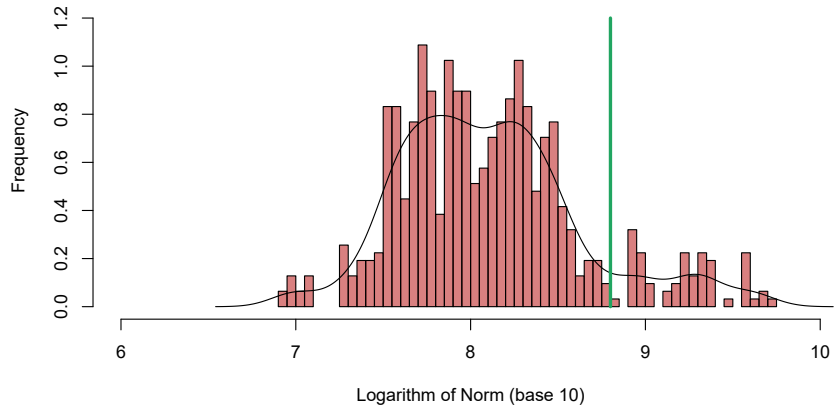
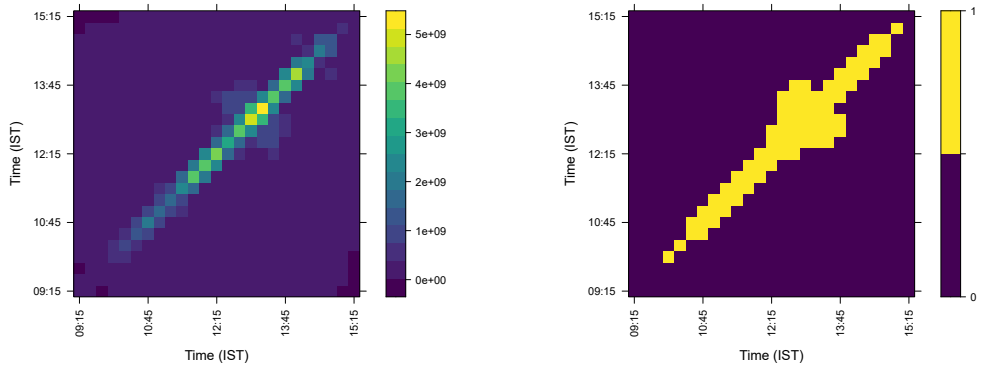


Figure 3.13: Histogram and density of the log-norms $\{\log_{10} \|\hat{\mathbf{P}}_{ij}\| : 1 \leq i, j \leq p\}$ for stock price data. The green line indicates the threshold ρ chosen for the graph in Figure 3.14 (b). It has been chosen to be an elbow of the density curve which corresponds to $\rho = 10^{8.8}$



(a) (b)

Figure 3.14: (a) The matrix of operator norms $[\|\mathbf{P}_{ij}\|]_{i,j=1}^{p}$ and (b) the graph $\tilde{\Omega}_X^{\pi}$ obtained for the threshold $\rho = 10^{8.8}$ for the stock price of Pfizer Limited.

4 Functional Graphical Lasso

Abstract

...In 1961 one of us (Tribus) asked Shannon what he had thought about when he had finally confirmed his famous measure. Shannon replied: "My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name. In the second place, and more importantly, no one knows what entropy really is, so in a debate you will always have the advantage.'"

Edward C. McIrvine and Myron T. Tribus, *Energy and Information* (1971)

Abstract

We consider the problem of recovering conditional independence relationships between a finite number of jointly distributed second-order Hilbertian random elements given multiple realizations thereof. We operate in the sparse high-dimensional regime and propose an infinite-dimensional generalization of the multivariate graphical lasso. We prove model selection consistency under natural assumptions and extend many classic results to infinite dimensions. Most importantly, our method can be understood as arising from a coherent maximum likelihood philosophy.

4.1 Introduction

We consider the problem of recovering conditional independence relationships between a finite number of jointly distributed second-order Hilbertian random elements given multiple realizations thereof. We operate in the sparse high-dimensional regime, where every random element is conditionally dependent on only a small number of other random elements, and the number of joint observations of the random elements is small compared to the number of random elements. We propose a plug-in procedure which recovers

the conditional independence graph from an estimate of the joint covariance of the random elements and can be thought of as an infinite-dimensional generalization of the multivariate graphical lasso.

Specifically, consider $X = (X_1, \ldots, X_p)^\top$ where $\{X_j\}_{j=1}^p$ are jointly distributed second-order random elements in the Hilbert spaces $\{\mathcal{H}_j\}_{j=1}^p$, respectively. The conditional independence structure of X can be thought of as an undirected graph G with the vertices $\{X_j\}_{j=1}^p$, where for $i \neq j$, X_i and X_j are adjacent unless they are conditionally independent given the rest of the vertices $\{X_k\}_{k \neq i,j}$, that is

$$X_i \perp\!\!\!\perp X_j \mid \{X_k\}_{k \neq i,j}.$$

The maximum degree d of G is defined as the maximum number of neighbours (adjacent vertices) of a vertex of G. We are interested in determining the edges of the graph G from n independent realizations $\{X^k\}_{k=1}^n$ of X in the sparse high-dimensional regime, where $n \ll p$ and $d \ll p$. The standard multivariate version of the problem can be seen as a special case where $\mathcal{H}_j = \mathbb{R}$ for every j and consequently $\{X_j\}_{j=1}^p$ are real-valued random variables.

In the multivariate setting, the problem has been studied comprehensively and many methods have been devised. Of these, *precision thresholding* is the simplest as it merely requires thresholding the entries of the inverse of the empirical covariance matrix $\hat{\mathbf{C}}$. If the absolute value of the (i,j)th entry of $\hat{\mathbf{C}}^{-1}$ is below the threshold, then the corresponding edge is understood as being absent in the graph. The motivation for this comes directly from a classic result in the theory of Gaussian graphical models that we shall call the *inverse zero characterization*, which states that if X is Gaussian with an invertible covariance $\mathbf{C}$, the (i,j)-th entry of $\mathbf{C}^{-1}$ is non-zero if and only if X_i and X_j are adjacent (Lauritzen, 1996; Meinshausen and Bühlmann, 2006; Drton and Maathuis, 2017). The method does not perform well in the sparse high-dimensional regime ($n \ll p$) because it cannot make use of the sparsity in the graph structure.

Fortunately, there are methods which are consistent in high-dimensions. One such method, known as *neighbourhood selection* (Meinshausen and Bühlmann, 2006), involves performing ℓ_1-penalized linear regression on each of the random variables against the rest with the non-zero coefficients in the regression corresponding to the neighbours of the random variable. A second such method, called the *graphical lasso* (Yuan and Lin, 2007; Friedman et al., 2008) combines the sparsity-exploiting properties of the ℓ_1 penalty along with the inverse zero characterization and is known to be consistent in high-dimensional settings (Rothman et al., 2008; Ravikumar et al., 2011). In practice, the graphical lasso is often the method of choice for Gaussian graphical models, likely due to its conceptual simplicity and ability to perform estimation and model selection (i.e., support estimation) in a single step (Yuan and Lin, 2007). The method involves estimating the precision matrix by maximizing the appropriately penalized Gaussian

log-likelihood:

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \operatorname{tr}\left(\hat{\mathbf{C}}\mathbf{Q}\right) - \log\det(\mathbf{Q}) + \lambda\|\mathbf{Q}\|_{1-}, \tag{4.1}$$

where $\mathbf{Q} = [q_{ij}]_{i,j=1}^{p}$ is positive-definite, $\lambda > 0$ is a tuning parameter and $\|\mathbf{Q}\|_{1-} = \sum_{i\neq j}|q_{ij}|$ is the penalty term which promotes sparsity in $\mathbf{Q}$ by driving the the less significant of its off-diagonal entries to zero.

We shall extend the graphical lasso to the general Hilbertian setting by reformulating the optimization problem (4.1) in infinite-dimensional terms. Our primary concern is the *multivariate functional data* setting in which X_j are real-valued random functions on compact intervals of the real line. The distinctive feature of functional data (Ramsay and Silverman, 2005; Hsing and Eubank, 2015), as opposed to multivariate data, is the fact that the covariance operator is trace-class and thus not boundedly invertible, obscuring the relationship between the graphical model and the support of the inverse covariance. While we focus on multivariate functional data, our approach can in principle be used to recover relationships between diverse types of random objects, be it variables, vectors, functions, surfaces, so long as they can be represented as second-order random elements in a Hilbert space. We shall mostly restrict ourselves to the classical setting where X_j are jointly Gaussian. For non-Gaussian X_j, our method recovers relationships based on an alternative notion of irrelevance, which is called conditional uncorrelation and is based purely on linear relationships between the random elements.

Recovering conditional independence graphs of multivariate functional data by means of extending the graphical lasso has been attempted before in the literature. Qiao et al. (2019) proposed an intuitive approach that proceeds by representing every random function X_j as a random vector of a chosen number of its principal component scores. Then the conditional independence graph of the resulting representations is recovered using the joint graphical lasso (Danaher et al., 2014), which ensures that the procedure recovers relationships between the different random functions while ignoring those between principal scores corresponding to the same random function. While the method is elegant in its conception, Zapata et al. (2022) noted that connecting conditional independence relationships between the random function with the zeros of the precision matrix of their principal component representations seems to require that every random function can be represented as a finite linear combination of a fixed number of deterministic functions with random coefficients. In other words, the functional data has to be exactly finite-dimensional. Observing that this assumption is impractical and unrealistic for certain applications, Zapata et al. (2022) advance a novel assumption of their own, called partial separability, under which they link the conditional independence graph of the random functions with the zero entries of a suitably defined precision matrix, while allowing the data to be infinite dimensional. As an interesting generalization of the separability assumption that is popular in multi-way functional data (Aston et al., 2017), partial separability could be of interest even in areas other than functional graphical models. However, it still constitutes a serious structural assumption as it postulates that

the covariance operators $\{\mathbf{C}_{jj}\}_{j=1}^{p}$ of the random functions $\{X_j\}_{j=1}^{p}$ are simultaneously diagonalizable, that is, they have the same eigenfunctions. Moreover, evaluating the plausibility of the assumption for a given data set on an intuitive basis is difficult and a statistical test for it has not yet been developed.

Both of these approaches to functional graphical models are based on functional principal components analysis, treating functional realizations in terms of their truncated principal component representations, and recovering conditional independence relationships between random elements from these representations demands imposing structural assumptions. The necessity of dimensionality reduction can be understood as stemming from the absence of a determinant-like functional on the space of covariance operators. Indeed, every functional defined as the product of eigenvalues must vanish everywhere because the eigenvalues of covariance operators converge to zero by virtue of compactness. In this chapter, we shall present an approach that circumvents this problem by reformulating (4.1) in terms of correlation operators, which are operator analogues of correlation matrices, and a regularized infinite-dimensional generalization of the matrix determinant known as the Carleman-Fredholm determinant (or alternatively the Hilbert-Carleman determinant). Although the use of correlation in place of covariance is not uncommon in functional data (Lee et al., 2023; Li and Solea, 2018b) and is, in fact, standard practice for the multivariate graphical lasso (Kovács et al., 2021), it arises quite naturally in our treatment, which can be understood as emerging from a coherent maximum likelihood philosophy. The key idea is to use the product measure of the "coordinates" $\{X_j\}_{j=1}^{p}$ as a reference measure.

4.1.1 Related Work

Other well-known methods of graph recovery for multivariate data that we mentioned above have also been generalized to the multivariate functional setting. Of these, inverse thresholding (Li and Solea, 2018b; Lee et al., 2023) is naturally the simplest, as it merely requires thresholding the entries of the inverse of a certain correlation operator that can be computed from the data. This requires the operator to be invertible. Like its multivariate counterpart and for the same reasons, this method is not supposed to perform well in the sparse high-dimensional setting. Naturally, all results proving model selection consistency for this approach assume that p is fixed.

Functional generalizations of the neighborhood selection approach (Zhao et al., 2021; Lee et al., 2022) involve performing appropriately penalized functional regression on each of the random elements against every other random element. Unlike inverse thresholding, these methods do work well in the high-dimensional settings and they also possess the computational advantage of being amenable to parallel implementation. But due to their reliance on functional regression, they require the regression operator to be Hilbert–Schmidt. This constitutes a substantial structural assumption since regression

operators are not bounded in general (Kneip and Liebl, 2020), and it does not hold if, for example, one of the random element is a linear combination of some other random elements since the corresponding regression operator will then be proportional to identity, which is not a Hilbert–Schmidt operator.

Many works in the literature, including those discussed here, deal with the more complicated setting where $\{X_j\}_{j=1}^p$ are non-Gaussian or have non-linear relationships, making them very different in flavor from the work presented here. They exhibit a complex variety in the details of the structural assumptions they make. As an interesting development, we mention here Solea and Dette (2022), but a comprehensive review of these different approaches is beyond the scope of this chapter.

4.1.2 Contribution

We extend the graphical lasso to a general infinite-dimensional Hilbertian setting. Under rather minimal and intuitive functional counterparts of the multivariate assumptions, we prove functional analogues of state-of-the-art results in the form of finite-sample guarantees concerning the family-wise error rate of model selection and the rates of convergence for precision estimation known for multivariate graphical lasso (Ravikumar et al., 2011) and, as a result, establish model selection consistency.

Our method can be motivated in a very natural manner from the maximum likelihood principle, which is uncommon, to say the least, for functional data due to the lack of a suitable replacement for the Lebesgue measure in function spaces. In doing so, we demonstrate what might be the right approach to applying the likelihood method to multivariate functional data. This development could be of wider interest.

Furthermore, we extend classic results concerning the equivalence of graphical lasso to penalized log-likelihood maximization, Kullback-Leibler divergence minimization and determinant maximization which were known in the multivariate setting, to the general setting of infinite-dimensional Hilbert spaces. From an analytical perspective, methods in functional data analysis are often infinite-dimensional reformulations of their counterparts in multivariate analysis. While reformulating functions such as the trace and Frobenius norm is almost trivial, we have managed to achieve the same for the nontrivial and rather tricky case of the determinant. This is significant given that the determinant is an important measure of the joint dispersion in multivariate analysis and appears in many other problems. We expect the development in this paper to bear other fruitful generalizations in the future.

Our treatment also clarifies certain elements of the multivariate functional data literature. For example, the hitherto ad-hoc concept of correlation operator arises naturally from the likelihood approach, while the assumption of eigenvalue gap, previously made only in order to ensure that the correlation operator is invertible, now admits a concrete

interpretation in terms of the supports of the measures involved. Moreover, we show that the inverse zero characterization holds in complete generality, without the need for any structural assumptions.

In the abstract tradition, the functional graphical lasso also contributes to our understanding of its multivariate counterpart by identifying the analytical properties of the random objects involved, that make the method work. It also suggests new ways of using the graphical lasso in the multivariate setting. In principle, using tools such as kernel or graph embeddings, the method can be extended to other classes of random objects such as distributions and networks, and to nonlinear relationships.

For all of its attractive theoretical properties, the approach is not without practical merits. It greatly eases the burden of parameter tuning, requiring only the lasso-type penalty parameter to be chosen, which is well-understood and can be easily interpreted using the method's divergence minimization characterization. The use of truncated representations is also completely optional in the moderately high-dimensional setting, where p is not so large as to compel some kind of dimensionality reduction due to computational constraints. Our simulations reveal that, in spite of its computational benefits, dimensionality reduction can be counterproductive if the underlying functions do not admit efficient representations, as is the case when the sample paths of the random functions $\{X_j\}_{j=1}^p$ are rough. Furthermore, the coordinate-free operator formulation of the method allows the user to choose whichever discretization scheme they deem fit, be it basis representation, point evaluation or cell averaging for reasons of representation accuracy or efficiency. For the same reason, working with heterogeneous data, where $\{X_j\}_{j=1}^p$ comprises of different kinds of random objects such as variables, vectors, curves or surfaces, is as simple as dealing with homogeneous data. Finally, the plug-in nature of the method permits the user to choose the covariance estimation procedure which is appropriate given the nature of the available observations, thus making it applicable to functional time series and sparsely observed functional data as well.

4.1.3 Structure of the Chapter

We begin by describing important concepts and introducing our notation in Section 4.2. This is followed by the problem formulation and a discussion of the assumptions of our method in Sections 4.3 and 4.4, respectively. In Section 4.5, we describe our methodology, its motivations and interpretations. Our main results, including finite sample results concerning the estimation of the precision operator and model selection consistency, are stated in Section 4.6. The proof are deferred to the supplementary material. Section 4.7 contains the details of how the method is implemented, and Section 4.8 presents simulation studies of our method's performance.

4.2 Background and Notation

The symbols $\mathbf{I}$ and $\mathbf{0}$ shall denote the identity and zero elements of the spaces they inhabit, which will be clear from the context, according to which they can be elements, operators or operator matrices. The symbols $x \vee y$ shall be used as a short-hand for maximum of $x, y \in \mathbb{R}$.

For $1 \leq j \leq p$, let $\mathcal{H}_j$ be separable Hilbert spaces equipped with the inner products $\langle \cdot, \cdot \rangle_j$. The subscript j shall always be clear from the context and we shall avoid writing it explicitly, preferring $\langle f, g \rangle$ instead, for $f, g \in \mathcal{H}_j$. We shall denote by $\mathcal{H}$, the *product Hilbert space* denoted by $\mathcal{H}_1 \times \cdots \times \mathcal{H}_p$ or $\times_{j=1}^{p} \mathcal{H}_j$ equipped with the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{j=1}^{p} \langle f_j, g_j \rangle$$

for $\mathbf{f}, \mathbf{g} \in \mathcal{H}$, where $\mathbf{f} = (f_1, \ldots, f_p)$ and $\mathbf{g} = (g_1, \ldots, g_p)$.

4.2.1 Operators and Operator Matrices on Hilbert Spaces

Operators between Hilbert spaces shall be denoted using boldface, as in $\mathbf{A}$ with the corresponding operator norm and adjoint being written as $\|\mathbf{A}\|$ and $\mathbf{A}^*$ as usual. We define the *spectrum* $\sigma(\mathbf{A})$ of $\mathbf{A}$ as the set of $\lambda \in \mathbb{R}$ for which the operator $\mathbf{A} - \lambda \mathbf{I}$ does not admit a bounded inverse. The notation $\mathbf{A}^{-1}$ shall denote the inverse of the operator $\mathbf{A}$ or its pseudoinverse, in case it is not invertible.

We shall mostly work with spaces of Hilbert–Schmidt operators. The space of Hilbert–Schmidt operators on a Hilbert space $\mathcal{H}$ shall be denoted as $\mathcal{L}_2(\mathcal{H})$. The space $\mathcal{L}_2(\mathcal{H})$ forms a Hilbert space under the inner product $\langle \cdot, \cdot \rangle_2$ induced by the Hilbert–Schmidt norm $\| \cdot \|_2$ given by

$$\|\mathbf{H}\|_2^2 = \sum_{j=1}^{\infty} \sigma_j^2(\mathbf{H})$$

where $\{\sigma_j\}_{j=1}^{\infty}$ are the singular values of $\mathbf{H}$, or equivalently, the eigenvalues of $|\mathbf{H}| = \sqrt{\mathbf{H}^*\mathbf{H}}$. It is a well-known fact that if $\mathcal{H}$ is the space of square-integrable functions, Hilbert–Schmidt operators can be elegantly represented as an integral operators corresponding to square-integrable kernels.

An *operator matrix* is a matrix of the form $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1}^{p}$ where the ijth entries are operators $\mathbf{A}_{ij} : \mathcal{H}_j \to \mathcal{H}_i$. For an operator matrix $\mathbf{A}$, we define the diagonal part $\operatorname{dg} \mathbf{A}$ of $\mathbf{A}$ as the diagonal matrix $\mathbf{D} = [\mathbf{D}_{ij}]_{i,j=1}^{p}$ given by $\mathbf{D}_{ij} = \mathbf{A}_{ij}$ for $i = j$ and $\mathbf{0}$ otherwise. The off-diagonal part $\mathbf{A} - \operatorname{dg} \mathbf{A}$ shall be denoted as $\mathbf{A}_0$. Operator matrices can be thought

of as operators on the product Hilbert space $\mathcal{H}$, as given by

$$\mathbf{A}\mathbf{f} = \left[\sum_{j=1}^{p} \mathbf{A}_{ij} f_j\right]_{i=1}^{p}$$

for $\mathbf{f} = (f_1, \ldots, f_p) \in \mathcal{H}$. The adjoint of an operator matrix $\mathbf{A}$ shall be denoted as $\mathbf{A}^\top$.

Note that the trace $\operatorname{tr}\mathbf{A}$ and Hilbert–Schmidt norm $\|\mathbf{A}\|_2$ of an operator matrix $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1}^{p}$ can be written in terms of the traces and Hilbert–Schmidt norms of the entries as

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{p} \operatorname{tr}(\mathbf{A}_{ii}) \quad \text{and} \quad \|\mathbf{A}\|_2^2 = \sum_{i,j=1}^{p} \|\mathbf{A}_{ij}\|_2^2.$$

To mirror the behaviour of Euclidean spaces in the product Hilbert space, we devise some additional norms. The operator counterparts $\|\cdot\|_{2,1}$ and $\|\cdot\|_{2,\infty}$ of the ℓ_1 and ℓ_∞ norms are given by $\|\mathbf{A}\|_{2,1} = \sum_{i,j=1}^{p} \|\mathbf{A}_{ij}\|_2$ and $\|\mathbf{A}\|_{2,\infty} = \max_{i,j} \|\mathbf{A}_{ij}\|_2$. In the same way, we define the operator analogues of matrix norms: $|||\mathbf{A}|||_{2,\infty} = \max_i \sum_j \|\mathbf{A}_{ij}\|_2$ (maximum column sum) and $|||\mathbf{A}|||_{2,1} = \max_j \sum_i \|\mathbf{A}_{ij}\|_2$ (maximum row sum). Note that $|||\mathbf{A}|||_{2,1} = \left|\left|\left|\mathbf{A}^\top\right|\right|\right|_{2,\infty}$ and that these norms are sub-multiplicative (see Appendix).

The tensor product $\mathbf{A} \otimes \mathbf{B}$ of operator matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as the linear map $\mathbf{D} \mapsto \mathbf{BDA}$ and can also be expressed as an array $[\mathbf{A}_{ij} \otimes \mathbf{B}_{kl}]_{i,j,k,l=1}^{p}$ of the tensor products of their entries. The action $\mathbf{D} \mapsto \mathbf{BDA}$ can be imitated by a matrix $[\mathbf{A}_{ij} \otimes \mathbf{B}_{kl}]_{(i,j),(k,l)}$ (indexed by the pairs (i,j) and (k,l)) acting on vectorized version of $\mathbf{D} = [\mathbf{D}_{ij}]_{(i,j)}$ (indexed by (i,j)). As a result, we can simultaneously think of the tensor product $\mathbf{A} \otimes \mathbf{B}$ as a linear map and as a matrix with tensor product entries.

4.2.2 Second-Order Random Elements in Hilbert Space

A random element X is said to be second-order if $\mathbb{E}[\|X\|^2] < \infty$. For such random elements, we can define the mean and the covariance operator

$$\mathbf{m} = \mathbb{E}[X] \quad \text{and} \quad \mathbf{C} = \mathbb{E}[(X - \mathbb{E}[X]) \otimes (X - \mathbb{E}[X])],$$

respectively. If $\mathcal{H}$ is the product Hilbert space of certain Hilbert spaces $\mathcal{H}_j$ for $1 \le j \le p$, then we can write X as a random tuple, as in $X = (X_1, \ldots, X_p)$ where $\{X_j\}_{j=1}^{p}$ are jointly distributed random elements on their respective Hilbert spaces. The covariance operator $\mathbf{C}$ can then be thought of as an operator matrix, as in $\mathbf{C} = [\mathbf{C}_{ij}]_{i,j=1}^{p}$ where the (i,j)-th entry is given by the operator $\mathbf{C}_{ij} = \mathbb{E}[(X_i - \mathbf{m}_i) \otimes (X_j - \mathbf{m}_j)]$.

By a well-known result of Baker (1973), for every $1 \le i, j \le p$ for $i \ne j$ there exists a unique bounded linear operator $\mathbf{R}_{ij} : \mathcal{H}_j \to \mathcal{H}_i$ with $\|\mathbf{R}_{ij}\| \le 1$ such that $\mathbf{C}_{ij} = \mathbf{C}_{ii}^{1/2} \mathbf{R}_{ij} \mathbf{C}_{jj}^{1/2}$

and $\mathbf{R}_{ij} = \Pi_i \mathbf{R}_{ij} \Pi_j$ where Π_i, Π_j are projections to the closures of the images of $\mathbf{C}_{ii}$ and $\mathbf{C}_{jj}$ in their respective co-domains, which is to say that $\mathbf{R}_{ij}$ maps the closure of the range of $\mathbf{C}_{jj}$ to that of $\mathbf{C}_{ii}$. Accordingly, we define the correlation operator matrix as $\mathbf{R} = [\mathbf{R}_{ij}]_{i,j=1}^{n}$ where $\mathbf{R}_{ii} = \mathbf{I}$ and $\mathbf{R}_{ij} = \mathbf{C}_{ii}^{-1/2} \mathbf{C}_{ij} \mathbf{C}_{jj}^{-1/2}$ for $i \neq j$, where $\mathbf{C}_{ii}^{-1/2}$ and $\mathbf{C}_{jj}^{-1/2}$ are understood to be the operator pseudoinverses of $\mathbf{C}_{ii}^{1/2}$ and $\mathbf{C}_{jj}^{1/2}$, respectively. It can be shown that $\mathbf{R}$ is always a positive semi-definite operator. Theorem 4.2 specifies a sufficient condition on the random element X, under which $\mathbf{R}$ is strictly positive-definite and hence invertible.

If $\mathbf{R}$ is invertible, we can write its inverse as $\mathbf{R}^{-1} = \mathbf{I} + \mathbf{H}$ where $\mathbf{I}$ is understood as the identity operator matrix and $\mathbf{H} = [\mathbf{H}_{ij}]_{i,j=1}^{p}$ is a bounded operator matrix. We shall refer to $\mathbf{H}$ as the *precision operator matrix* or simply, the *precision operator* of X.

We shall describe the diffusedness of the distributions of our random elements using the notion of sub-Gaussian and sub-exponential norms of random variables. The sub-Gaussian and sub-exponential norms of a random variable Z are given by

$$\|Z\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\Big[\exp(Z^2/t^2)\Big] \leq 2\} \text{ and } \|Z\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}\Big[\exp(|Z|/t)\Big] \leq 2\},$$

respectively.

4.2.3 Conditional Independence Graphs of Random Elements

Let $X = (X_1, X_2, X_3)$ be a random element in the product Hilbert space $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$. We say that X_1 and X_2 are conditionally independent given X_3, or alternatively, $X_1 \perp\!\!\!\perp X_2 \mid X_3$ if the conditional measures $\mathbb{P}_{X_1|X_3}$, $\mathbb{P}_{X_2|X_3}$ and $\mathbb{P}_{X_1,X_2|X_3}$ satisfy

$$\mathbb{P}_{X_1,X_2|X_3} = \mathbb{P}_{X_1|X_3} \otimes \mathbb{P}_{X_2|X_3}.$$

The σ-algebra generated by the random variables $\{\langle h, X_3 \rangle : h \in \mathcal{H}_3\}$ is same as the Borel σ-algebra associated with $\mathcal{H}_3$ (Hsing and Eubank, 2015, Theorem 7.1.1). Therefore, the above statement can be interpreted in terms of the familiar notion of conditional independence for real-valued random variables as follows: for every $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$, we have that

$$\langle f, X_1 \rangle \perp\!\!\!\perp \langle g, X_2 \rangle \mid \{\langle h, X_3 \rangle : h \in \mathcal{H}_3\}. \tag{4.2}$$

Thus X_1 and X_2 are conditionally independent given X_3 if and only if any two linear functionals of X_1 and X_2 are conditionally independent given every linear functional of X_3. If X is second-order, we can define a purely second-order counterpart of the notion of conditional independence called *conditional uncorrelatedness* which we denote as $X_1 \perp\!\!\!\perp_2 X_2 \mid X_3$ and define as

$$\mathbb{E}\Big[\Big(\langle f, X_1 \rangle - \mathbb{E}_2\Big[\langle f, X_1 \rangle \Big| L(X_3)\Big]\Big)\Big(\langle g, X_2 \rangle - \mathbb{E}_2\Big[\langle g, X_2 \rangle \Big| L(X_3)\Big]\Big)\Big] = 0 \tag{4.3}$$

or equivalently,

$$\mathbb{E}\Big[\langle f, X_1\rangle\langle g, X_2\rangle\Big] = \mathbb{E}\Big[\mathbb{E}_2\Big[\langle f, X_1\rangle \mid L(X_3)\Big] \cdot \mathbb{E}_2\Big[\langle g, X_2\rangle \mid L(X_3)\Big]\Big] \tag{4.4}$$

where $\mathbb{E}_2[Z \mid L(X_3)]$ denotes the best linear unbiased predictor of the random variable Z from the closed linear span $L(X_3)$ of the random variables $\{\langle h, X_3\rangle : h \in \mathcal{H}_3\}$. For zero-mean Gaussian random elements, the two notions of independence and uncorrelatedness coincide (see Loeve, 2017).

Consider a random element $X = (X_1, \ldots, X_p)$ on a product Hilbert space $\mathcal{H}$. Let G be an undirected graph with the vertex set $\{1, \ldots, p\}$. By convention, every vertex is understood to be adjacent to itself. We say that X has the graph G if it satisfies the *pairwise Markov property*, that is, for every $1 \leq i, j \leq p$ such that i and j are not adjacent in G we have

$$X_i \perp\!\!\!\perp X_j \mid X_k : k \neq i, j \tag{4.5}$$

or equivalently, we have

$$\langle f_i, X_i\rangle \perp\!\!\!\perp \langle f_j, X_j\rangle \mid \{\langle f_k, X_k\rangle : f_k \in \mathcal{H}_k, k \neq i, j\}$$

for every $f_i \in \mathcal{H}_i$ and $f_j \in \mathcal{H}_j$. We refer to G, thus defined, as the conditional independence graph of X. For a second-order X, we can similarly define the conditional uncorrelation graph of X, by simply replacing $\perp\!\!\!\perp$ in (4.5) with $\perp\!\!\!\perp_2$. We shall see eventually that the graph of a second-order random element X is intimately related to the entries of the precision operator matrix $\mathbf{H}$.

4.2.4 The Carleman-Fredholm Determinant

In order to properly generalize the multivariate graphical lasso to random elements, we will need to reformulate the graphical lasso objective function in terms of operator matrices. Although, one can think of fairly straightforward extensions to the definitions of trace and the ℓ_1 penalty from matrices to operators (or operator matrices), doing the same for the determinant is slightly more involved and non-standard.

Definition 4.1. *Let $\mathbf{H} \in \mathcal{L}_2(H)$ with the eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$. We define the Carleman-Fredholm determinant of $\mathbf{H}$ as*

$$\det{}_2(\mathbf{I} + \mathbf{H}) = \prod_{j=1}^{\infty} (1 + \lambda_j) e^{-\lambda_j} \tag{4.6}$$

It can be shown that the infinite product converges when $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$ and thus, the Carleman-Fredholm determinant is well-defined for all Hilbert–Schmidt operators. It is also known that the map $\mathbf{H} \mapsto \det_2(\mathbf{I} + \mathbf{H})$ is strictly log-concave, continuous everywhere

in $\|\cdot\|_2$ norm and Gateaux differentiable on $\{\mathbf{H} : -1 \notin \sigma(\mathbf{H})\} \subset \mathcal{L}_2(H)$ (see Appendix). Note that it is not simply the product of the eigenvalues $1 + \lambda_j$ of $\mathbf{I} + \mathbf{H}$. Defining the determinant simply as the product of the eigenvalues leads to what is known as the *Fredholm determinant*, which is defined only for trace-class operators. We shall see that the Carleman-Fredholm determinant appears most naturally when one attempts to correctly generalize the multivariate graphical lasso optimization function to covariance operators. For a more in-depth discussion on the generalization of determinants to operators, the interested reader is invited to consult Gohberg et al. (2000) and Simon (1977).

4.2.5 The Big $\mathcal{O}$ and Ω Notation

For $f, g : \mathbb{N} \to \mathbb{R}$, we write $f(n) = \mathcal{O}(g(n))$ if for some $C > 0$ and $n_0 \geq 1$, $f(n) \leq Cg(n)$ for $n \geq n_0$. The notation $f(n) = \Omega(g(n))$ denotes the inverse statement, that is, for some $c > 0$ and $n_0 \geq 1$, $f(n) \geq cg(n)$ for $n \geq n_0$.

4.3 Problem Statement

Let $X = (X_1, \ldots, X_n)$ be a second-order random element in $\mathcal{H}$ and $\{X^k\}_{k=1}^n$ be (not necessarily independent) realizations of X. Given an estimate $\hat{\mathbf{C}} = \hat{\mathbf{C}}_n(X^1, \ldots, X^n)$ of the covariance $\mathbf{C}$ of X, we are interested in estimating the graph G of the random elements $\{X_j\}_{j=1}^p$, which is given by the adjacency matrix $A = [A_{ij}]_{i,j=1}^p$ where

$$A_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 1_{\{\mathbf{H}^*_{ij} \neq \mathbf{0}\}} & \text{otherwise} \end{cases}$$

and $\mathbf{H}^* = \mathbf{R}^{-1} - \mathbf{I}$, with $\mathbf{R}$ being the correlation operator matrix of X. Essentially, we are interested in determining the non-zero off-diagonal entries of $\mathbf{H}^*$. Of particular interest to us is the sparse high-dimensional setting, where the number p of random elements X_j can be much larger than the number n of samples, and the graph G is known to be sparse in the sense that the maximum degree d of a vertex in G is much smaller than p.

When X is Gaussian, the graph G is identical to the conditional independence graph of $\{X_j\}_{j=1}^p$ and if X is not Gaussian, we can still interpret the off-diagonal zero entries of $\mathbf{H}$ in terms of the alternative notion of conditional uncorrelatedness (see Theorem 4.2). In either case, the graph describes the dependence structure of the random elements $\{X_j\}_{j=1}^p$ in the following sense: X_i is adjacent to X_j if and only if X_j can tell us something about X_i that other elements $\{X_k\}_{k \neq i,j}$ put together cannot. We use linear relationships between the random elements to judge what they tell us about each other and therefore, the graph G can be regarded as the graph of linear relationships between X_j. For

Gaussian X, the relationships are always linear and as a result the graph G is equal to the conditional independence graph.

4.4 Assumptions

In this section, we discuss the conditions we assume in order to prove the consistency and rates of convergence of functional graphical lasso. In particular, we explain how they can be interpreted in terms of properties of the distribution of X and how they may break down in certain cases.

4.4.1 Equivalence

The *support* of a measure (denoted by supp) is intuitively understood as the set on which the measure lives. It is formally defined as the largest closed set such that each of its open subsets have positive measure. Two measures are said to be *equivalent*, if they have the same support, and *singular*, if they have disjoint supports. In general, it is possible for two measures to be neither equivalent nor singular. But according to a classical result known as the Feldman-Hájek theorem, two Gaussian measures on a locally convex space must be either equivalent or singular.

Let $\mathbb{P}_{X_j}$ and $\mathbb{P}_X$ denote the random measures corresponding to X_j in the space $\mathcal{H}_j$ for every $1 \leq j \leq p$ and X in the product space $\mathcal{H} = \otimes_{j=1}^{p} \mathcal{H}_j$ respectively. We can also view the components X_j of X separately, and they would correspond to the product measure $\otimes_{j=1}^{p} \mathbb{P}_{X_j}$. If X is Gaussian, we shall make the following assumption:

Assumption 3 (Equivalence)**.** $\mathbb{P}_X$ *is equivalent to the product measure* $\otimes_{j=1}^{p} \mathbb{P}_{X_j}$*, that is,*

$$\text{supp } \mathbb{P}_X = \text{supp } \otimes_{j=1}^{p} \mathbb{P}_{X_j}.$$

According to Corollary 6.4.11 of Bogachev (1998), this seemingly innocuous statement is actually equivalent to saying that: (a) the off-diagonal entries of the correlation operator matrix $\mathbf{R}$ are Hilbert–Schmidt and (b) that there is a gap between the eigenvalues of $\mathbf{R}$ and 0, that is, $1 + \inf_j \lambda_j(\mathbf{R}_0) > 0$. This ensures that the correlation operator matrix $\mathbf{R}$ is invertible and that the operator $\mathbf{H} = \mathbf{R}^{-1} - \mathbf{I}$ is Hilbert–Schmidt (Lemma 4.1), implying that our optimization functional is well-defined at $\mathbf{H}$. It is important to note that for us this is a consequence of a "first principles" assumption imposed upon the observed random element X itself, namely the support condition. In contrast, previously this was a convenient assumption imposed upon intermediate quantities such as $\mathbf{R}$ so as to make certain operations (such as operator inversion or evaluation of the Hilbert–Schmidt norm), that one intends to perform, well-defined. To put it differently, Assumption 3 is a fundamental assumption as opposed to an operational assumption. It allows us to see the Hilbert–Schmidtness of $\mathbf{H}$ in a new light, whereas otherwise it would be seen merely

as the operator matrix counterpart of the familiar non-singularity of the covariance or correlation matrix assumed elsewhere in the multivariate literature (for example Rothman et al., 2008).

If X is not Gaussian, we shall assume the properties (a) and (b) directly through the following assumption instead:

Assumption 1* (Eigenvalue Gap)**.** *The cross-correlation operator matrix* $\mathbf{R}_0 = \mathbf{R} - \mathbf{I}$ *is Hilbert–Schmidt and the eigenvalues* $\{\lambda_j(\mathbf{R}_0)\}_{j=1}^{\infty}$ *of* $\mathbf{R}_0$ *satisfy*

$$1 + \inf_j \lambda_j(\mathbf{R}_0) > 0.$$

Define $\rho = 1 + |||\mathbf{R}_0|||_{2,\infty}$. It is worth pointing out that the assumption of an eigenvalue gap is not so harsh considering that $1 + \lambda_k(\mathbf{R}_0) \geq 0$ for $k \geq 1$ anyway since $\mathbf{R}$ is non-negative and $\lambda_k(\mathbf{R}_0) \to 0$ as $k \to \infty$ because $\mathbf{R}_0$ is Hilbert–Schmidt.

Lemma 4.1. *If* $1 + \inf_j \lambda_j(\mathbf{R}_0) > 0$ *and* $\mathbf{R}_0 = \mathbf{R} - \mathbf{I}$ *is Hilbert–Schmidt, then so is* $\mathbf{H}^* = \mathbf{R}^{-1} - \mathbf{I}$.

Proof of Lemma 4.1. Let $c = 1 + \inf_j \lambda_j(\mathbf{R}_0)$. By the spectral mapping theorem, $\lambda_k(\mathbf{H}^*) = [1 + \lambda_k(\mathbf{R}_0)]^{-1} - 1 = -\lambda_k(\mathbf{R}_0)[1 + \lambda_k(\mathbf{R}_0)]^{-1}$ and therefore,

$$\|\mathbf{H}^*\|_2^2 = \sum_{k=1}^{\infty} \frac{\lambda_k^2(\mathbf{R}_0)}{[1 + \lambda_k(\mathbf{R}_0)]^2} \leq \frac{1}{c^2} \sum_{k=1}^{\infty} \lambda_k^2(\mathbf{R}_0) = \frac{\|\mathbf{R}_0\|_2^2}{c^2} < \infty.$$

Hence proved. □

In fact, we shall see in Section 4.5.1 that Assumption 3 allows us to treat the product measure $\otimes_{j=1}^{p} \mathbb{P}_{X_j}$ as a *reference measure* to describe the distribution of X much like the Lebesgue measure serves to do the same in Euclidean spaces.

Remark 4.1. *It is not difficult to imagine a scenario where Assumption 3 fails to hold. Consider a Gaussian process* X *on the unit interval* $[0, 1]$ *with continuous sample paths which corresponds to a Gaussian measure in the space* $L^2[0, 1]$. *Then* X *can be thought of as a pair* (X_1, X_2) *where* X_1 *and* X_2 *are the processes (and random elements) corresponding to the restrictions of* X *to the intervals* $[0, 1/2]$ *and* $(1/2, 1]$ *respectively. Let* Y *be the random element (process) corresponding to the product measure* $\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. *The sample paths of the process* Y *are almost surely discontinuous at* $t = 1/2$ *while that of the process* X *are almost surely continuous throughout. Thus the measures* $\mathbb{P}_X$ *and* $\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$ *are singular.*

Remark 4.2. *Assumption* 1^* *here is strictly weaker than Assumption 1 of Lee et al. (2022) which states that for every* $1 \leq i \leq p$, *the regression operator* $\mathbf{C}_{-i,-i}^{\dagger}\mathbf{C}_{-i,i}$ *is Hilbert-Schmidt (see the appendix for a proof).*

4.4.2 Incoherence

Let Γ denote the outer product of the operator matrix $\mathbf{R}$ with itself, i.e.

$$\Gamma = \mathbf{R} \otimes \mathbf{R} = [\mathbf{R}_{ij} \otimes \mathbf{R}_{kl}]_{i,j,k,l=1}^{p} .$$

Equivalently, Γ can be thought of as an operator matrix indexed by the pairs (i, j) and (k, l) with $\Gamma_{(i,j)(k,l)} = \mathbf{R}_{ij} \otimes \mathbf{R}_{kl}$ just as $\mathbf{R}$ is indexed by the vertices i, j in $\mathbf{R}_{ij}$. For two sets A and B of vertex pairs we can write the *submatrix* Γ_{AB} as

$$\Gamma_{AB} = \left[\Gamma_{(i,j)(k,l)}\right]_{(i,j)\in A,(k,l)\in B} .$$

Finally, observe that Γ can be thought of as an operator on the product space of the tensor product spaces $\mathcal{H}_i \otimes \mathcal{H}_j$ with $\Gamma\mathbf{A} = \mathbf{RAR}$. Because $\mathbf{R}$ is invertible (under Assumption 3), it follows that so is Γ with $\Gamma^{-1} = \mathbf{R}^{-1} \otimes \mathbf{R}^{-1}$.

Let S denote the set of (i, j) such that i and j are adjacent in G or equivalently, (i, j) corresponds to an edge. Naturally, S^c denotes its complement. Then Γ_{SS} can be shown to be invertible by virtue of being a principal submatrix of Γ. The following assumption shall serve as the functional analogue of the familiar mutual incoherence condition from Ravikumar et al. (2011).

Assumption 4 (Incoherence)**.** *For some $\alpha > 0$, we have*

$$\max_{e\in S^c} \|\Gamma_{eS}\Gamma_{SS}^{-1}\|_{2,1} \leq 1 - \alpha. \tag{4.7}$$

Notice that like $\mathbf{R}$, the inverse of submatrix Γ_{SS} can be written as the sum of identity and a Hilbert–Schmidt operator matrix. Indeed, for $\mathbf{A} = (\mathbf{R}_0)_S$, where $\mathbf{R}_0 = \mathbf{R} - \mathbf{I}$, we can write

$$\Gamma_{SS}^{-1} = [\mathbf{I} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I} + \mathbf{A} \otimes \mathbf{A}]^{-1} = \mathbf{I} + \Lambda_{SS}$$

where Λ_{SS} is Hilbert–Schmidt by Lemma 4.1 because the operator $\mathbf{I}\otimes\mathbf{A} + \mathbf{A}\otimes\mathbf{I} + \mathbf{A}\otimes\mathbf{A}$ inside the inverse is itself Hilbert–Schmidt. Define $\gamma = 1 + |||\Lambda_{SS}|||_{2,\infty}$.

Intuitively speaking, if we could think of $\mathbf{R}$ as the covariance operator of a zero mean random element $Z = (Z_j)_{j=1}^{p}$ with $\mathbf{R}_{ij} = \mathbb{E}[Z_i \otimes Z_j]$, we would consider the random elements

$$Y_{(i,j)} = Z_i \otimes Z_j - \mathbb{E}[Z_i \otimes Z_j]$$

for $1 \leq i, j \leq p$. Using the same tools as in multivariate analysis (doing a Taylor expansion of the moment generating function), it can be shown that $\Gamma_{(i,j)(k,l)} = \mathbb{E}\left[Y_{(i,j)} \otimes Y_{(k,l)}\right]$ when Z_j are Gaussian. Let $Y_S = \{Y_e : e \in S\}$. Assumption 4 can now be expressed as

$$\max_{e\in S^c} \left\|\mathbb{E}[Y_e \otimes Y_S]\,\mathbb{E}[Y_S \otimes Y_S]^{-1}\right\|_{2,1} \leq 1 - \alpha.$$

Notice that Γ_{eS} is the cross-covariance of Y_e with Y_S and Γ_{SS} is the covariance of Y_S. If we were to find the best linear predictor of Y_e using Y_S, the linear coefficients would be given by $\Gamma_{eS}\Gamma_{SS}^{-1}$. Assumption 4 is essentially saying that these coefficients cannot be to large: none of the "non-edges" Y_e (with $e \in S^c$) are highly correlated with the "edges" Y_S and therefore one cannot predict the "non-edges" Y_e from the "edges" Y_S too well.

Of course, strictly speaking, the operator $\mathbf{R}$ is not the covariance operator of any random element in Hilbert space due to not being trace-class. We believe that this explanation can be made rigorous by treating $\mathbf{R}$ as the covariance operator of Gaussian random element on a suitably chosen locally convex topological vector space (where covariance operators do not have to be trace-class). Even without the technical details formalizing this, the intuition is just as valuable.

Incoherence is the assumption that enables us to exploit sparsity. It seems that incoherence is an indispensable assumption for multivariate graphical lasso and weaker assumptions lead to substantially weaker rates of convergence (Rothman et al., 2008). It would thus be unreasonable to expect functional graphical lasso to work for anything less.

4.4.3 Regularity

To execute our method, we shall need to estimate the correlation operator matrix $\mathbf{R}$. We do this by first estimating the covariance operator $\mathbf{C}$ and then resolving the following linear problem for $\mathbf{R}$:

$$[\operatorname{dg}\mathbf{C}]^{1/2}\mathbf{R}[\operatorname{dg}\mathbf{C}]^{1/2} = \mathbf{C}. \tag{4.8}$$

Note that the problem is ill-posed because $\mathbf{C}$ (and $\operatorname{dg}\mathbf{C}$) are compact operators. To ensure reasonable rates of convergence for this estimation procedure we need to impose the following condition on $\mathbf{R}$:

Assumption 5 (Regularity)**.** *For some $0 < \beta \leq 1$, we have $\mathbf{R} = [\operatorname{dg}\mathbf{C}]^{\beta}\Phi_0[\operatorname{dg}\mathbf{C}]^{\beta}$ for some Hilbert–Schmidt operator matrix Φ_0, whose diagonal entries are zero.*

In principle, it is possible that $\mathbf{R} = [\operatorname{dg}\mathbf{C}]^{\beta}\Phi_0[\operatorname{dg}\mathbf{C}]^{\beta}$ with $\beta > 1$ but for our purpose this situation is essentially identical to the case $\beta = 1$. Note that for $0 < \beta' < \beta$, the condition $\mathbf{R} = [\operatorname{dg}\mathbf{C}]^{\beta}\Phi_0[\operatorname{dg}\mathbf{C}]^{\beta}$ implies $\mathbf{R} = [\operatorname{dg}\mathbf{C}]^{\beta'}\Phi_0'[\operatorname{dg}\mathbf{C}]^{\beta'}$ for some Φ_0', and therefore, Assumption 5 holds for $\beta = 1$ if it holds for $\beta > 1$. Note that Assumption 5 is equivalent to saying that for every $i \neq j$,

$$\sum_{k,l=1}^{\infty}\left[\frac{1}{\mu_k\lambda_l}\right]^{1+2\beta}|\langle e_k, \mathbf{C}_{ij}f_l\rangle|^2 < \infty$$

for some $\beta > 0$, where $\{(\mu_k, e_k)\}_{k=1}^{\infty}$ and $\{(\lambda_l, f_l)\}_{l=1}^{\infty}$ are the eigenpairs of $\mathbf{C}_{ii}$ and $\mathbf{C}_{jj}$, respectively. Essentially, this means that $\mathbf{C}_{ij}$ admits an efficient or sparse representation in the eigenbases of $\mathbf{C}_{ii}$ and $\mathbf{C}_{jj}$.

In fact, it is a classical result in inverse problem theory that in the absence of such *source conditions*, the rate of convergence for the solution of an infinite-dimensional linear inverse problem can be arbitrarily slow (Hanke, 2017). In the language of numerical linear algebra, Assumption 5 means that the operator $\mathbf{C}$ is intrinsically preconditioned for inversion by $\operatorname{dg}\mathbf{C}$. Moreover, the usage of such regularity conditions is standard in the literature (c.f. Li and Solea, 2018b)). We shall see that the performance of our procedure shall depend critically on the maximum degree d of the graph and that this dependence is mediated by β.

4.5 Methodology and Philosophy

We now describe our two-step estimation procedure to recover the graph G of X given an estimate $\hat{\mathbf{C}}$ of the covariance $\mathbf{C}$ of X.

Firstly, we estimate the correlation operator matrix $\mathbf{R}$ of X. Because $\mathbf{C}$ is known only approximately, estimating $\mathbf{R}$ using Equation (4.8) presents an ill-posed linear problem. We use the regularized estimator $\hat{\mathbf{R}} = [\hat{\mathbf{R}}_{ij}]_{i,j=1}^{p}$ given by

$$\hat{\mathbf{R}}_{ij} = \begin{cases} \mathbf{I} & \text{for } i = j, \text{ and} \\ [\epsilon_n \mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}_{ii}]^{-1/2} \hat{\mathbf{C}}_{ij} [\epsilon_n \mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}_{jj}]^{-1/2} & \text{for } i \neq j, \end{cases}$$

where ϵ_n serves as a tuning parameter. Recall that $\mathbf{R}_{ij} = \mathbf{I}$ when $i = j$, so we need not burden ourselves with its estimation.

Secondly, we minimize the proposed objective functional $\mathcal{F}$ over the space of Hilbert–Schmidt operators $\mathbf{H}$ on $\mathcal{H}$ given by

$$\mathcal{F}[\mathbf{H}] = \begin{cases} \operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I}+\mathbf{H}) + \lambda_n \|\mathbf{H}_0\|_{2,1} & \text{if } \mathbf{I}+\mathbf{H} > \mathbf{0}, \text{ and} \\ \infty & \text{otherwise.} \end{cases} \tag{4.9}$$

where the trace $\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0)$ can be expressed as $\sum_{i\neq j} \operatorname{tr}(\mathbf{H}_{ij}\hat{\mathbf{R}}_{ij})$ and $\|\mathbf{H}_0\|_{2,1} = \sum_{i\neq j} \|\mathbf{H}_{ij}\|_2$ is the ℓ_1-norm of the Hilbert–Schmidt norms of the off-diagonal entries of $\mathbf{H}$ and can be likened to a group lasso penalty proposed in Yuan and Lin (2006). Note that the trace is well-defined since both $\mathbf{H}$ and $\hat{\mathbf{R}}_0$ are Hilbert–Schmidt implying that the product $\mathbf{H}\hat{\mathbf{R}}_0$ is trace-class. Thus, $\mathcal{F}[\mathbf{H}]$ is well-defined for a Hilbert–Schmidt operator $\mathbf{H}$. Furthermore as an eigenvalue of $\mathbf{H}$ approaches -1 from above, $\det_2(\mathbf{I}+\mathbf{H})$ converges to 0 and its logarithm grows without bound implying that $\mathcal{F}[\mathbf{H}] \to \infty$. The piece-wise definition is thus quite reasonable and in fact, makes $\mathcal{F}$ into a coercive, strictly convex functional which is continuous in the extended sense. This will ensure that $\mathcal{F}$ always has a unique minimum and minimizer (Theorem 4.3). The nonzero entries of the minimizer $\hat{\mathbf{H}} = \arg\min_{\mathbf{H}} \mathcal{F}[\mathbf{H}]$ describe the graph G in that $\hat{\mathbf{H}}_{ij} \neq \mathbf{0}$ if and only if i and j are adjacent.

To summarize, given an estimator $\hat{\mathbf{C}}$ of the covariance $\mathbf{C}$ we have the following procedure to estimate the graph G of X:

Step 1. Estimation. Estimate the correlation operator matrix $\hat{\mathbf{R}}$ as follows:

$$\hat{\mathbf{R}} \quad = \mathbf{I} + [\epsilon_n \mathbf{I} + \operatorname{dg} \hat{\mathbf{C}}]^{-1/2} \hat{\mathbf{C}}_0 [\epsilon_n \mathbf{I} + \operatorname{dg} \hat{\mathbf{C}}]^{-1/2}. \tag{4.10}$$

Step 2. Minimization. Compute $\hat{\mathbf{H}}$:

$$\hat{\mathbf{H}} \quad = \arg\min \left[\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I} + \mathbf{H}) + \lambda_n \|\mathbf{H}_0\|_{2,1} \right], \tag{4.11}$$

where the minimum is taken over all Hilbert–Schmidt operators $\mathbf{H}$ such that $\mathbf{I} + \mathbf{H} > \mathbf{0}$. The adjacency matrix $\hat{A} = [\hat{A}_{ij}]_{i,j=1}^{p}$ of the estimate $\hat{G}$ of the graph G is given by $\hat{A}_{ij} = 1_{\{\hat{\mathbf{H}}_{ij} \neq \mathbf{0}\}}$ for $i \neq j$ and 1 otherwise.

A variety of methods have been used in the literature for choosing the tuning parameter ϵ_n. For example, Li and Solea (2018b) use generalized cross validation. We prefer a simpler approach inspired by Waghmare and Panaretos (2023) which involves minimizing (3.14). The tuning parameter λ_n can be chosen using stability selection proposed by Meinshausen and Bühlmann (2010). We show in Section 4.6.2, that the rate at which $\epsilon_n \to 0$ and $\lambda_n \to 0$ is related to how well $\hat{\mathbf{C}}$ concentrates around $\mathbf{C}$ together with the regularity β.

In Section 4.7, we describe how the quantities involved are actually calculated in practice and how the minimization procedure is implemented using the alternating direction method of multipliers (ADMM) algorithm.

4.5.1 Penalized Log-likelihood Maximization

The usage of likelihood maximization techniques in functional data analysis is largely impeded by the absence of a compelling reference measure in infinite dimensional function spaces, with respect to which the likelihood can be defined. In multivariate statistics, we are mostly concerned with finite dimensional Euclidean spaces where the Lebesgue measure serves as a *de facto* reference measure due to its translation invariance and the accompanying *indifference* to points in the space, even though it is not a probability measure. However, according to a classical result, every nontrivial translation invariant Borel measure on an infinite dimensional separable Banach space, is bound to assign infinite measure to every open set. Consequently, translation invariance proves to be an excessively harsh criterion for a reference measure to satisfy in infinite dimensions.

We propose to use the product $\mathbb{Q} = \otimes_{j=1}^{p} \mathbb{P}_{X_j}$ as our reference measure. It is not translation invariant like the Lebesgue measure. However, unlike the Lebesgue measure, it is a

probability measure and represents an actually possible scenario, which is when X_j are all independent —arguably the simplest of all possible scenarios under consideration. Of course, we do not know $\mathbb{Q}$ a priori but it turns out that we can, in a certain sense, evaluate the log-likelihood without knowing $\mathbb{Q}$ exactly using what amounts to a renormalization technique.

Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be two zero-mean Gaussian measures with the marginals $\{\mathbb{P}_{X_j}\}_{j=1}^p$ which are equivalent to $\mathbb{Q} = \otimes_{j=1}^p \mathbb{P}_{X_j}$. By Corollary 6.4.11 of Bogachev (1998), we can write the covariance operators of $\mathbb{P}$ and $\tilde{\mathbb{P}}$ as

$$\begin{aligned} \mathbf{C}_{\mathbb{P}} &= \mathbf{C}_{\mathbb{Q}}^{1/2}(\mathbf{I} + \mathbf{R}_0)\mathbf{C}_{\mathbb{Q}}^{1/2} \\ \mathbf{C}_{\tilde{\mathbb{P}}} &= \mathbf{C}_{\mathbb{Q}}^{1/2}(\mathbf{I} + \tilde{\mathbf{R}}_0)\mathbf{C}_{\mathbb{Q}}^{1/2} \end{aligned}$$

where $\mathbf{C}_{\mathbb{Q}} = \operatorname{dg}\mathbf{C} = \operatorname{dg}\mathbf{C}_{\mathbb{P}} = \operatorname{dg}\mathbf{C}_{\tilde{\mathbb{P}}}$, and $\mathbf{R}_0, \tilde{\mathbf{R}}_0$ are Hilbert–Schmidt operators with diagonal entries all zero and eigenvalues separated from -1 in the sense of Assumption 1^*. The average log-likelihood of $\mathbb{P}$ with respect to $\mathbb{Q}$ evaluated with an infinite number of samples drawn from $\tilde{\mathbb{P}}$ evaluates to the following expectation:

Lemma 4.2. *We have*

$$\int \log\left[\frac{d\mathbb{P}}{d\mathbb{Q}}\right] d\tilde{\mathbb{P}} = -\frac{1}{2}[\operatorname{tr}(\mathbf{H}\tilde{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I} + \mathbf{H})] \tag{4.12}$$

where $\mathbf{H} = (\mathbf{I} + \mathbf{R}_0)^{-1} - \mathbf{I}$.

If we think of $\tilde{\mathbb{P}}$ as the empirical measure generated from the samples $\{X^k\}_{k=1}^n$ drawn from $\mathbb{P}$, the expression on the left of Equation (4.12) is exactly the log-likelihood of $\mathbb{P}$ with respect to $\mathbb{Q}$. Of course, we require $\tilde{\mathbb{P}}$ to be a Gaussian measure which an empirical measure cannot be. So, we treat $\tilde{\mathbb{P}}$ as the zero-mean Gaussian measure with the covariance operator $\mathbf{C}_{\tilde{\mathbb{P}}} \approx \hat{\mathbf{C}}$. Roughly speaking, this would mean that $\tilde{\mathbf{R}}_0 \approx \hat{\mathbf{R}}_0$ and the right hand side of Equation (4.12) becomes

$$-\frac{1}{2}[\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I} + \mathbf{H})] \tag{4.13}$$

which makes for a compelling substitute for the sought after log-likelihood and corresponds to the first two terms of our objective functional $\mathcal{F}[\mathbf{H}]$. The idea of using a Gaussian measure $\tilde{\mathbb{P}}$ corresponding approximately to the empirical covariance operator $\hat{\mathbf{C}}$ instead of the empirical measure to evaluate the log-likelihood is reminiscent of the idea of parametric bootstrap.

Essentially, we have conditioned the log-likelihood on the prior knowledge that $\tilde{\mathbb{P}}$ is a Gaussian. Deriving expression (4.13) is also possible with a more direct approach using the empirical measure, but requires usingsome renormalization techniques such as truncation and regularization needed to deal with the infinities arising from the difference

in the supports of the empirical and true versions of the measures involved, all of which we have managed to avoid here.

Since we know that the sparse edge structure of the graph is encoded in the non-zero off-diagonal entries of the Hilbert–Schmidt operator matrix $\mathbf{H}$, it seems natural to penalize (4.13) with the ℓ_1 norm of the norms $\|\mathbf{H}_{ij}\|_2$. This gives

$$-\frac{1}{2}\mathcal{F}[\mathbf{H}] = -\frac{1}{2}\Big[\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I}+\mathbf{H})\Big] - \frac{\lambda_n}{2}\sum_{i\neq j}\|\mathbf{H}_{ij}\|_2. \tag{4.14}$$

We have thus shown that our method can be thought of as penalized log-likelihood maximization, albeit with a few leaps of imagination.

4.5.2 Constrained Divergence Minimization

The continuity and convexity of the functional $\mathcal{F}$ implies that the optimization problem (4.11) has an equivalent dual formulation.

In fact, minimizing $\mathcal{F}$ is actually equivalent to evaluating the convex conjugate $\mathcal{G}^*$ of the functional $\mathcal{G}[\mathbf{A}] = \log\det{}_2(\mathbf{I}+\mathbf{A}) - \lambda_n\|\mathbf{A}_0\|_{2,1}$ at $-\hat{\mathbf{R}}_0$. Indeed,

$$\begin{aligned}\min{}_{\mathbf{A}}\,\mathcal{F}[\mathbf{A}] &= -\max_{\mathbf{A}}\Big[\operatorname{tr}(\mathbf{A}(-\hat{\mathbf{R}}_0)) + \log\det{}_2(\mathbf{I}+\mathbf{A}) - \lambda_n\|\mathbf{A}_0\|_{2,1}\Big]\\ &= -\mathcal{G}^*[-\hat{\mathbf{R}}_0].\end{aligned}$$

Now, the convex conjugate of $\mathbf{A}\mapsto -\log\det{}_2(\mathbf{I}+\mathbf{A})$ is actually twice the Kullback-Leibler divergence of the Gaussian measure with the correlation operator $\mathbf{I}-\mathbf{B}$ assuming $\operatorname{dg}\mathbf{B} = \mathbf{0}$. This can be verified from Equation (4.12) which yields for $\tilde{\mathbb{P}} = \mathbb{P}$, the Kullback-Leibler divergence $\mathcal{D}[\mathbf{R}_0]$ of a Gaussian measure $\mathbb{P}$ with correlation operator $\mathbf{R} = \mathbf{I}+\mathbf{R}_0$ with respect to its product measure $\mathbb{Q}$, to be

$$\mathcal{D}[\mathbf{R}_0] = -\frac{1}{2}\log\det{}_2(\mathbf{I}+\mathbf{R}_0)$$

Using infimal convolution we can combine this convex conjugate with that of $\mathbf{A}\mapsto\|\mathbf{A}_0\|_{2,\infty}$ and rewrite the convex conjugate of $\mathcal{G}$ as the solution of a constrained minimization or maximization problem.

Theorem 4.1. *The optimization problem (4.11) satisfies*

$$\min_{\mathbf{A}}\mathcal{F}[\mathbf{A}] = -2\min_{\mathbf{B}}\mathcal{D}[\mathbf{B}_0] = \max_{\mathbf{B}}\log\det{}_2(\mathbf{I}+\mathbf{B}_0), \tag{4.15}$$

$$\arg\min_{\mathbf{A}}\mathcal{F}[\mathbf{A}] = \arg\min_{\mathbf{B}}\mathcal{D}[\mathbf{B}_0] = \arg\max_{\mathbf{B}}\det{}_2(\mathbf{I}+\mathbf{B}_0) \tag{4.16}$$

where $\mathbf{B}$ *are Hilbert–Schmidt operator matrices with* $\operatorname{dg}\mathbf{B} = \mathbf{0}$ *such that* $\|\mathbf{B}_0 - \hat{\mathbf{R}}_0\|_{2,\infty} \leq \lambda_n$.

In other words, the optimization problem (4.11) is equivalent to minimizing the Kullback-Leibler divergence $\mathcal{D}[\mathbf{B}_0]$ with respect to $\mathbb{Q}$ or maximizing the Carleman-Fredholm determinant $\det_2(\mathbf{I}+\mathbf{B}_0)$ under the constraint that every off-diagonal entry $\mathbf{B}_{ij}$ stays within λ_n in Hilbert–Schmidt distance from $\hat{\mathbf{R}}_{ij}$. Our method can thus be seen as constrained minimization of the Kullback-Leibler divergence.

4.5.3 Limit of the Multivariate Graphical Lasso

Our functional graphical lasso can also be motivated from its multivariate counterpart by correcting the multivariate objective function so as to obtain meaningful limiting behaviour as the grid resolution increases and finding an appropriate replacement for the penalty term.

Suppose that the spaces $\mathcal{H}_i$ are composed of continuous functions on the sets U_i and $X=(X_1,\dots,X_p)$ is a Gaussian random element on $\mathcal{H}$. We construct a grid $\{u_{ij}\}_{j=1}^m$ of m points on each set U_i. A natural way to study the graph of $X=(X_1,\dots,X_p)$ is to study the graph of the random vector

$$\mathsf{X}=(X_1(u_{11}),\dots,X_1(u_{1m}),X_2(u_{21}),\dots,X_2(u_{2m}),\dots,X_p(u_{p1}),\dots,X_p(u_{pm})).$$

Thus we can apply the multivariate graphical lasso to the random vectors X^k corresponding to the independent realizations X^k of X and see what happens as $m\to\infty$ and the grids $\{u_{ij}\}_{j=1}^m$ populate the sets U_i more and more densely.

To this end, we evaluate the empirical covariance estimator $\hat{\mathsf{C}}$ of the covariance C of X. Notice that C and $\hat{\mathsf{C}}$ are simply restrictions to the grid $\{u_{ij}\}_{i=1,j=1}^{p,m}$ of the continuous integral kernels of the operators $\mathbf{C}$ and $\hat{\mathbf{C}}$. Now, consider the objective function given by

$$\mathsf{F}(\Theta)=\mathsf{tr}(\Theta\hat{\mathsf{C}})-\mathsf{log\ det}(\Theta)+\lambda\sum\nolimits_{\mathsf{i}\neq\mathsf{j}}|\Theta_{\mathsf{ij}}|,$$

where Θ is a possible candidate for the precision matrix C^{-1}. Observe that we can write

$$\hat{\mathsf{C}}=\tfrac{1}{\mathsf{m}}\mathsf{D}_{\mathsf{c}}^{1/2}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\hat{\mathsf{R}}_0)\mathsf{D}_{\mathsf{c}}^{1/2}$$

where $\mathsf{D_c}$ and $\hat{\mathsf{R}}_0$ are approximately the restrictions to the grid of the integral kernels of the Hilbert–Schmidt operators $\operatorname{dg}\hat{\mathbf{C}}$ and $\hat{\mathbf{R}}_0$. The factors $\frac{1}{\mathsf{m}}$ are a result of having to replicate the operations $\mathbf{A}_{ij}f_j(x)=\int_{U_j}A(x,y)f(y)dy$ whose discrete approximation is given by $l\mapsto\frac{1}{m}\sum_{k=1}^m A(u_{il},u_{jk})f(u_{jk})$. This suggests that we should parametrize Θ as in

$$\Theta^{-1}=\tfrac{1}{\mathsf{m}}\mathsf{D}_{\mathsf{c}}^{1/2}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H})^{-1}\mathsf{D}_{\mathsf{c}}^{1/2}$$

in terms of a rough approximation H to the grid $\{u_{ij}\}_{i=1,j=1}^{p,m}$ of $\mathbf{H}$. We can now write

the first two terms as

$$\begin{aligned}
\mathsf{tr}(\Theta\hat{\mathsf{C}}) - \mathsf{log\ det}(\Theta) &= \mathsf{tr}([\tfrac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c}^{1/2}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H})^{-1}\mathsf{D}_\mathsf{c}^{1/2}]^{-1}\hat{\mathsf{C}}) - \mathsf{log\ det}([\tfrac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c}^{1/2}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H})^{-1}\mathsf{D}_\mathsf{c}^{1/2}]^{-1}) \\
&= \mathsf{tr}((\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H})[\mathsf{m}\mathsf{D}_\mathsf{c}^{-1/2}\hat{\mathsf{C}}\mathsf{D}_\mathsf{c}^{-1/2}]) - \mathsf{log\ det}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H}) + \mathsf{log\ det}(\tfrac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c}) \\
&= \mathsf{tr}((\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H})(\mathsf{I}+\tfrac{1}{\mathsf{m}}\hat{\mathsf{R}}_0)) - \mathsf{log\ det}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H}) + \mathsf{log\ det}(\tfrac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c}) \\
&= \mathsf{tr}(\tfrac{1}{\mathsf{m}^2}\mathsf{H}\hat{\mathsf{R}}_0) + \mathsf{tr}(\tfrac{1}{\mathsf{m}}\mathsf{H}) - \mathsf{log\ det}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H}) + \mathsf{pm} + \mathsf{log\ det}(\tfrac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c})
\end{aligned}$$

using the fact that $\operatorname{tr}\hat{\mathsf{R}}_0 \approx 0$ and $\mathsf{tr}(\mathsf{I}) = \mathsf{pm}$. Assuming polynomial decay of eigenvalues, $\mathsf{log\ det}(\frac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c}) \approx \mathsf{C}\mathsf{p}\sum_{\mathsf{j}=1}^{\mathsf{m}}\mathsf{log}(1/\mathsf{j}^\alpha) \approx -\mathsf{C}\alpha\mathsf{pm}\ \mathsf{log}(\mathsf{m})$ which diverges to $-\infty$ faster than the second last term pm diverges to ∞. This suggests that the above expression diverges to $-\infty$ and explains why the multivariate graphical lasso is not stable with respect to grid resolution. If we ignore the terms pm and $\mathsf{log\ det}(\frac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c})$ with ill-defined limits or alternatively, introduce the correction term $-\mathsf{pm} - \mathsf{log\ det}(\frac{1}{\mathsf{m}}\mathsf{D}_\mathsf{c})$, we can obtain a nontrivial limiting behaviour from the above expression, which gives

$$\begin{aligned}
\mathsf{tr}(\tfrac{1}{\mathsf{m}^2}\mathsf{H}\hat{\mathsf{R}}_0) + \mathsf{tr}(\tfrac{1}{\mathsf{m}}\mathsf{H}) - \mathsf{log\ det}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H}) &= \mathsf{tr}(\tfrac{1}{\mathsf{m}^2}\mathsf{H}\hat{\mathsf{R}}_0) - \left[\mathsf{log\ det}(\mathsf{I}+\tfrac{1}{\mathsf{m}}\mathsf{H}) - \mathsf{tr}(\tfrac{1}{\mathsf{m}}\mathsf{H})\right] \\
&\to \operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I}+\mathbf{H})
\end{aligned}$$

as $\mathsf{m} \to \infty$. Because we know from Theorem 4.2 how the information about the graph is in the off-diagonal entries of $\mathbf{H}$, it now makes sense to penalize the above expression accordingly, thus recovering our objective functional $\mathcal{F}[\mathbf{H}]$. Section 4.7.1 contains a longer discussion on how operators are discretized and on how the above formulas may be obtained.

4.6 Theoretical Guarantees

4.6.1 Identifiability and Well-posedness

If X is a Gaussian random element and the correlation operator matrix $\mathbf{R}$ is invertible, the pairwise Markov property can be expressed in a particularly elegant way in terms of the precision operator matrix $\mathbf{H}$. Even if X is not Gaussian, the same applies for the conditional uncorrelatedness version of the pairwise Markov property.

Theorem 4.2 (Precision Operator and Conditional Independence)**.** *Let $X = (X_1, \ldots, X_p)$ be a second-order random element in the product Hilbert space $\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_p$.*

1. *Under Assumption 3, if X is Gaussian then the correlation operator matrix $\mathbf{R}$ is invertible and for $1 \le i, j \le p$ with $i \ne j$, we have the correspondence*

$$X_i \perp\!\!\!\perp X_j \mid X_k : k \ne i, j \quad \text{if and only if} \quad \mathbf{H}_{ij}^* = \mathbf{0},$$

2. *Under Assumption 1*, the correlation operator matrix $\mathbf{R}$ is invertible and for*

$1 \leq i, j \leq p$ with $i \neq j$, we have the correspondence

$$X_i \perp\!\!\!\perp_2 X_j \mid X_k : k \neq i, j \quad \text{if and only if} \quad \mathbf{H}^*_{ij} = \mathbf{0},$$

where $\mathbf{H}^* = [\mathbf{H}^*_{ij}]_{i,j=1}^p$ is the precision operator matrix of X.

In other words, the off-diagonal zero entries of the adjacency matrix of G (which represent edges) correspond precisely to the off-diagonal zero entries of the precision operator matrix $\mathbf{H}^*$. This is the Hilbert space generalization of the familiar result for Gaussian graphical models where the role of the precision matrix is served by the inverse of the covariance instead.

As mentioned before, $\mathcal{F}$ is coercive, strictly convex and continuous in the extended sense and these are sufficient conditions for a functional to admit a unique minimum and minimizer in Hilbert space. Our theoretical analysis depends critically on exploiting the stationary condition (4.17).

Theorem 4.3. *The optimization problem (4.11) admits a unique solution $\hat{\mathbf{H}}$ for every $\lambda_n > 0$ and estimated correlation operator $\hat{\mathbf{R}}$ which satisfies*

$$\hat{\mathbf{R}} - (\mathbf{I} + \hat{\mathbf{H}})^{-1} + \lambda_n \hat{\mathbf{Z}} = \mathbf{0} \tag{4.17}$$

for some $\hat{\mathbf{Z}} \in \partial \|\hat{\mathbf{H}}_0\|_{2,1}$, where $\partial \|\hat{\mathbf{H}}_0\|_{2,1}$ denotes the subdifferential (the set of subgradients) of $\mathbf{H} \mapsto \|\mathbf{H}_0\|_{2,1}$ at $\mathbf{H} = \hat{\mathbf{H}}$.

4.6.2 Finite Sample Theory

We begin by introducing some language from Ravikumar et al. (2011) which will be useful for describing the tail behaviour of our estimators. For $\delta_* > 0$ and a function $f : \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}_+$, which is monotonically increasing in both arguments, we say that an estimator $\hat{\mathbf{A}} = [\hat{\mathbf{A}}_{ij}]_{i,j=1}^p$ of an operator matrix $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1}^p$ satisfies a *tail condition with the parameters f and δ_** if for every $n \geq 1$ and $0 < \delta < \delta_*$ we have

$$\mathbb{P}[\|\hat{\mathbf{A}}_{ij} - \mathbf{A}_{ij}\|_2 \geq \delta] \leq 1/f(n, \delta)$$

To handle the behaviour of $\hat{\mathbf{A}}$ under such tail conditions, we define

$$\bar{n}_f(\delta, r) = \max\{n : f(n, \delta) \leq r\} \quad \text{and} \quad \bar{\delta}_f(r, n) = \max\{\delta : f(n, \delta) \leq r\}.$$

Essentially, $\bar{n}_f(\delta, r)$ is the smallest n and $\bar{\delta}_f(r, n)$ is the smallest δ for which $\|\hat{\mathbf{A}} - \mathbf{A}\|_2 < \delta$ with probability at least $1 - 1/r$.

Let $X = (X_1, \ldots, X_p)$ be a second-order random element in $\mathcal{H}$ with the covariance $\mathbf{C}$. Let $\{X^k\}_{k=1}^n$ be n independent realizations of X and $\hat{\mathbf{C}} = \hat{\mathbf{C}}(X^1, \ldots, X^n)$ be an estimator

of $\mathbf{C}$ which satisfies the tail condition with the parameters f and δ_*. Then these tail conditions together with the regularity conditions of Assumption 5, naturally lead to similar conditions on the corresponding estimator $\hat{\mathbf{R}}$ of the correlation operator $\mathbf{R}$. Recall that $\rho = 1 + |||\mathbf{R}_0|||_{2,\infty}$ and $\gamma = 1 + |||\Lambda_{SS}|||_{2,\infty}$.

Theorem 4.4. *Let $\hat{\mathbf{C}}$ be an estimator of $\mathbf{C}$ satisfying the tail condition with the parameters f and δ_* such that $\mathrm{dg}(\hat{\mathbf{C}})$ is non-negative. Under Assumption 5, for $\epsilon_n = \delta^{\frac{1}{1+\beta}}$, the corresponding estimator $\hat{\mathbf{R}}$ for the correlation $\mathbf{R}$ satisfies for $i \neq j$*

$$\mathbb{P}\left[\|\hat{\mathbf{R}}_{ij} - \mathbf{R}_{ij}\|_2 \geq \kappa\delta^{\frac{\beta}{1+\beta}}\right] \leq \frac{1}{f(n,\delta)}$$

for $0 < \delta \leq \delta_*$, where $\kappa = 16\sqrt{2}\,([1 \vee \max_{i\neq j}\|\mathbf{R}_{ij}\|_2] \vee [\max_{i\neq j}\{\|\Phi_{ij}\|_2\}[2 \vee 2\max_j\{\|\mathbf{C}_{jj}\|\}]])$. Consequently, $\|\hat{\mathbf{R}}_{ij} - \mathbf{R}_{ij}\|_2 \leq \kappa\bar{\delta}_f(n,r)^{\frac{\beta}{1+\beta}}$ with probability at least $1 - 1/r$ when $\epsilon_n = \bar{\delta}_f(n,r)^{\frac{1}{1+\beta}}$.

In other words, if $\hat{\mathbf{C}}$ satisfies tail conditions with the parameters f and δ_*, then $\hat{\mathbf{R}}$ satisfies a tail condition with the parameters $g(n,\delta) = f(n,[\delta/\kappa]^{1+1/\beta})$ and $[\delta_*/\kappa]^{1+1/\beta}$. Note that

$$\bar{\delta}_g(n,r) = \kappa\bar{\delta}_f(n,r)^{\frac{\beta}{1+\beta}} \quad \text{and} \quad \bar{n}_g(\delta,r) = \bar{n}_f([\delta/\kappa]^{1+1/\beta},r).$$

Theorem 4.4 tells us how well we can estimate the correlation operator $\mathbf{R}$, given an estimator $\hat{\mathbf{C}}$ of the covariance operator $\mathbf{C}$. The performance depends crucially on the regularity β, with smaller values of β requiring higher sample sizes $n = \bar{n}_g(\delta,r)$ to estimate $\mathbf{R}_{ij}$ up to the same error with high probability.

Estimation of the Precision Operator

We begin by describing the entry-wise Hilbert–Schmidt deviation of the estimator $\hat{\mathbf{H}}$. In the following results, the parameter τ is user-defined and can be increased to get better concentration of $\hat{\mathbf{H}}$ near $\mathbf{H}^*$ in exchange for more demanding requirements on the sample size n.

Theorem 4.5. *Let $X = (X_1,\ldots,X_p)$ be a second-order random element in the Hilbert space $\mathcal{H}$ with the covariance $\mathbf{C}$ and let $\hat{\mathbf{C}}$ be an estimator of $\mathbf{C}$ satisfying the tail condition with parameters f and $\delta_* > 0$, and let $\tau > 2$. Under Assumptions 3/1*, 4 and 5, and conditions for Theorem 4.4, if $\epsilon_n = \bar{\delta}_f(n,p^\tau)^{\frac{1}{1+\beta}}$, $\lambda_n = \frac{8}{\alpha}\kappa\bar{\delta}_f(n,p^\tau)^{\frac{\beta}{1+\beta}}$ and the sample size n satisfies*

$$n \geq \bar{n}_f\left(1/\max\left\{\frac{1}{\delta_*}, \left[12d\kappa\left(1+\frac{8}{\alpha}\right)^2\left[\rho\gamma \vee \rho^3\gamma^2\right]\right]^{1+\frac{1}{\beta}}\right\}, p^\tau\right), \tag{4.18}$$

then with probability at least $1 - 1/p^{\tau-2}$, we have:

1. *The estimate* $\hat{\mathbf{H}}$ *satisfies*

$$\|\hat{\mathbf{H}} - \mathbf{H}^*\|_{2,\infty} \leq 2\gamma \left(1 + \frac{8}{\alpha}\right) \kappa \bar{\delta}_f(n, p^\tau)^{\frac{\beta}{1+\beta}}. \tag{4.19}$$

2. *If for some* (i, j), $\|\mathbf{H}^*_{ij}\|_2 > 2\gamma(1 + 8/\alpha)\kappa\bar{\delta}_f(n, p^\tau)^{\frac{\beta}{1+\beta}}$, *then* $\hat{\mathbf{H}}_{ij}$ *is nonzero.*

Notice how the sample size requirement (4.18) depends on the incoherence parameter α, the degree d of the graph, and the parameters ρ and γ in essentially the same way as Theorem 1 of Ravikumar et al. (2011), except for the coefficient κ and the power $1 + 1/\beta$. As the regularity $\beta \to 0$, the sample size requirement increases while the bound (4.19) on the entry-wise deviation $\|\hat{\mathbf{H}} - \mathbf{H}^*\|_{2,\infty}$ weaken. The parameters κ, ρ and γ can be said to capture the sizes of involved quantities while the degree d describes the sparsity of the graph. The factor $(1 + 8/\alpha)$ quantifies the dependence of our sample size requirement and bound on entry-wise deviation on Assumption 4 and decreasing the incoherence α unsurprisingly weakens the bound. Interestingly, the dependence on the degree d of the graph is through the regularity β.

The bound on the maximum deviation of an entry of $\hat{\mathbf{H}}$ from the corresponding entry of $\mathbf{H}^*$ in Hilbert–Schmidt norm, naturally yields a bound on the deviation of $\hat{\mathbf{H}}$ itself from $\mathbf{H}^*$ in Hilbert–Schmidt norm.

Corollary 4.1. *Let* s *denote the total number of nonzero off-diagonal entries in* $\mathbf{H}^*$. *Under the same conditions and choices of* ϵ_n, λ_n *and* n *as Theorem 4.5, we have*

$$\|\hat{\mathbf{H}} - \mathbf{H}^*\|_2 \leq 2\gamma \left(1 + \frac{8}{\alpha}\right) \kappa \sqrt{p + s} \bar{\delta}_f(n, p^\tau)^{\frac{\beta}{1+\beta}}.$$

with probability at least $1 - 1/p^{\tau-2}$.

Proof. Clearly, $\|\hat{\mathbf{H}} - \mathbf{H}^*\|_2^2 = \sum_{i=1}^p \|\hat{\mathbf{H}}_{ii} - \mathbf{H}^*_{ii}\|_2^2 + \sum_{i \neq j} \|\hat{\mathbf{H}}_{ij} - \mathbf{H}^*_{ij}\|_2^2 \leq (p + s)\|\hat{\mathbf{H}} - \mathbf{H}^*\|_{2,\infty}^2$. □

Model Selection Consistency

If we increase the sample size enough, we can ensure that we recover the whole graph exactly with high probability. Specifically, the sample size has to high enough to ensure that the $\hat{\mathbf{H}}_{ij}$ is nonzero for the smallest $\|\mathbf{H}^*_{ij}\|_2$ with high probability. Naturally, a smaller θ leads to a more stringent sample size requirement.

Corollary 4.2 (Model Selection Consistency)**.** *Let* $\theta = \min\{\|\mathbf{H}^*_{ij}\|_2 : \mathbf{H}^*_{ij} \neq \mathbf{0}\}$ *and* $\tau > 2$. *Under the same conditions as Theorem 4.5, if* $\epsilon_n = \bar{\delta}_f(n, p^\tau)^{\frac{1}{1+\beta}}$, $\lambda_n = \frac{8}{\alpha}\kappa\bar{\delta}_f(n, p^\tau)^{\frac{\beta}{1+\beta}}$

and the sample size n satisfies

$$n \geq \bar{n}_f \left(1/\max \left\{ \frac{1}{\delta_*}, \left[\frac{2\gamma\kappa}{\theta} \left(1 + \frac{8}{\alpha} \right) \right]^{1+\frac{1}{\beta}}, \left[12d\kappa \left(1 + \frac{8}{\alpha} \right)^2 \left[\rho\gamma \vee \rho^3\gamma^2 \right] \right]^{1+\frac{1}{\beta}} \right\}, p^\tau \right)$$

then $\mathbb{P}[\hat{G} = G] \geq 1 - 1/p^{\tau-2}$.

Proof. In addition to the sample size requirement of Theorem 4.5, we require that $\theta > 2\gamma(1 + 8/\alpha)\kappa\bar{\delta}_f(n, p^\tau)^{\frac{\beta}{1+\beta}}$. By Theorem 4.5 (2), this ensures that the entry $\hat{\mathbf{H}}_{ij}$ is nonzero for every nonzero entry $\mathbf{H}^*_{ij}$ thus implying exact recovery of the graph G. □

4.6.3 Sub-Gaussian Random Elements

In this section, we shall work out the finite sample theory for sub-Gaussian random elements $X = (X_1, \ldots, X_p)$, if the covariance $\mathbf{C}$ is estimated using the empirical covariance operator $\hat{\mathbf{C}} = \frac{1}{n}\sum_{k=1}^n X^k \otimes X^k - \bar{X} \otimes \bar{X}$ where $\bar{X} = \frac{1}{n}\sum_{k=1}^n X^k$.

There are many definitions of sub-Gaussianity for random elements in Hilbert spaces (cf. Chen and Yang (2021), Antonini (1997)). For our purpose, the sub-Gaussianity of the norms of the constituent random elements provides a natural generalization of the definition used in Ravikumar et al. (2011) which required the coordinates to be sub-Gaussian random variables.

Definition 4.2. *We shall say that $X = (X_1, \ldots, X_p)$ is a sub-Gaussian random element in the product space $\mathcal{H}$ if the norms $\|X_j\|$ are sub-Gaussian random variables for $1 \leq j \leq p$. Equivalently, X is a sub-Gaussian random element if*

$$\|X\|_\infty = \max_j \|\|X_j\|\|_{\psi_2} < \infty.$$

The above definition is weaker than an alternative definition proposed by Chen and Yang (2021) but stronger than the one suggested by Vershynin (2018). Using the Karhunen-Loève expansion, it can be shown that for a Gaussian random element $X = (X_1, \ldots, X_p)$, the sub-Gaussian norms of X_j satisfy

$$\|\|X_j\|\|^2_{\psi_2} = \|\|X_j\|^2\|_{\psi_1} \leq \tfrac{8}{3}\operatorname{tr}(\mathbf{C}_{jj})$$

and therefore, $\|X\|_\infty \leq \sqrt{8/3}\max_j [\operatorname{tr}(\mathbf{C}_{jj})]^{1/2} \leq \sqrt{8/3}\,[\operatorname{tr}(\mathbf{C})]^{1/2}$. Thus our definition includes all Gaussian random elements X as sub-Gaussian.

Using Bernstein's inequality (cf. Theorem 2.8.1 of Vershynin (2018)), we can show that

Lemma 4.3. *If $\hat{\mathbf{C}} = \frac{1}{n}\sum_{k=1}^n X^k \otimes X^k - \bar{X} \otimes \bar{X}$ is the empirical covariance estimator,*

then for $0 < \delta \leq \delta_*$, *we have*

$$\mathbb{P}\{\|\hat{\mathbf{C}}_{ij} - \mathbf{C}_{ij}\|_2 \geq \delta\} \leq 2\exp\left[-\frac{cn\delta^2}{\|X\|_\infty^4}\right]$$

where

$$\delta_* = \min_{ij} \| \|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] - \bar{X}_i \otimes \bar{X}_j + \mathbb{E}[X_i] \otimes \mathbb{E}[X_j] \| \|_{\psi_1}.$$

If $\mathbb{E}[X] = \mathbf{0}$, *the statement continues to hold even for* $\hat{\mathbf{C}} = \frac{1}{n}\sum_{k=1}^n X^k \otimes X^k$ *and* $\delta_* = \min_{ij} \| \|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] \| \|_{\psi_1}$.

If the mean is zero, which is the case addressed in Ravikumar et al. (2011), then δ_* simply does not depend on n. If the mean $\mathbb{E}[X]$ is not zero, the dependence of the quantity δ_* on n is not as pronounced as it may seem because for large n, $\delta_* \approx \min_{ij} \| \|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] \| \|_{\psi_1}$ which can be shown to be always greater than or equal to $\min_{ij} \| \|X_i\| \|X_j\| - \mathbb{E}\left[\|X_i\| \|X_j\|\right] \|_{\psi_1}$ by Jensen's inequality.

Now, Lemma 4.3 essentially says that $\hat{\mathbf{C}}$ satisfies the tail condition for $f(n, \delta) = \frac{1}{2}\exp\left[\frac{cn\delta^2}{\|X\|_\infty^4}\right]$ and $\delta_* > 0$, which implies that

$$\bar{n}_f(\delta, r) = \left\lfloor \frac{\|X\|_\infty^4 \log(2r)}{c\delta^2} \right\rfloor \quad \text{and} \quad \bar{\delta}_f(n, r) = \sqrt{\frac{\|X\|_\infty^4 \log(2r)}{cn}}.$$

Applying Theorem 4.5 to our special case now yields explicit parameter choices, sample size requirements and upper bounds on the entry-wise deviations.

Theorem 4.6 (Sub-Gaussian Random Elements)**.** *Assume that* $X = (X_1, \ldots, X_p)$ *is such that the norms* $\|X_j\|$ *are sub-Gaussian and let* $\hat{\mathbf{C}} = \frac{1}{n}\sum_{k=1}^n X^k \otimes X^k - \bar{X} \otimes \bar{X}$. *Under the same conditions as Theorem 4.5, if the parameters* ϵ_n *and* λ_n *are chosen as*

$$\epsilon_n = \left[\frac{\|X\|_\infty^4(\log 2 + \tau \log p)}{cn}\right]^{\frac{1}{2(1+\beta)}}, \quad \lambda_n = \frac{8}{\alpha}\kappa\left[\frac{\|X\|_\infty^4(\log 2 + \tau \log p)}{cn}\right]^{\frac{\beta}{2(1+\beta)}}$$

and the sample size n *satisfies*

$$n > [\log 2 + \tau \log p] \max\left\{\frac{1}{\delta_*^2}, \left[12d\kappa\left(1 + \frac{8}{\alpha}\right)^2\left[\rho\gamma \vee \rho^3\gamma^2\right]\right]^{2+\frac{2}{\beta}}\right\}\frac{\|X\|_\infty^4}{c}$$

we have with probability at least $1 - 1/p^{\tau-2}$ *that*

$$\|\hat{\mathbf{H}} - \mathbf{H}^*\|_{2,\infty} \leq 2\gamma\left(1 + \frac{8}{\alpha}\right)\kappa\|X\|_\infty^2\left[\frac{\log 2 + \tau \log p}{cn}\right]^{\frac{\beta}{2(1+\beta)}}$$

and $\|\hat{\mathbf{H}} - \mathbf{H}^*\|_2 \leq \sqrt{s+p}\|\hat{\mathbf{H}} - \mathbf{H}^*\|_{2,\infty}$ *where* s *denotes the number of nonzero off-diagonal entries of* $\mathbf{H}$. *Here* δ_* *is as in Lemma 4.3.*

Assuming the parameters κ, ρ, γ and α do not change very much with respect to p, this suggests that

$$\|\hat{\mathbf{H}} - \mathbf{H}^*\|_2 = \mathcal{O}\left(\sqrt{s+p}\left[\frac{\log p}{n}\right]^{\frac{\beta}{2(1+\beta)}}\right).$$

It must be noted that even under the most favourable regularity conditions, which is when $\beta = 1$, we cannot recover the bound for the multivariate case, which is $\mathcal{O}(\sqrt{(s+p)(\log p)/n})$. It appears that this an enduring consequence of having to use correlation operator matrices, which is necessary in the functional setting because covariance operators (which are compact) do not admit bounded inverses, but optional in the multivariate setting where the inverse of a full-rank covariance matrix is always bounded. Fortunately, the sample size requirement is still reasonable in that it only requires

$$n = \Omega((\delta_*^{-2} + d^{2+2/\beta})\tau \log p),$$

which implies that estimation with a relatively small sample size is still feasible so long as $d \ll p$ and $s \ll p^2$.

Corollary 4.3 (Model Selection Consistency for Sub-Gaussians)**.** *Let* $\theta = \min\{\|\mathbf{H}_{ij}^*\|_2 : \mathbf{H}_{ij}^* \neq \mathbf{0}\}$. *Under the same conditions and parameter choices of* ϵ_n *and* λ_n *as in Theorem 4.6, if the sample size* n *satisfies,*

$$n > [\log 2 + \tau \log p] \max\left\{\frac{1}{\delta_*^2}, \left[\frac{2\gamma\kappa}{\theta}\left(1+\frac{8}{\alpha}\right)\right]^{2+\frac{2}{\beta}}, \left[12d\kappa\left(1+\frac{8}{\alpha}\right)^2\left[\rho\gamma \vee \rho^3\gamma^2\right]\right]^{2+\frac{2}{\beta}}\right\}\frac{\|X\|_\infty^4}{c}$$

we have $\mathbb{P}[\hat{G} = G] \geq 1 - 1/p^{\tau-2}$.

The sample size requirement for model selection consistency is thus higher. In Big-Ω notation, we need

$$n = \Omega((\delta_*^{-2} + \theta^{-2-2/\beta} + d^{2+2/\beta})\tau \log p)$$

samples to recover the true graph with at least $1 - p^{\tau-2}$ probability. The dependence on θ and d is mediated by the regularity β.

Remark 4.3. *Our theoretical analysis also gives insight into the multivariate case by showing that a partial recovery might still be feasible even when the graph is not sparsely connected so long as it can be partitioned into sparsely connected subgraphs. Given a random vector* $Y = \{Y_j\}_{j=1}^P$ *for some* $P \gg 1$*, this would correspond to partitioning* Y *into sparsely related smaller random* subvectors $X_j \subset Y$ *while the entries of* Y *withing an individual random vector* X_j *are allowed to be densely related. This ensures that the maximum degree* d *for the graph of* $\{X_j\}_{j=1}^p$ *is small even if that of* $\{Y_j\}_{j=1}^P$ *is not, thus creating the sufficient conditions for our finite sample arguments to work. We would expect the standard multivariate graphical lasso to fail here because it does not leverage the latent sparsity in the form of sparsely related subvectors of* Y*.*

4.7 Implementation

4.7.1 Discretization

The operator formalism used in (4.11) has so far allowed us to put off the delicate issue of how the involved quantities are discretized for the purpose of computation. It turns out that this is mostly a question of finding the correct discrete equivalents of the objects and operations involved. As in Section 4.5.3, we shall denote the discrete counterparts of functional quantities (eg. X or $\mathbf{C}$) using the sans serif font (eg. X or C).

We shall discretize every element $\mathbf{f} = (f_1, \ldots, f_p)$ in $\mathcal{H}$ as a column vector f of length K indexed by $\mathsf{J} = [1, \ldots, \mathsf{K}]$. The coordinates f_i in $\mathcal{H}_i$ shall be discretized as subvectors $\mathsf{f_i} = [\mathsf{f_{ij}} : \mathsf{j} \in \mathsf{J_i}]$ of length $\mathsf{K_i}$ indexed by $\mathsf{J_i} \subset \mathsf{J}$ such that J is the concatenation of the sets $\mathsf{J_i}$. There are many different ways of doing this. For example, if $\mathcal{H}_i$ is the space $L^2(U_i, \mu_i)$ of square-integrable functions on some space U_i equipped with the measure μ_i, we can generate a mesh $\{U_{ij}\}$ of $\mathsf{K_i}$ cells of roughly equal measure on U_i and take $\mathsf{f_{ij}}$ to be the average value $\frac{1}{\mu_i(U_{ij})} \int_{U_{ij}} f_i$ of f_i in the jth cell. Often, $\mathcal{H}_i$ is composed of continuous functions on U_i and we can take $\mathsf{f_{ij}}$ to be the value $f_i(u_{ij})$ at some fixed point $u_{ij} \in U_{ij}$. These discretizations schemes can be described as *discretization by cell averaging* and *discretization by point evaluation* respectively (c.f. Masak and Panaretos, 2022).

Another recourse is to take $\mathsf{f_{ij}}$ to be the jth coefficient $\langle f_i, e_j \rangle$ in the basis expansion of f_i with respect to a fixed basis $\{e_j\}_{j=1}^{\infty}$ on $\mathcal{H}_i$. This is *discretization by basis representation*. An element $\mathbf{f} \in \mathcal{H}$ can thus be represented in terms of the tensor product basis formed from bases on the spaces $\mathcal{H}_j$. There are plenty of different ways of doing this. One can use pre-specified bases such as B-splines or empirical bases corresponding to Karhunen-Loève type expansions, be it a one-dimensional expansion in every node like in Qiao et al. (2019), a two-dimensional expansion under additional structural assumptions like in Zapata et al. (2022), or any other version of multivariate functional PCA (Chiou et al., 2014).

Let $\mathbf{f}, \mathbf{g}, \mathbf{h} \in \mathcal{H}$ where $\mathbf{f} = (f_1, \ldots, f_p)$ and $\mathbf{g} = (g_1, \ldots, g_p)$, with the discretizations f, g and h. The tensor or outer product $\mathbf{f} \otimes \mathbf{g}$ of $\mathbf{f}, \mathbf{g} \in \mathcal{H}$ is to be discretized simply as the matrix $\mathsf{f}\mathsf{g}^\top$. The inner product $\langle \mathbf{f}, \mathbf{g} \rangle$, however, is to be discretized as $\mathsf{f}^\top \mathsf{M} \mathsf{g}$, where the $\mathsf{K} \times \mathsf{K}$ matrix M is the discrete equivalent of the inner product operation, which actually depends on the scheme of discretization employed. If we are averaging on cells or using point evaluations as discussed before, M is given by $\mathsf{M_{ij}} = 1/\mathsf{K_l}$ if both $\mathsf{i} = \mathsf{j} \in \mathsf{J_l}$ and is 0 otherwise. On the other hand, if we are using a basis representation, M is same as the $\mathsf{K} \times \mathsf{K}$ identity matrix $\mathsf{I_K}$. This difference follows from the observation that for $f_i, g_i \in \mathcal{H}_i$ and their discretizations $\mathsf{f_i}, \mathsf{g_i}$, we have under the former schemes of observation

$$\langle f_i, g_i \rangle = \int_{U_i} f_i(u) g_i(u) d\mu_i(u) \approx \frac{1}{\mathsf{K_i}} \sum_{\mathsf{j}=1}^{\mathsf{K_i}} \mathsf{f_{ij}} \mathsf{g_{ij}}$$

while under the latter scheme, we have

$$\langle f_i, g_i\rangle = \sum_{j=1}^{\infty}\langle f_i, e_j\rangle\langle g_i, e_j\rangle \approx \sum_{\mathsf{j}=1}^{\mathsf{K_i}} \mathsf{f_{ij}g_{ij}}$$

instead. The correct way to represent $(\mathbf{f}\otimes\mathbf{g})\mathbf{h} = \langle\mathbf{g},\mathbf{h}\rangle\mathbf{f}$ is thus $(\mathsf{g}^\top\mathsf{Mh})\mathsf{f} = (\mathsf{fg}^\top)\mathsf{Mh}$.

Because compact operators are infinite sums of tensor products of elements, we can discretize them in essentially the same way as elements themselves. Thus the discretization of a Hilbert–Schmidt operator matrix $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1}^{p}$ is a $\mathsf{K}\times\mathsf{K}$ matrix A, with the entries of $\mathbf{A}$ being represented by the submatrices of A in the same way as with the element $\mathbf{f}$ and its coordinates. And just like the outer products, we can represent $\mathbf{Ah}$ as AMh. The same applies to operator-operator multiplication and the correct representation of the product $\mathbf{AB}$ is AMB, where $\mathbf{B}$ is another Hilbert–Schmidt operator with the discretization B.

Non-compact operators such has $\mathbf{I}$ on the other hand, cannot be discretized like compact operators. For such operations, it is best to find the discrete equivalent of their action on the elements directly. For $\mathbf{I}$, notice that $\mathbf{If} = \mathbf{f}$. What linear operation when applied to f would return f? Of course, that's the $\mathsf{K}\times\mathsf{K}$ identity matrix $\mathsf{I_K}$. So the correct way to represent the operation $\mathbf{If}$ is $\mathsf{I_K f}$. Trivial as it may appear, understanding this is what allows us to arrive at the correct representation of the operation $(\mathbf{I}+\mathbf{A})^{-1/2}\mathbf{f}$ which is $(\mathsf{I_K}+\mathsf{AM})^{-1/2}\mathsf{f}$, as can be inferred from the binomial expansion of $(\mathbf{I}+\mathbf{A})^{-1/2}$.

Using the same principle, we can work out that the trace $\operatorname{tr}(\mathbf{A})$ and Carleman-Fredholm determinant $\det_2(\mathbf{I}+\mathbf{A})$ of $\mathbf{A}$ can be represented as

$$\mathsf{tr}[\mathsf{MA}] \qquad \text{and} \qquad \mathsf{det}[\mathsf{I_K}+\mathsf{MA}]\cdot\mathsf{exp}\left(-\mathsf{tr}[\mathsf{MA}]\right),$$

respectively. Note also that the action of taking the diagonal part $\operatorname{dg}(\mathbf{A})$ of $\mathbf{A}$ is equivalent to taking the Hadamard product $\mathsf{D}\circ\mathsf{A}$ with matrix $\mathsf{D} = [\mathsf{D_{ij}}]$ where $\mathsf{D_{ij}} = 1$ if both $\mathsf{i},\mathsf{j}\in\mathsf{J_l}$ for some $1\le\mathsf{l}\le\mathsf{p}$ and is 0 otherwise. We are now going to describe the discretized version of our algorithm.

Given n realizations of X in the form of vectors $\{\mathsf{X_k} : \mathsf{k} = 1,\ldots,\mathsf{n}\}$ we compute the discretized version C of the estimated covariance $\hat{\mathbf{C}}$. For example, if $\hat{\mathbf{C}}$ is the empirical covariance estimator, we get:

$$\mathsf{C} = \tfrac{1}{\mathsf{n}}\sum_{\mathsf{k}=1}^{\mathsf{n}}\mathsf{X_k}\mathsf{X_k}^\top - \left[\tfrac{1}{\mathsf{n}}\sum_{\mathsf{k}=1}^{\mathsf{n}}\mathsf{X_k}\right]\left[\tfrac{1}{\mathsf{n}}\sum_{\mathsf{k}=1}^{\mathsf{n}}\mathsf{X_k}\right]^\top$$

The off-diagonal part $\hat{\mathbf{C}}_0 = \hat{\mathbf{C}} - \operatorname{dg}\hat{\mathbf{C}}$ is discretized as $\mathsf{C}-\mathsf{D}\circ\mathsf{C}$. Since $\operatorname{dg}\hat{\mathbf{C}}\,\mathbf{f}$ is given by $(\mathsf{D}\circ\mathsf{CM})\mathsf{f}$, the estimated cross-correlation operator matrix $\hat{\mathbf{R}}_0$ (which is compact) thus corresponds to $\mathsf{R_0}$ given by

$$\mathsf{R_0} = \left[\epsilon_n\mathsf{I_K}+\mathsf{D}\circ\mathsf{CM}\right]^{-1/2}\cdot\left[\mathsf{C}-\mathsf{D}\circ\mathsf{C}\right]\cdot\left[\epsilon_n\mathsf{I_K}+\mathsf{D}\circ\mathsf{MC}\right]^{-1/2}.$$

The operator trace $\mathrm{tr}(\mathbf{H}\hat{\mathbf{R}}_0)$ and the Carleman-Fredholm determinant $\det_2(\mathbf{I}+\mathbf{H})$ can be evaluated in terms of the matrix trace tr and determinant det as

$$\mathsf{tr}[\mathsf{MHMR}_0] \qquad \text{and} \qquad \mathsf{det}[\mathsf{I}_\mathsf{K}+\mathsf{MH}]\cdot\exp\left(-\mathsf{tr}[\mathsf{MH}]\right),$$

respectively. Finally, recall that the discretization of $\mathbf{A}=[\mathbf{A}_{ij}]_{i,j=1}^{p}$ is defined as a $\mathsf{K}\times\mathsf{K}$ matrix $\mathsf{A}=[\mathsf{A}_{\mathsf{ij}}]_{\mathsf{i,j}=1}^{\mathsf{K}}$ where each of the operators $\mathbf{A}_{ij}$ is discretized as $\mathsf{A}[\mathsf{J_i},\mathsf{J_j}]=[\mathsf{A}_{\mathsf{kl}}]_{\mathsf{k}\in\mathsf{J_i},\mathsf{l}\in\mathsf{J_j}}$. Because $\|\mathbf{A}\|_{2,1}=\sum_{i,j=1}^{p}\|\mathbf{A}_{ij}\|_2=\sum_{i,j=1}^{p}[\mathrm{tr}(\mathbf{A}_{ij}\mathbf{A}_{ij}^*)]^{1/2}$, the discretized counterpart is given by

$$\begin{aligned}
&\sum_{\mathsf{i,j}=1}^{\mathsf{p}}\mathsf{tr}\left[\mathsf{M}[\mathsf{J_i},\mathsf{J_i}]\mathsf{A}[\mathsf{J_i},\mathsf{J_j}]\mathsf{M}[\mathsf{J_j},\mathsf{J_j}]\mathsf{A}[\mathsf{J_i},\mathsf{J_j}]^\top\right]\\
=&\sum_{\mathsf{i,j}=1}^{\mathsf{p}}\mathsf{tr}\left[(\mathsf{M}[\mathsf{J_i},\mathsf{J_i}]^{1/2}\mathsf{A}[\mathsf{J_i},\mathsf{J_j}]\mathsf{M}[\mathsf{J_j},\mathsf{J_j}]^{1/2})(\mathsf{M}[\mathsf{J_i},\mathsf{J_i}]^{1/2}\mathsf{A}[\mathsf{J_i},\mathsf{J_j}]\mathsf{M}[\mathsf{J_j},\mathsf{J_j}]^{1/2})^\top\right]\\
=&\|\mathsf{M}^{1/2}\mathsf{AM}^{1/2}\|_{2,1}
\end{aligned}$$

where the norm $\|\cdot\|_{2,1}$ is defined as $\|\mathsf{A}\|_{2,1}=\sum_{\mathsf{i,j}=1}^{\mathsf{p}}\|\mathsf{A}[\mathsf{J_i},\mathsf{J_j}]\|_\mathsf{F}$. Altogether, the optimization functional $\mathcal{F}$ can thus be written as

$$\mathsf{F}[\mathsf{H}] = \mathsf{tr}[\mathsf{MHMR}_0]+\mathsf{tr}[\mathsf{MH}]-\log\,\mathsf{det}[\mathsf{I}_\mathsf{K}+\mathsf{MH}]+\lambda_n\cdot\|\mathsf{M}^{1/2}(\mathsf{H}-\mathsf{D}\circ\mathsf{H})\mathsf{M}^{1/2}\|_{2,1}.$$

Now, using the cyclic property of the trace and multiplicativity of the determinant, we can write

$$\begin{aligned}
\mathsf{tr}[\mathsf{MHMR}_0]+\mathsf{tr}[\mathsf{MH}] &= \mathsf{tr}[(\mathsf{M}^{1/2}\mathsf{HM}^{1/2})(\mathsf{M}^{1/2}\mathsf{R}_0\mathsf{M}^{1/2})]+\mathsf{tr}[\mathsf{M}^{1/2}\mathsf{HM}^{1/2}]\\
&= \mathsf{tr}[(\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{HM}^{1/2})(\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{R}_0\mathsf{M}^{1/2})]-\mathsf{tr}[\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{R}_0\mathsf{M}^{1/2}]\\
\log\,\mathsf{det}[\mathsf{I}_\mathsf{K}+\mathsf{MH}] &= \log\,\mathsf{det}[\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{HM}^{1/2}]\\
\|\mathsf{M}^{1/2}(\mathsf{H}-\mathsf{D}\circ\mathsf{H})\mathsf{M}^{1/2}\|_{2,1} &= \|(\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{HM}^{1/2})-\mathsf{D}\circ(\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{HM}^{1/2})\|_{2,1}
\end{aligned}$$

Ignoring the constant term $-\mathsf{tr}[\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{R}_0\mathsf{M}^{1/2}]$ in the second equation, the problem reduces to minimizing

$$\tilde{\mathsf{F}}(\mathsf{Q})=\mathsf{tr}[\mathsf{QR}]-\log\,\mathsf{det}[\mathsf{Q}]+\lambda_n\cdot\|\mathsf{Q}-\mathsf{D}\circ\mathsf{Q}\|_{2,1} \tag{4.20}$$

with respect to Q under the constraint $\mathsf{Q}>0$, where

$$\mathsf{Q}=\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{HM}^{1/2} \text{ and } \mathsf{R}=\mathsf{I}_\mathsf{K}+\mathsf{M}^{1/2}\mathsf{R}_0\mathsf{M}^{1/2}.$$

Note that the matrix determinant should be evaluated directly as the product of the eigenvalues of the matrix calculated using the eigendecomposition rather than using cofactor expansion. The former is vastly superior in terms of computational efficiency and numerical precision.

The operator formalism used in (4.11) faithfully encapsulates the three different discretization techniques (averaging on cells, evaluation at points, or an orthonormal basis representation) discussed above, in a *coordinate-free* way (Stone, 1987). As a consequence of this faithful representation, the above formulas are comparable across (high enough) resolutions. Their value does not change drastically if one increases the number of points, cells or basis functions without bound and in fact tends to the exact values of the corresponding quantities as the number of samples increases.

4.7.2 Optimization

We use the alternating direction method of multipliers (ADMM, Boyd et al., 2011) to solve the convex optimization problem 4.20. The basic idea of ADMM is to introduce an auxiliary variable Z to separate the loss or likelihood term and from the penalty term,

$$\arg\min_{\mathsf{Q},\mathsf{Z}} \mathsf{tr}[\mathsf{QR}] - \mathsf{log\ det}[\mathsf{Q}] + \lambda_n \cdot \sum_{\mathsf{i}\neq\mathsf{j}} \left[\sum_{\mathsf{u}\in\mathsf{I_i}}\sum_{\mathsf{v}\in\mathsf{I_j}} \mathsf{Z}_{\mathsf{uv}}^2\right]^{1/2} \quad \text{s.t.} \quad \mathsf{Q} = \mathsf{Z}$$

The augmented Lagrangian can then be written as

$$\arg\min_{\mathsf{Q},\mathsf{Z}} \mathsf{tr}[\mathsf{QR}] - \mathsf{log\ det}[\mathsf{Q}] + \lambda_n \cdot \sum_{\mathsf{i}\neq\mathsf{j}} \left[\sum_{\mathsf{u}\in\mathsf{I_i}}\sum_{\mathsf{v}\in\mathsf{I_j}} \mathsf{Z}_{\mathsf{uv}}^2\right]^{1/2} + \frac{\rho}{2}\|\mathsf{Q}-\mathsf{Z}\|_{\mathsf{F}}^2 + \langle \mathsf{Y}, \mathsf{Q}-\mathsf{Z}\rangle$$

and subsequently minimized w.r.t. Q and Z in an alternating fashion, with the dual variable Y updated after every iteration. In the above, ρ is a small positive constant affecting the convergence speed, not the convergence itself, which is guaranteed irrespective of the choice (Boyd et al., 2011). We use the default $\rho = 1$ in our applications of the algorithm. It is customary to perform another variable change: $\mathsf{U} := \mathsf{Y}/\rho$. The augmented Lagrangian then becomes

$$\mathsf{L}_\rho(\mathsf{Q},\mathsf{Z},\mathsf{U}) = \mathsf{tr}[\mathsf{QR}] - \mathsf{log\ det\ Q} + \lambda_n \cdot \textstyle\sum_{\mathsf{i}\neq\mathsf{j}} \left[\sum_{\mathsf{u}\in\mathsf{I_i}}\sum_{\mathsf{v}\in\mathsf{I_j}} \mathsf{Z}_{\mathsf{uv}}^2\right]^{1/2} + \frac{\rho}{2}\|\mathsf{Q}-\mathsf{Z}+\mathsf{U}\|_{\mathsf{F}}^2,$$

which is equal to the one above up to a constant, and hence the optimal H can be obtained easily from the optimal Q, indeed providing a solution to the original problem 4.20. Overall, the l-th iteration of the ADMM algorithm consists of the following three steps, iterated until confergence for $\mathsf{m} = 1, 2, \ldots$ starting from an initial point $\mathsf{Z}^{(0)}, \mathsf{U}^{(0)}$:

$$\begin{aligned}
\mathsf{Q}^{(\mathsf{m})} &:= \arg\min_{\mathsf{Q}} \mathsf{L}_\rho(\mathsf{Q}, \mathsf{Z}^{(\mathsf{m}-1)}, \mathsf{U}^{(\mathsf{m}-1)}) \\
\mathsf{Z}^{(\mathsf{m})} &:= \arg\min_{\mathsf{Z}} \mathsf{L}_\rho(\mathsf{Q}^{(\mathsf{m})}, \mathsf{Z}, \mathsf{U}^{(\mathsf{m}-1)}) \\
\mathsf{U}^{(\mathsf{m})} &:= \mathsf{U}^{(\mathsf{m}-1)} + (\mathsf{Q}^{(\mathsf{m})} - \mathsf{Z}^{(\mathsf{m})})
\end{aligned}$$

The first step has an analytic solution. Equating the derivative of $\mathsf{L}_\rho(\mathsf{Q}, \mathsf{Z}^{(\mathsf{m}-1)}, \mathsf{U}^{(\mathsf{m}-1)})$ w.r.t. Q to zero, one obtains the following non-linear system:

$$\rho\mathsf{Q} - \mathsf{Q}^{-1} = \rho(\mathsf{Z}^{(\mathsf{m}-1)} - \mathsf{U}^{(\mathsf{m}-1)}) - \mathsf{R}.$$

Denoting by $\mathsf{E}^{(\mathsf{m}-1)}\Gamma^{(\mathsf{m}-1)}(\mathsf{E}^{(\mathsf{m}-1)})^\top$ the eigendecomposition on the right-hand side and changing the variable to $\widetilde{\mathsf{Q}} = (\mathsf{E}^{(\mathsf{m}-1)})^\top\mathsf{Q}\mathsf{E}^{(\mathsf{m}-1)}$, the non-linear system becomes

$$\rho\widetilde{\mathsf{Q}} - \widetilde{\mathsf{Q}}^{-1} = \Gamma^{(\mathsf{m}-1)}.$$

Note that $\widetilde{\mathsf{Q}}$ and $\widetilde{\mathsf{Q}}^{-1}$ have the same eigenvectors, and $\Gamma^{(\mathsf{l}-1)}$ is diagonal, i.e. the eigenvectors are forming the canonical basis of $\mathbb{R}^K$. Hence the solution is given by matching the eigenvalues only: for $i = 1, \ldots, K$ it is sufficient to have $\rho\widetilde{q}_{ii} - 1/\widetilde{q}_{ii} = \gamma_{ii}^{(l-1)}$. These quadratic equations are solved, respectively, by

$$\widetilde{\mathsf{q}}_{ii}^{(\mathsf{m})} = \frac{\gamma_{ii}^{(\mathsf{m}-1)} + \sqrt{\left[\gamma_{ii}^{(\mathsf{m}-1)}\right]^2 + 4\rho}}{(2\rho)}.$$

With these forming the diagonal of $\widetilde{\mathsf{Q}}^{(\mathsf{m})}$, we obtain $\mathsf{Q}^{(\mathsf{m})} = \mathsf{E}^{(\mathsf{m}-1)}\widetilde{\mathsf{Q}}^{(\mathsf{m})}(\mathsf{E}^{(\mathsf{m}-1)})^\top$. Note that we chose the negative sign above to obtain a positive semi-definite solution, which is naturally the one sought even though we do not make this constraint explicit.

In the second step, the problem separates in variables $\mathsf{Z}_{\mathsf{i},\mathsf{j}} := \mathsf{Z}[\mathsf{I}_\mathsf{i}, \mathsf{I}_\mathsf{j}]$ with the group lasso penalizing only off-diagonal blocks. Hence, using the shorthand notation, the solution is given by

$$\mathsf{Z}_{\mathsf{i},\mathsf{j}}^{(\mathsf{m})} = \begin{cases} \mathsf{Q}_{\mathsf{i},\mathsf{j}}^{(\mathsf{m})} + \mathsf{U}_{\mathsf{i},\mathsf{j}}^{(\mathsf{m}-1)} & \text{for} \quad \mathsf{i} = \mathsf{j}, \\ \mathcal{S}_{\lambda_\mathsf{n}/\rho}(\mathsf{Q}_{\mathsf{i},\mathsf{j}}^{(\mathsf{m})} + \mathsf{U}_{\mathsf{i},\mathsf{j}}^{(\mathsf{m}-1)}) & \text{for} \quad \mathsf{i} \neq \mathsf{j}, \end{cases}$$

where $\mathcal{S}_\mathsf{t}(\mathsf{M}) = (1 - \frac{\mathsf{t}}{\|\mathsf{M}\|_F})_+\mathsf{M}$ is the group-wise soft-thresholding operator (Friedman et al., 2010). We always use $\mathsf{Z}^{(0)} = \mathsf{U}^{(0)} = \mathrm{diag}(\mathsf{R})$ as the starting point and iterate until the relative residual is small, namely until $\|\mathsf{Q}^{(m)} - \mathsf{Z}^{(m)}\|_F / \|\mathsf{Q}^{(m)}\|_F \leq 10^{-4}$.

4.8 Simulation Study

The finite sample performance of the proposed methodology is explored in a small simulation study. We devise three simulation setups underlining several claims we intend to make. Below, we describe the three setups briefly, while a full description is available in the supplementary material.

Setup 1 is closely related to Model 1 of Qiao et al. (2019), which generates the functional datum in every node as a zero-mean Gaussian with the covariance being rank 5 with Fourier eigenfunctions and equal eigenvalues, with the precision matrix chosen

such that a functional AR(2) process is formed between the nodes. This is done in a perfectly regular way such that all rank-one projections of the processes form AR(2) processes on their own, and the dependencies are created over the whole functional domain. We change this slightly to rank 10, eigenvalues $\lambda_j = 1/l$ for $k = l, \ldots, 10$, and create the AR(2) dependencies only between the eigenfunctions corresponding to $\lambda_6, \ldots, \lambda_{10}$.

Setup 2 also utilizes Fourier eigenfunctions but differs from Setup 1 in two aspects. Firstly, the rank is not finite, with eigenvalues decaying quadratically as $\lambda_l = 1/l^2$. Secondly, the functional AR(2) dependencies are not flat, they are created only locally on one tenth of the functional domain corresponding to every node. This makes them harder to discover after a projection. We consider these local dependencies in the time domain more realistic as opposed to the perfectly global spectral dependencies in Setup 1, where eigenfunctions directly influence themselves across different nodes.

Setup 3 superposes independent Fourier rank-5 processes with fractional Brownian motions (with the parameter $H = 0.2$, i.e. a relatively slow eigendecay). But here, the dependency is only formed between the fractional Brownian motions. In other words, every functional datum has an independent smooth component and a dependent but rough component. We believe such rough short-scale dependencies could be interesting e.g. in portfolio optimization (Carvalho et al., 2007) with a short time horizon (Lin and SenGupta, 2021).

The proposed functional graphical lasso is implemented using the ADMM algorithm described in Section 4.7 and compared against the functional graphical lasso of Qiao et al. (2019) implemented using a block coordinate gradient descent. The computer code for the latter was kindly provided to us by the authors, and we slightly modified it to allow for non-regular settings (namely Setup 3). Note that the very fact that this modification can be done and is guaranteed to work, stems from the theoretical development in this paper. We do not compare against other, possibly non-functional approaches, since these have been shown inferior by Qiao et al. (2019).

The results are averages of 24 independent simulation runs. They are reported in terms of mean ROC curves, showing the performance across all values of the the lasso penalty parameter λ_n leading to different sparsity levels. This not only leads to fair comparisons, but also note that λ_n is typically chosen in practice in order to obtain a desired sparsity level (Danaher et al., 2014). Alternatively, the stability selection approach of Meinshausen and Bühlmann (2010) can be used.

The competing projection-based approach of Qiao et al. (2019) requires a user to choose the projection levels, i.e. the rank and the number of B-splines. The authors provide a standard prediction-based cross-validation approach to choose first the number of

B-splines and then the rank. While their approach works reasonably well for the former, we see no reason why it should work for the latter. In fact, it not always does in our experience, e.g. in Setup 2. Hence we show two versions of the algorithm in our simulations, one with the cross-validated tuning parameters, and the other with the no. of B-splines fixed at 15 and the rank fixed at 5. Those are two arbitrary and rather low choices one might think about given the other parameters of the problems, and can be interpret easily for comparison purposes. On the other hand, the proposed methodology does not require a choice of any tuning parameters, which is a genuine practical advantage.

We fix $n = p = 100$ in order to facilitate comparisons with Qiao et al. (2019) or even Zapata et al. (2022). Still, we do not emulate specifically the simulation setups of Zapata et al. (2022) or include their method in our comparisons for the following reasons. While the approach of Qiao et al. (2019) needs a choice of two tuning parameters, they are both easily interpretable, and a relatively simple way of choosing them is provided. On the other hand, the approach of Zapata et al. (2022) also requires a choice of two tuning parameters: the no. of partially separable components and a tuning parameter weighing their sparsity levels together. But the first one is chosen arbitrarily (as the proportion of variance explained) while the second one is chosen in an oracle fashion, and does not have a straightforward interpretation. The point of this simulation study is not to demonstrate a general superiority of our approach, there is in fact no reason why our methodology should outperform that of Qiao et al. (2019) or Zapata et al. (2022) for well chosen values of their respective tuning parameters. But we rather aim to demonstrate the advantages of not being forced to choose any tuning parameters, which is an implication of the theoretical development in this paper, free of any structural assumptions. And this point is self-evident in the case of Zapata et al. (2022).

Figure 4.1 displays the results of our simulation study. We can see that in Setup 1, the proposed methodology matches that of Qiao et al. (2019). Even though we increased the number of Fourier eigenfunctions to 10 and only created dependencies between the second group of five, the cross-validation approach of Qiao et al. (2019) correctly identifies the number of components needed, and matches the performance of the proposed method back. On the other hand, the poor performance of the fixed pre-chosen projection in Setup 1 shows the dangers of choosing the projection level too low. In Setup 2, on the other hand, the cross-validation approach of Qiao et al. (2019) underestimates the rank, leading to a worse performance than with the pre-chosen values of the projection levels. Still, the proposed approach clearly outperforms both of its competitors. Finally, in Setup 3, the proposed approach vastly outperform its competitors, because this simulation setup generally disfavors projections. While cross-validation leads to higher projection levels than the pre-chosen ones in this case, it does not retain a sufficient number of components.

Overall, Setup 1 constitutes an example where not performing projections even in

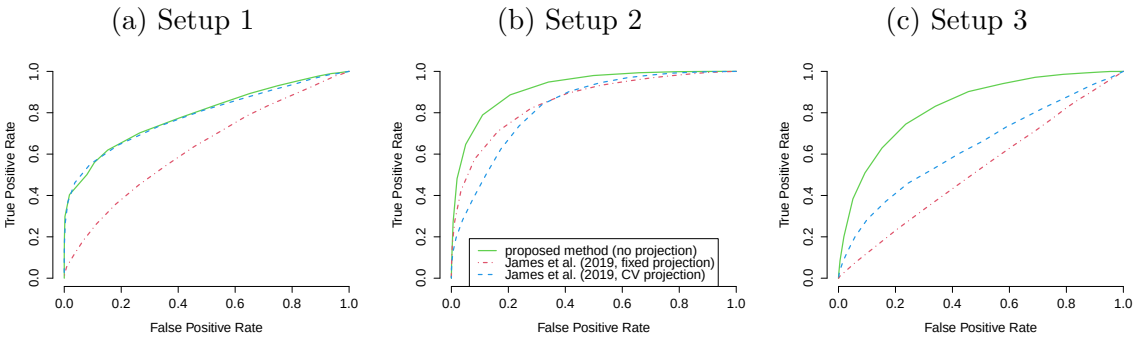


Figure 4.1: ROC curves in the three simulation setups for the proposed method and the projection approach of Qiao et al. (2019) with pre-chosen values for the projection dimensions and with cross-validated choice of projection dimensions.

a perfectly low-dimensional case poses no issues. Secondly, Setup 2 illustrates that projecting data in a not perfectly low-dimensional case can lead to a loss of information. Finally, Setup 3 constitutes a case where projections are simply not advisable.

5 Future Directions

The thesis raises many interesting problems which remain unsolved in spite of our best efforts. At the same time, our methods are somewhat alien to how things are usually done and consequently, there are many problems to which our methods provide elegant solutions. We discuss below some of the ways in which the work in this thesis can be built upon, in an increasing order of difficulty.

5.1 Extreme Points of the Set of Completions

A sizeable amount of literature in linear algebra (Li and Tam, 1994; Grone et al., 1990) and statistics (Parthasarathy, 2002) is dedicated to the study of the extreme points of so-called correlation matrices, which are really positive semi-definite matrices with diagonal entries all equal to one. Because correlation matrices can also be regarded as positive semi-definite completions of the identity matrix over off-diagonal entries, Theorem 2.2 of Chapter 2 provides a remarkably simple characterization of their extreme points as well as those of completions of any partially specified matrix. Building on this characterization could help us understand the set of completions in greater detail, which seems to be a difficult problem even for matrices.

5.2 Covariance Selection and Estimation from Incomplete Observations

There are many classical problems concerning covariance estimation which are well-understood for multivariate data, such as covariance selection and estimation with incomplete observation, but whose resolution for multivariate functional data has been obstructed for want of an infinite-dimensional counterpart of the likelihood function. It appears that the likelihood approach devised in Chapter 4 can address these problems with relative ease. Covariance selection, which can be understood as using the *a priori*

knowledge of certain conditional independence relationships for the purpose of improving covariance estimation, can be performed by merely maximizing the proposed multivariate functional likelihood under certain constraints. Similarly, an expectation-maximization algorithm can be formulated based on the proposed likelihood for estimating the covariance from incomplete observations as it is done in a multivariate setting.

5.3 Large-scale Nonparametric Covariance Modelling

For many applications such as those in geostatistics, estimation of covariance is a computationally challenging task and as a result one often defers to parametric methods of estimation. Using results from Chapter 2, we can extend the ideas of Chapter 1 towards a method which can pool local nonparametric estimates of the covariance on small regions, which can be computed with relative ease, to construct a global covariance estimate which propagates the local estimate using canonical completion. This can be thought of as nonparametric model of covariance and appears particularly natural for geostatistical applications dealing with natural processes which are known to propagate locally.

5.4 Speed of Spatio-Temporal Processes

Chapter 3 makes it possible to think of spatio-temporal processes as graphical models. In particular, we can think of the temporal evolution of a process in terms of changes propagating through a graph on space and time. Say we are interested in quantifying how quickly changes at a point in space affect nearby points for a given spatio-temporal process $\{X(s,t) : s,t \in \mathbb{R}\}$. We can formalize this as asking whether for a given $\delta s, \delta t > 0$ and every $s' \in (s - \delta s, s + \delta s)$, there is an edge between $X(s,t)$ and $X(s', t + \delta t)$. If there is an edge, then it would mean that change from s has propagated to a region of radius δs by δt amount of time. Furthermore, we can say that the speed v at which the change is propagating satisfies $v > \delta s / \delta t$.

5.5 A Complete Theory of Positive-Definite Completion

Arguably, the biggest shortcoming of this thesis has been the failure to prove the uniqueness of canonical completion for regular domains except for the stationary case under technical conditions which are admittedly opaque and convenient. Although no new hints to the solution have been found as of yet, this remains an obvious venue for the direction of future research efforts, in addition to being the principal impediment towards achieving a more complete theory of positive-definite completion. Furthermore, there are plenty of results that are known to hold for completions of matrices which we haven't been able to prove for kernels. Results from the discrete case suggest that canonical

completions exist for every partially reproducing kernel which admits a completion not just those considered in Chapter 2. A related problem is that of characterizing the entire set of completions in terms of certain bounded linear maps, as was done in Gohberg et al. (1989) for completions of matrix specified on a band around the diagonal.

A Appendix A. The Completion of Covariance Kernels

This appendix collects some additional graphs and the proofs of the formal statements in Chapter 1.

A.1 Graphs

Figure A.1 depicts plots of covariance completions in the case of regular and sparse observations for K_1, K_2 and K_3.

A.2 Proofs of Formal Statements

We will arrange our proofs into subsections that parallel the corresponding sections of the paper. We shall make extensive use of the projection theorem as well as certain isometries between Hilbert spaces, such as the Loève isometry. For tidiness, we introduce some shorthand notation for the restrictions $K_\Omega(t,\cdot)$ or $K(t,\cdot)$: for $u \in I$ and $J \subset I$, we denote by $k_{u,J}$ the function $k_{u,J} : J \to \mathbb{R}$ given by $k_{u,J}(v) = K(u,v)$ or $K_\Omega(u,v)$ according to the context. Similarly, we denote by $k^\star_{u,J}$ the function $k^\star_{u,J} : J \to \mathbb{R}$ given by $k^\star_{u,J}(v) = K_\star(u,v)$.

Moreover, for every covariance K we have $k_{u,J} \in \mathcal{H}(K_J)$ for every $u \in I$ and $J \subset I$. This is known as the restriction theorem (Paulsen and Raghupathi, 2016, Corollary 5.8). For $f, g \in \mathcal{H}(K_J)$ we shall denote the norm $\|f\|_{\mathcal{H}(K_J)}$ and the inner product $\langle f, g\rangle_{\mathcal{H}(K_J)}$ in $\mathcal{H}(K_J)$ simply as $\|f\|$ and $\langle f, g\rangle$, since the Hilbert space can always be inferred form the domain of the involved function.

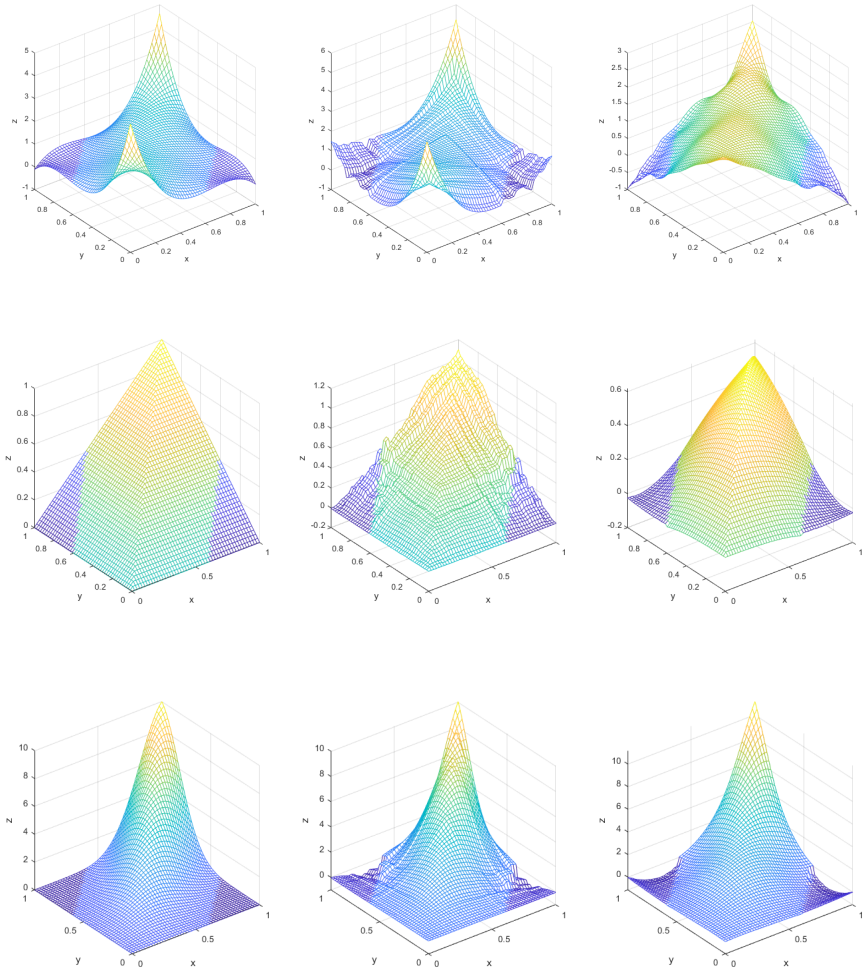


Figure A.1: Covariance Completions of K_1 (top), K_2 (middle) and K_3 (bottom) for $m = 9$ and $N = 300$. For every row, the plot on the left is the true covariance, the plot in the middle is the completion from regular observations using the pairwise estimator on the right is the completion for sparse observations.

A.2.1 The Canonical Completion

The completion $K_\star$ as defined in Equation (1.2) is well-defined because the restrictions of $K_\Omega(s,\cdot)$ and $K_\Omega(t,\cdot)$ to $I_1 \cap I_2$ belong to the RKHS $\mathcal{H}(K_{I_1\cap I_2})$ thanks to the restriction theorem (Paulsen and Raghupathi, 2016, Corollary 5.8). We shall however make use of a stronger result in order to prove our theorems.

Let H_J denote the closed subspace spanned by $\{k_x : x \in J\}$ in $\mathcal{H}$ and let Π_J denote the projection from $\mathcal{H}$ to H_J. Then the subspace H_J is isomorphic to the reproducing kernel Hilbert space $\mathcal{H}(K_J)$.

Theorem A.1. *There exists an isometry $\rho : H_J \to \mathcal{H}(K_J)$ such that its adjoint ρ^* satisfies $\rho^* g|_J = g$ for $g \in H_J$.*

Proof. Define a linear map $\sigma_0 : \mathrm{Span}\{k_{x,J} : x \in J\} \to H_J$ by $\sigma_0(k_{x,J}) = k_x$ for $x \in J$. For $f = \sum_{j=1}^n c_j k^J_{x_j}$, we have

$$\|\sigma_0(f)\|^2 = \sum_{i,j=1}^n c_i c_j K(x_i, x_j) = \|f\|^2$$

Therefore, if $f = 0$, then $\sigma_0(f) = 0$. It follows that the map σ_0 is well-defined, injective and continuous. Moreover, it maps $\mathrm{Span}\{k_{x,J} : x \in E\}$ onto $\mathrm{Span}\{k_{x,J} : x \in J\}$.

Extending σ_0 by continuity from $\mathrm{Span}\{k_{x,J} : x \in J\}$ to $\mathcal{H}(K_J)$ gives $\sigma : \mathcal{H}(K_J) \to H_J$ such that $\sigma(f) = \sigma_0(f)$ for $f \in \mathrm{Span}\{k_{x,J} : x \in J\}$. Additionally, $\|\sigma(f)\| = \|f\|$ for every $f \in \mathcal{H}(K_J)$ and therefore σ is also well-defined, injective and continuous.

To show that σ is surjective, pick any $g \in H_J$. Then there exists a sequence $\{g_j\}_{j=1}^\infty \subset \mathrm{Span}\{k_x : x \in J\}$ such that $g_j \to g$ in $\mathcal{H}$. Let $f_j \in \mathrm{Span}\{k_{x,J} : x \in J\}$ be such that $\sigma f_j = g_j$ for $j \geq 1$. Since $\{g_j\}_{j=1}^\infty$ is Cauchy, so is $\{f_j\}_{j=1}^\infty$ and therefore it converges to some $f \in \mathcal{H}(K_J)$ such that $\sigma f = g$.

Now, notice that for every $x \in J$,

$$\sigma^* g(x) = \langle \sigma^* g, k_{x,J} \rangle = \langle g, \sigma k_{x,J} \rangle = \langle g, k_x \rangle = g(x)$$

Define $\rho = \sigma^*$ and the conclusion follows. □

We shall see that the above isometry enables a sort of infinite-dimensional matrix algebra with the partial covariance in order to recover its unknown values. We shall refer to it as the *subspace isometry*, in contrast to another isometry we often make use of, which is the *Loève isometry*.

Remark A.1. *Notice that for $f, g \in \mathcal{H}(K)$ and $\Pi_J : \mathcal{H}(K) \to \mathcal{H}(K)$ the closed subspace*

spanned by $\{k_u : u \in J\}$*, we have*

$$\begin{aligned}\langle \Pi_J f, \Pi_J g\rangle_{\mathcal{H}(K)} &= \langle \rho\Pi_J f, \rho\Pi_J g\rangle_{\mathcal{H}(K_J)} \\ &= \langle \Pi_J f|_J, \Pi_J g|_J\rangle_{\mathcal{H}(K_J)} \\ &= \langle f|_J, g|_J\rangle_{\mathcal{H}(K_J)}\end{aligned}$$

since for $t \in J$, $\Pi_J f(t) = \langle \Pi_J f, k_t\rangle = \langle f, \Pi_J k_t\rangle = \langle f, k_t\rangle = f(t)$*. Thus, projection boils down to restriction.*

Proof of Theorem 1.1. It suffices for us to construct a Hilbert space $\mathcal{H}$ with a set of vectors $\{\varphi_x\}_{x\in I} \subset \mathcal{H}$ such that $K_\star(s,t) = \langle \varphi_s, \varphi_t\rangle$ for every $s, t \in I$. Accordingly, we let $\mathcal{H} = \mathcal{H}(K_{I_1}) \oplus \mathcal{H}(K_{I_2})$, the direct sum of the reproducing kernel Hilbert spaces of K_{I_1} and the space K_{I_2}.

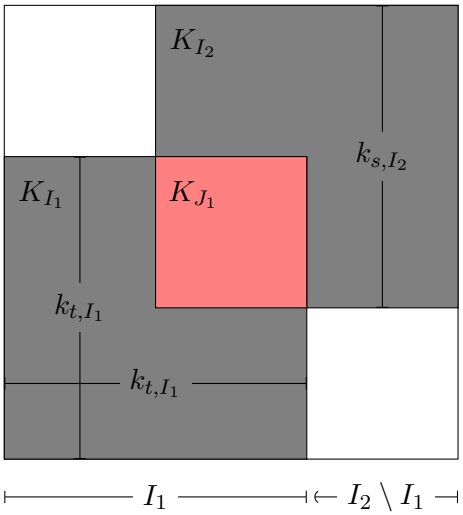


Figure A.2: The Partial Covariance K_Ω

Let H_1 denote the closed subspace in $\mathcal{H}(K_{I_1})$ spanned by $\{k_{t,I_1} : t \in J_1\}$ and similarly, let H_2 denote the closed subspace in $\mathcal{H}(K_{I_2})$ generated by $\{k_{t,I_2} : t \in J_1\}$. By Theorem A.1, both H_1 and H_2 are isomorphic to $\mathcal{H}(K_{J_1})$ with the restrictions $\rho_1 : H_1 \to \mathcal{H}(K_{J_1})$ and $\rho_2 : H_2 \to \mathcal{H}(K_{J_1})$ given by $\rho_1 f = f|_{J_1}$ and $\rho_2 g = g|_{J_1}$, serving as isometries. It follows that H_1 and H_2 are isomorphic, with the isometry $\rho_1^*\rho_2 : H_2 \to H_1$. Also, let $\Pi_{J_1} : \mathcal{H}(K_{I_2}) \to \mathcal{H}(K_{I_2})$ denote the projection to H_2.

Define φ_t as follows,

$$\varphi_t = \begin{cases} k_{t,I_1} \oplus 0 & \text{if } t \in I_1 \\ \rho_1^*\rho_2 \Pi_{J_1} k_{t,I_2} \oplus [k_{t,I_2} - \Pi_{J_1} k_{t,I_2}] & \text{if } t \in I_2 \setminus I_1 \end{cases}$$

All that remains now, is for us to verify that $\langle \varphi_s, \varphi_t \rangle$ is indeed equal to $K_\star(s,t)$ for every $s,t \in I$. We shall do this on a case-by-case basis as follows:

Case 1. If both s and $t \in I_1$, then $\langle \varphi_s, \varphi_t \rangle = \langle k_{s,I_1}, k_{t,I_1} \rangle + 0 = K_{I_1}(s,t) = K_\star(s,t)$.

Case 2. If $s \in I_1$ and $t \in I_2 \setminus I_1$, then by the projection theorem and Theorem A.1 we get

$$\langle \varphi_s, \varphi_t \rangle = \langle k_{s,I_1}, \rho_1^* \rho_2 \Pi_{J_1} k_{t,I_2} \rangle + 0 = \langle \rho_1 k_{s,I_1}, \rho_2 \Pi_{J_1} k_{t,I_2} \rangle = \langle k_{s,J_1}, k_{t,J_1} \rangle$$

If $s \in J_1$, then $\langle k_{s,J_1}, k_{t,J_1} \rangle = K_{J_1}(s,t) = K_\star(s,t)$. On the other hand, if $s \in I_1 \setminus J$, $\langle k_{s,J_1}, k_{t,J_1} \rangle = K_\star(s,t)$ by definition.

Case 3. Covered by Case 2 by symmetry.

Case 4. If both s and $t \in I_2 \setminus I_1$, then

$$\langle \varphi_s, \varphi_t \rangle = \langle \rho_1^* \rho_2 \Pi_{J_1} k_{s,I_2}, \rho_1^* \rho_2 \Pi_{J_1} k_{t,I_2} \rangle + \langle k_{s,I_2} - \Pi_{J_1} k_{s,I_2}, k_{t,I_2} - \Pi_{J_1} k_{t,I_2} \rangle$$

By Theorem A.1, $\langle \rho_1^* \rho_2 \Pi_{J_1} k_{s,I_2}, \rho_1^* \rho_2 \Pi_{J_1} k_{t,I_2} \rangle = \langle \Pi_{J_1} k_{s,I_2}, \Pi_{J_1} k_{t,I_2} \rangle = \langle k_{s,J_1}, k_{t,J_1} \rangle$. And using the projection theorem,

$$\begin{aligned}
&\langle k_{s,I_2} - \Pi_{J_1} k_{s,I_2}, k_{t,I_2} - \Pi_{J_1} k_{t,I_2} \rangle \\
&\qquad = \langle k_{s,I_2} - \Pi_{J_1} k_{s,I_2}, k_{t,I_2} \rangle - \langle k_{s,I_2} - \Pi_{J_1} k_{s,I_2}, \Pi_{J_1} k_{t,I_2} \rangle \\
&\qquad = \langle k_{s,I_2}, k_{t,I_2} \rangle - \langle \Pi_{J_1} k_{s,I_2}, k_{t,I_2} \rangle - 0 \\
&\qquad = \langle k_{s,I_2}, k_{t,I_2} \rangle - \langle \Pi_{J_1} k_{s,I_2}, \Pi_{J_1} k_{t,I_2} \rangle \\
&\qquad = K_{I_2}(s,t) - \langle k_{s,J_1}, k_{t,J_1} \rangle
\end{aligned}$$

Thus, $\langle \varphi_s, \varphi_t \rangle = K_{I_2}(s,t) = K_\star(s,t)$.

Now that we have established that $K_\star(s,t) = \langle \varphi_s, \varphi_t \rangle$ is a covariance extension of K_Ω, we need only verify that $K(s,t) = \langle k_{s,J_1}, k_{t,J_1} \rangle$ for $(s,t) \in \Omega^c$. For $s \in I_2 \setminus I_1$ and $t \in I_1 \setminus I_2 \subset I_1$

$$\begin{aligned}
\langle \varphi_s, \varphi_t \rangle &= \langle \rho_1^* \rho_2 \Pi_{J_1} k_{s,I_2} \oplus [k_{s,I_2} - \Pi_{J_1} k_{s,I_2}], k_{t,I_1} \oplus 0 \rangle \\
&= \langle \rho_1^* \rho_2 \Pi_{J_1} k_{s,I_2}, k_{t,I_1} \rangle \\
&= \langle \rho_2 \Pi_{J_1} k_{s,I_2}, \rho_1 k_{t,I_1} \rangle \\
&= \langle k_{s,J_1}, k_{t,J_1} \rangle
\end{aligned}$$

This completes the proof. □

Remark A.2. *Nothing in the proof above requires that I_1 and I_2 have to be intervals of the real line. In fact, the result holds true so long as I_1 and I_2 are any two sets with a non-empty intersection.*

Proof of Theorem 1.2. By Theorem 1.1, the result of a 2-serrated completion is a valid completion and therefore, the same should be true for successive 2-serrated completions.

Let $K_\star$ denote the completion obatined by using Algorithm 1. In order to show that the resulting completion is independent of the order in which the completion is carried out, it suffices to show that for $s, t \in I$ separated by J_p and J_q, we have that $\langle k_{s,J_p}, k_{t,J_p} \rangle = \langle k_{s,J_q}, k_{t,J_q} \rangle$.

We proceed by induction. The statement is vacuously true for the case $m = 2$ by Theorem 1.1. We shall prove it for $m = 3$ and on. For $m = 3$, it suffices for us to show that for $(s, t) \in (I_3 \setminus I_2) \times (I_1 \setminus I_2)$,

$$\langle k^\star_{s,J_1}, k^\star_{t,J_1} \rangle = \langle k^\star_{s,J_2}, k^\star_{t,J_2} \rangle$$

since only such s and t are separated by both J_1 and J_2. By definition, $k^\star_{s,J_1}(u) = \langle k^\star_{s,J_2}, k^\star_{u,J_2} \rangle$ for $u \in J_1$. $k^\star_{s,J_1}$ can be written in terms of $k^\star_{s,J_2}$ in a more concise way, as an image of a linear operator.

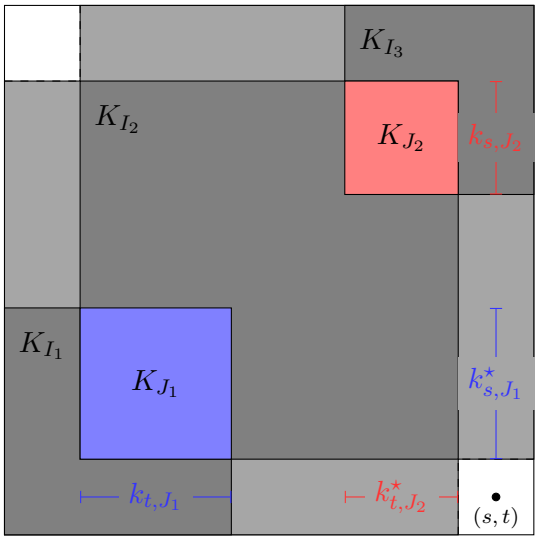


Figure A.3: The Partial Covariance K_Ω

Let $\rho_1 : \mathcal{H}(K_{I_2}) \to \mathcal{H}(K_{J_1})$ and $\rho_2 : \mathcal{H}(K_{I_2}) \to \mathcal{H}(K_{J_2})$ denote the restrictions given by $\rho_1 f = f|_{J_1}$ and $\rho_2 f = f|_{J_2}$ respectively. Let $\Pi_1 : \mathcal{H}(K_{I_2}) \to \mathcal{H}(K_{I_2})$ denote the

projection to the closed subspace generated by $\{k_{u,I_2} : u \in J_1\}$. Now, for $u \in J_1$,

$$\begin{aligned}\rho_1\Pi_1\rho_2^* k^\star_{s,J_2}(u) &= \langle \rho_1\Pi_1\rho_2^* k^\star_{s,J_2}, k_{u,J_1}\rangle \\ &= \langle k^\star_{s,J_2}, \rho_2\Pi_1\rho_1^* k_{u,J_1}\rangle \\ &= \langle k^\star_{s,J_2}, \rho_2\Pi_1 k^\star_{u,I_2}\rangle \\ &= \langle k^\star_{s,J_2}, \rho_2 k^\star_{u,I_2}\rangle \\ &= \langle k^\star_{s,J_2}, k^\star_{u,J_2}\rangle.\end{aligned}$$

because $\rho_1^* k_{u,J_1} = k_{u,I_2}$, $\rho_2 k^\star_{u,I_2} = k^\star_{u,J_2}$, and $\Pi_1 k^\star_{u,I_2} = k^\star_{u,I_2}$ as $u \in J_1$. Therefore, $k^\star_{s,J_1} = \rho_1\Pi_1\rho_2^* k^\star_{s,J_2}$. Using this representation,

$$\begin{aligned}\langle k^\star_{s,J_1}, k^\star_{t,J_1}\rangle &= \langle \rho_1\Pi_1\rho_2^* k^\star_{s,J_2}, k^\star_{t,J_1}\rangle \\ &= \langle k^\star_{s,J_2}, \rho_2\Pi_1\rho_1^* k^\star_{t,J_1}\rangle \\ &= \langle k^\star_{s,J_2}, k^\star_{t,J_2}\rangle.\end{aligned}$$

Thus, $K_\star$ indeed satisfies the separation condition. Uniqueness follows by observing that the separation condition uniquely determines $K_\star$ given $K_\star|_{I_3\times I_3}$ and $K_\star|_{S_1^2\times S_1^2}$, which is in turn uniquely determined given $K_\star|_{I_1\times I_1}$ and $K_\star|_{I_2\times I_2}$.

Now, assuming the statement for $m \le q$, we consider the case $m = q+1$. Thus, $K_\star|_{[\cup_{j=1}^k I_j]\times[\cup_{j=1}^k I_j]}$ and $K_\star|_{[\cup_{j=2}^{k+1} I_j]\times[\cup_{j=2}^{k+1} I_j]}$ are uniquely determined. It suffices to verify the separating condition for the remaining part. So we need to show that for $(s,t) \in (I_{q+1} \setminus I_q) \times (I_1 \setminus I_2)$, that,

$$\langle k^\star_{s,J_1}, k^\star_{t,J_1}\rangle = \langle k^\star_{s,J_2}, k^\star_{t,J_2}\rangle = \cdots = \langle k^\star_{s,J_q}, k^\star_{t,J_q}\rangle.$$

Pick $1 \le p < q$, and let $I_1' = [\cup_{j=1}^p I_j]$, $I_2' = [\cup_{j=p+1}^q I_j]$, and $I_3' = I_{q+1}$. Then $\Omega' = \cup_{j=1}^3 I_j' \times I_j'$ is a serrated domain of three intervals and we are back to the case when $m = 3$. This implies that, $\langle k^\star_{s,J_1'}, k^\star_{t,J_1'}\rangle = \langle k^\star_{s,J_2'}, k^\star_{t,J_2'}\rangle$ for $J_1' = I_1' \cap I_2' = J_p$ and $J_2' = I_2' \cap I_3' = J_q$. Since p was chosen arbitrarily, it follows that, $\langle k^\star_{s,J_p}, k^\star_{t,J_p}\rangle = \langle k^\star_{s,J_q}, k^\star_{t,J_q}\rangle$ for $1 \le p < q$. Uniqueness follows the same way as in the case $m = 3$. Hence proved. □

Lemma A.1. *Let $J \subset I$ be a separator of Ω containing J_p for some p. Then K as defined above, satisfies*

$$\langle k_{s,J_p}, k_{t,J_p}\rangle_{\mathcal{H}(K_{J_p})} = \langle k_{s,J}, k_{t,J}\rangle_{\mathcal{H}(K_J)} \tag{A.1}$$

for every $s,t \in I$ separated by J.

Proof. By Theorem 1.2, $K_\star(s,t) = \langle k^\star_{s,J_p}, k^\star_{t,J_p}\rangle$ and by Remark A.1, $\langle k^\star_{s,J_p}, k^\star_{t,J_p}\rangle = \langle \Pi_p k^\star_{s,J}, \Pi_p k^\star_{t,J}\rangle$, where $\Pi_p : \mathcal{H}(K_\star|_{J\times J}) \to \mathcal{H}(K_\star|_{J\times J})$ denotes the projection to the closed subspace spanned by $\{k^\star_{u,J} : u \in J_p\}$. Thus, all we need to show is for $s,t \in I$

separated by $J \subset I$,

$$\begin{aligned}\langle k^\star_{s,J} - \Pi_p k^\star_{s,J}, k^\star_{t,J} - \Pi_p k^\star_{t,J}\rangle &= \langle k^\star_{s,J}, k^\star_{t,J}\rangle - \langle \Pi_p k^\star_{s,J}, \Pi_p k^\star_{t,J}\rangle \\ &= \langle k^\star_{s,J}, k^\star_{t,J}\rangle - \langle k^\star_{s,J_p}, k^\star_{t,J_p}\rangle \\ &= 0.\end{aligned}$$

Recall that $J_p = [a_{p+1}, b_p]$. Define $J_- = \{u \in J : u < v \text{ for some } v \in J_p\}$ and

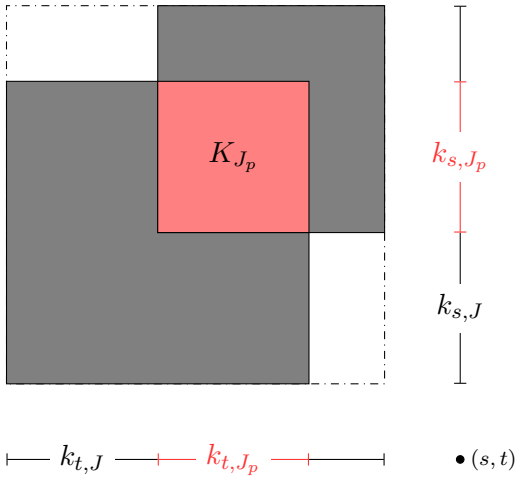


Figure A.4: The Partial Covariance K_J

$J_+ = \{u \in J : u > v \text{ for some } v \in J_p\}$. Thus, $J = J_- \cup J_p \cup J_+$. Notice that for $u \in J \setminus J_+$,

$$\begin{aligned}\langle k^\star_{s,J} - \Pi_p k^\star_{s,J}, k^\star_{u,J} - \Pi_p k^\star_{u,J}\rangle &= \langle k^\star_{s,J} - \Pi_p k^\star_{s,J}, k^\star_{u,J}\rangle - \langle k^\star_{s,J} - \Pi_p k^\star_{s,J}, \Pi_p k^\star_{u,J}\rangle \\ &= K_\star(s,u) - \langle \Pi_p k^\star_{s,J}, k^\star_{u,J}\rangle - 0 \\ &= \langle k^\star_{s,J_p}, k^\star_{u,J_p}\rangle - \langle \Pi_p k^\star_{s,J}, k^\star_{u,J}\rangle \\ &= \langle \Pi_p k^\star_{s,J}, \Pi_p k^\star_{u,J}\rangle - \langle \Pi_p k^\star_{s,J}, k^\star_{u,J}\rangle \\ &= \langle \Pi_p k^\star_{s,J}, \Pi_p k^\star_{u,J} - k^\star_{u,J}\rangle = 0\end{aligned}$$

Therefore, $k^\star_{s,J} - \Pi_p k^\star_{s,J}$ belongs to the closed subspace spanned by $\{k^\star_{u,J} - \Pi_p k^\star_{u,J} : u \in J_+\}$. Similarly, it can be shown that $k^\star_{t,J} - \Pi_p k^\star_{t,J}$ belongs to the closed subspace spanned by $\{k^\star_{u,J} - \Pi_p k^\star_{u,J} : u \in J_-\}$. If we are able to show that these subspaces themselves are mutually orthogonal, we would be done. Arguing as before, for $u \in J_-$ and $v \in J_+$,

$$\begin{aligned}\langle k^\star_{u,J} - \Pi_p k^\star_{u,J}, k^\star_{v,J} - \Pi_p k^\star_{v,J}\rangle &= \langle k^\star_{u,J} - \Pi_p k^\star_{u,J}, k^\star_{v,J}\rangle \\ &= K_\star(u,v) - \langle \Pi_p k^\star_{u,J}, k^\star_{v,J}\rangle \\ &= \langle k^\star_{u,J_p}, k^\star_{v,J_p}\rangle - \langle \Pi_p k^\star_{u,J}, k^\star_{v,J}\rangle \\ &= \langle \Pi_p k^\star_{u,J}, \Pi_p k^\star_{v,J}\rangle - \langle \Pi_p k^\star_{u,J}, k^\star_{v,J}\rangle \\ &= \langle \Pi_p k^\star_{u,J}, \Pi_p k^\star_{v,J} - k^\star_{v,J}\rangle = 0\end{aligned}$$

The conclusion follows. □

Let $J \subset I$ separating $s, t \in I$ be such that $(s,t) \in \Omega^c$. Then for some $1 \leq p, q < m$, $J_p \subset I$ separates $s, t \in I$, since $(s,t) \in D_p \times S_p$ from some $1 \leq p < m$ and $J_q \subset J$ separates $s, t \in I$. By Lemma A.1 and Theorem 1.2,

$$\langle k_{s,J}, k_{t,J} \rangle_{\mathcal{H}(K_J)} = \langle k_{s,J_q}, k_{t,J_q} \rangle_{\mathcal{H}(K_{J_q})}$$
$$\langle k_{s,J_q}, k_{t,J_q} \rangle_{\mathcal{H}(K_{J_q})} = \langle k_{s,J_p}, k_{t,J_p} \rangle_{\mathcal{H}(K_{J_p})}$$

And again by Theorem 1.2, $K_\star(s,t) = \langle k_{s,J_p}, k_{t,J_p} \rangle_{\mathcal{H}(K_{J_p})}$. We have thus shown the following:

Theorem A.2. *If K_Ω is a partial covariance on a serrated domain Ω, then K_Ω has a unique covariance completion $K_\star$ to I which possesses the separation property: for every $s, t \in I$ separated by $J \subset I$,*

$$K_\star(s,t) = \langle k^\star_{s,J}, k^\star_{t,J} \rangle_{\mathcal{H}(K_J)}$$

where $k^\star_{u,J} : J \to \mathbb{R}$ is given by $k^\star_{u,J}(v) = K_\star(u,v)$ for $v \in J$. Furthermore, $K_\star$ can be recursively computed using Algorithm 1.

A.2.2 Canonicity and Graphical models

Proof of Theorem 1.3. Simply use Theorem A.2 in conjunction with Theorem 1.4. □

Proof of Theorem 1.4. The process X is said to form a graphical model with $([0,1], \Omega)$ precisely when for every $s, t \in I$ separated by $J \subset I$, we have

$$\mathrm{Cov}(X_s, X_t | X_J) \equiv \mathbb{E}\left[(X_s - \mathbb{E}\left[X_s | X_J\right])(X_t - \mathbb{E}\left[X_t | X_J\right]) | X_J\right] = 0 \quad \text{a.s.}$$

which is equivalent to saying that $\mathbb{E}\left[X_s X_t | X_J\right] = \mathbb{E}\left[X_s | X_J\right] \mathbb{E}\left[X_t | X_J\right]$ almost surely. According to Loeve (2017), for a Gaussian process X, the conditional expectation $\mathbb{E}\left[X_t | X_J\right]$ is same as the projection $\Pi(X_t | X_J)$ as described in Section 1.5. Because the mean of the Gaussian process is zero, we can write the above equation as

$$K(s,t) = \Pi(X_s | X_J) \Pi(X_t | X_J)$$

By the Loève isometry,

$$K(s,t) = \langle \Pi_J k_s, \Pi_J k_t \rangle$$

which reduces to

$$K(s,t) = \langle k_{s,J}, \Pi_J k_{t,J} \rangle.$$

by the subspace isometry from Theorem A.1. Thus, $K(s,t) = \langle K(s,\cdot), K(t,\cdot)\rangle_{\mathcal{H}(K_J)}$ and the conclusion follows. □

A.2.3 Necessary and Sufficient Conditions for Unique Completion

Naturally, we begin by dealing with the 2-serrated case. To this end we prove another existence result which captures how much the completion of a partial covariance on a 2-serrated domain can vary at a given point.

Lemma A.2. *Let K_Ω be a partial covariance on a serrated domain Ω of two intervals, $s \in I_1 \setminus J_1$, $t \in I_2 \setminus J_1$ and $\alpha \in \mathbb{R}$. There exists a covariance extension K of K_Ω such that*

$$K(s,t) = \alpha + \langle k_{s,J_1}, k_{t,J_1}\rangle$$

if any only if

$$|\alpha| \leq \sqrt{K_{I_1}/K_{J_1}(s,s) \cdot K_{I_2}/K_{J_1}(t,t)}.$$

Proof. We begin by getting rid of the part of the covariance that is due to J_1. Let $J_- = I_1 \setminus J_1$, $J_+ = I_2 \setminus J_1$, $J^c = J_- \cup J_+$, $\Omega_0 = [J_- \times J_-] \cup [J_+ \times J_+]$ and define $L_{\Omega_0} : \Omega_0 \to \mathbb{R}$ as $L_{\Omega_0}(s,t) = K_\Omega(s,t) - \langle k_{s,J_1}, k_{t,J_1}\rangle$. This is similar to taking a Schur complement with respect to J_1. Strictly speaking, L_{Ω_0} is not a partial covariance, but it possesses the necessary structure of one and hence we can talk of its extension, which would be a covariance L_0 on J^c such that $L_0|_{\Omega_0} = L_{\Omega_0}$.

Notice that K_Ω has an extension if and only if L_{Ω_0} does. Indeed, if K is an extension of K_Ω then K/K_{J_1} is an extension of L_{Ω_0}. Conversely, if L_0 is an extension of L_{Ω_0}, then $L : I \times I \to \mathbb{R}$ given by $L|_{J^c \times J^c} = L_0$ and 0 otherwise, is a covariance, and so is $K : I \times I \to \mathbb{R}$ given by $K(s,t) = L(s,t) + \langle k_{s,J_1}, k_{t,J_1}\rangle$ for $s,t \in I$. Thus, there is a clear one-one correspondence between the extensions K of K_Ω and the "extensions" L_0 as defined above. Let $s \in I_1 \setminus J_1$ and $t \in I_2 \setminus J_1$.

($\Longrightarrow$) If K is a covariance extension of K_Ω, then $\alpha = K(s,t) - \langle k_{s,J_1}, k_{t,J_1}\rangle = L(s,t)$. Since, L is a covariance,

$$|\alpha| = |L(s,t)| \leq \sqrt{L(s,s) \cdot L(t,t)} = \sqrt{K_{I_1}/K_{J_1}(s,s) \cdot K_{I_2}/K_{J_1}(t,t)}$$

($\Longleftarrow$) Let $\alpha \in \mathbb{R}$ with the given property. It suffices to show that there is an extension L_0 of L_{Ω_0} such that $L_0(s,t) = \alpha$. To do this, we shall rearrange the points of $J_- \cup J_+$ so that the region over which L_0 is known, resembles a serrated domain.

Let $I_- = J_- \setminus \{u\}$ and $I_+ = J_+ \setminus \{v\}$. Now consider the sets $J_- = I_- \cup \{u\}$, $\{u, v\}$ and $\{v\} \cup I_+ = J_+$. These exhibit an overlapping pattern resembling that of the intervals of a serrated domain, although they are not intervals themselves – see Figure A.5. In light of Remark A.2, by applying the completion procedure in Equation 1.2 twice or equivalently using Algorithm 1 we can complete the partial covariance on this "serrated-type-domain". This permits us to conclude that there exists a covariance L such that $L(u, v) = \alpha$ for every α as described above and the conclusion for K follows from the correspondence between the two.

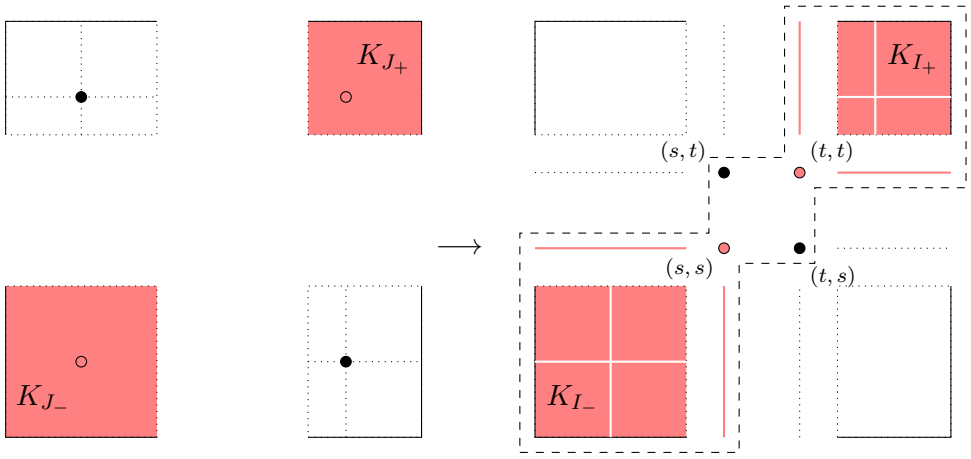


Figure A.5: Rearrangement of $J^c = J_- \cup J_+$

□

The following corollary is immediate.

Corollary A.1. *Let K_Ω be a partial covariance on a 2-serrated domain Ω. Then K_Ω has a unique extension if and only if $K_{I_1}/K_{J_1} = 0$ or $K_{I_2}/K_{J_1} = 0$.*

To extend this result to all serrated domains by induction we need to understand the effect uniqueness has on Schur complements.

Lemma A.3. *Let K_Ω be a partial covariance on a 2-serrated domain. If K_Ω has a unique extension $K_\star$, then $K_{I_1}/K_{J_1} = K_\star/K_{I_2}$ and $K_{I_2}/K_{J_1} = K_\star/K_{I_1}$.*

Proof. For $s, t \in I_1 \setminus J_1 = I \setminus I_2$,

$$\begin{aligned} K_{I_1}/K_{J_1}(s,t) - K_\star/K_{I_2}(s,t) &= K_{I_1}(s,t) - \langle k_{s,J_1}, k_{t,J_1} \rangle - K_\star(s,t) + \langle k_{s,I_2}, k_{t,I_2} \rangle \\ &= \langle k_{s,I_2}, k_{t,I_2} \rangle - \langle k_{s,J_1}, k_{t,J_1} \rangle \\ &= \langle k_{s,I_2}, k_{t,I_2} \rangle - \langle k_{s,I_2}, \Pi_{J_1} k_{t,I_2} \rangle \\ &= \langle k_{s,I_2}, k_{t,I_2} - \Pi_{J_1} k_{t,I_2} \rangle \end{aligned}$$

where $\Pi_{J_1} : \mathcal{H}(K_{I_2}) \to \mathcal{H}(K_{I_2})$ denotes the projection to the closed subspace spanned by $\{k_{u,I_2} : u \in J_1\}$. To see why this term vanishes we reason as follows.

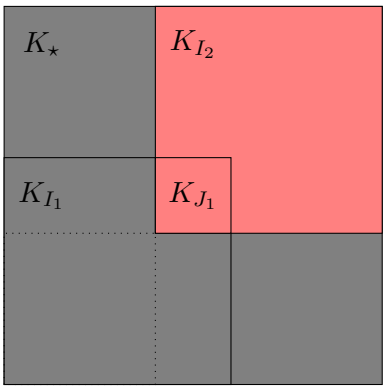


Figure A.6: The covariances K_{J_1}, K_{I_1}, K_{I_2} and $K_\star$

Observe that for $u \in I_2$,

$$\begin{aligned}\langle k_{u,I_2}, k_{t,I_2} - \Pi_{J_1} k_{t,I_2} \rangle &= \langle k_{u,I_2}, k_{t,I_2} \rangle - \langle k_{u,I_2}, \Pi_{J_1} k_{t,I_2} \rangle \\ &= k_{t,I_2}(u) - \langle \Pi_{J_1} k_{u,I_2}, \Pi_{J_1} k_{t,I_2} \rangle \\ &= K_\star(u,t) - \langle k_{u,J_1}, k_{t,J_1} \rangle \\ &= 0.\end{aligned}$$

Therefore, $k_{t,I_2} - \Pi_{J_1} k_{t,I_2} = 0$ and the conclusion follows. Similarly, we can show that $K_{I_2}/K_{J_1} = K_\star/K_{I_1}$. □

In other words, uniqueness causes certain Schur complements to reduce to “smaller" Schur complements. We are now ready to prove Theorem 1.5.

Proof of Theorem 1.5. We shall use induction on m. The base case $m = 2$ follows from Corollary A.1. Assume that the result holds for some $m \geq 2$ and consider a partial covariance K_Ω on an $(m+1)$-serrated domain Ω. Let

$$\begin{aligned}&\bar{I}_1 = \cup_{j=1}^{m} I_j, && \Omega_1 = \cup_{j=1}^{m} I_j \times I_j \subset \bar{I}_1 \times \bar{I}_1, && K_{\Omega_1} = K_\Omega|_{\Omega_1}, \\ &\bar{I}_2 = \cup_{j=2}^{m+1} I_j, && \Omega_2 = \cup_{j=2}^{m+1} I_j \times I_j \subset \bar{I}_2 \times \bar{I}_2, && K_{\Omega_2} = K_\Omega|_{\Omega_2},\end{aligned}$$

and $\bar{\Omega} = (\bar{I}_1 \times \bar{I}_1) \cup (\bar{I}_2 \times \bar{I}_2)$.

If K_Ω admits a unique extension then so do the partial covariances K_{Ω_1} and K_{Ω_2}, for otherwise using Theorem 1.1, one can complete two distinct completions of K_{Ω_1} and K_{Ω_2} to get two distinct completions of K_Ω. By the induction hypothesis, there exist

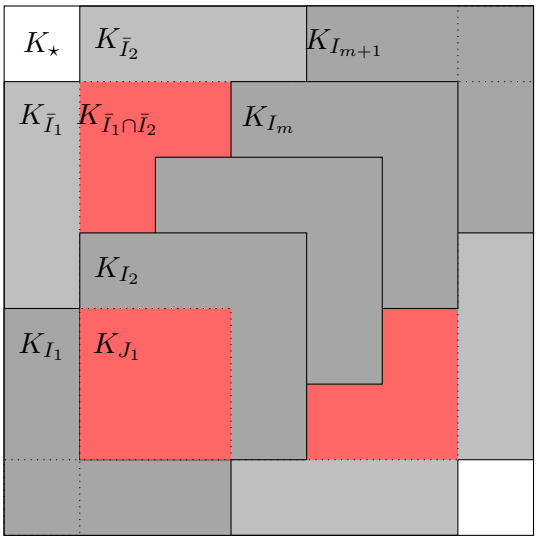


Figure A.7: The covariance extensions $K_{\bar{I}_1}$, $K_{\bar{I}_2}$, $K_{\bar{I}_1\cap\bar{I}_2}$ and $K_\star$

$r_1 \in \{1,\dots,m\}$ and $r_2 \in \{2,\dots,m+1\}$ such that $K_{I_p}/K_{J_p} = 0$ for $1 \le p < r_1 \vee r_2$ and $K_{I_{q+1}}/K_{J_q} = 0$ for $r_1 \wedge r_2 \le q < m+1$. Pick any r such that $r_1 \wedge r_2 \le r \le r_1 \vee r_2$. Then $K_{I_p}/K_{J_p} = 0$ for $1 \le p < r$ and $K_{I_{q+1}}/K_{J_q} = 0$ for $r \le q < m+1$.

To show the converse, assume that $K_{I_p}/K_{J_p} = 0$ for $1 \le p < r$ and $K_{I_{q+1}}/K_{J_q} = 0$ for $r \le q < m+1$ for some $r \in \{1,\dots,m+1\}$. Then $K_{I_p}/K_{J_p} = 0$ for $1 \le p < r \wedge m$ and $K_{I_{q+1}}/K_{J_q} = 0$ for $r \wedge m \le q < m$ and $K_{I_p}/K_{J_p} = 0$ for $2 \le p < r \vee 2$, and that $K_{I_{q+1}}/K_{J_q} = 0$ for $r \vee 2 \le q < m+1$. By the induction hypothesis, it follows that K_{Ω_1} and K_{Ω_2} both admit unique completions, say $K_{\bar{I}_1}$ and $K_{\bar{I}_2}$ respectively.

Due to uniqueness, $K_{\bar{I}_1}(s,t) = K_{\bar{I}_2}(s,t)$ for $s,t \in \bar{I}_1 \cap \bar{I}_2$, so together they form a partial covariance $K_{\bar{\Omega}}$ on $\bar{\Omega}$ given by $K_{\Omega'}(s,t) = K_{\bar{I}_1}(s,t)$ if $(s,t) \in \Omega_1$ and $K_{\bar{I}_2}(s,t)$ if $(s,t) \in \Omega_2$. To prove that K_Ω admits a unique completion, it suffices to show that $K_{\bar{\Omega}}$ admits a unique completion. Since $\bar{\Omega}$ is a 2-serrated domain, we can use the base case and this reduces to showing that $K_{\bar{I}_1}/K_{\bar{I}_1\cap\bar{I}_2} = 0$ or $K_{\bar{I}_2}/K_{\bar{I}_1\cap\bar{I}_2} = 0$. By applying Lemma A.3 to the 2-serrated domains

$$(I_1 \times I_1) \cup [(\bar{I}_1 \cap \bar{I}_2) \times (\bar{I}_1 \cap \bar{I}_2)] \text{ and } [(\bar{I}_1 \cap \bar{I}_2) \times (\bar{I}_1 \cap \bar{I}_2)] \cup (I_{m+1} \times I_{m+1})$$

we get that $K_{\bar{I}_1}/K_{\bar{I}_1\cap\bar{I}_2} = K_{I_1}/K_{J_1}$ and $K_{\bar{I}_2}/K_{\bar{I}_1\cap\bar{I}_2} = K_{I_{m+1}}/K_{J_m}$, at least one of which has to be zero by our assumption. The conclusion follows. □

A.2.4 Characterisation of All Completions

Proof of Theorem 1.6. Let K be a completion of K_Ω and $\Pi_J : \mathcal{H}(K) \to \mathcal{H}(K)$ denote the projection to the closed subspace spanned by $\{k_u : u \in J\}$. Then

$$\begin{aligned} L(s,t) = K(s,t) - \Big\langle K_\Omega(s,\cdot), K_\Omega(\cdot,t)\Big\rangle_{\mathcal{H}(K_{I_1\cap I_2})} &= \langle k_s, k_t\rangle - \langle k_{s,J}, k_{t,J}\rangle \\ &= \langle k_s, k_t\rangle - \langle \Pi_J k_s, \Pi_J k_t\rangle \\ &= \langle k_s - \Pi_J k_s, k_t - \Pi_J k_t\rangle \end{aligned}$$

is a completion of $L_{\Omega'}$. The converse is obvious. □

Remark A.3. *Notice that if K_Ω is continuous, then so is K and so is the term $\Big\langle K_\Omega(s,\cdot), K_\Omega(\cdot,t)\Big\rangle_{\mathcal{H}(K_{I_1\cap I_2})}$ as a function of s and t. This is because the mapping $t \to k_t$ is continuous since*

$$\|k_{t+h} - k_t\|^2 = K(t+h,t+h) - 2K(t,h) + K(t,t) \to 0$$

as $h \to 0$ and the same would apply to the mapping $t \to \Pi_J k_t$. It follows that L is continuous.

Proof of Lemma 1.1. For $f \in L^2(I)$ we have

$$\langle f, \mathbf{K}f\rangle_2 = \sum_{j=1}^{m} \langle f|_{I_j}, \mathbf{K}_j f|_{I_j}\rangle_2 + \sum_{p=1}^{m-1} \left[2\langle f|_{S_p}, \mathbf{R}_p f|_{D_p}\rangle_2 - \langle f|_{J_p}, \mathbf{J}_p f|_{J_p}\rangle_2\right].$$

Let $g \in L^2(I)$. Then we can write

$$\begin{aligned} \langle g, \mathbf{K}f\rangle_2 &= \tfrac{1}{4}\left[\langle f+g, \mathbf{K}(f+g)\rangle_2 - \langle f-g, \mathbf{K}(f-g)\rangle_2\right] \\ &= \sum_{j=1}^{m} \langle g|_{I_j}, \mathbf{K}_j f|_{I_j}\rangle_2 + \sum_{p=1}^{m-1} \left[\langle g|_{S_p}, \mathbf{R}_p f|_{D_p}\rangle_2 + \langle \mathbf{R}_p g|_{D_p}, f|_{S_p}\rangle_2 - \langle g|_{J_p}, \mathbf{J}_p f|_{J_p}\rangle_2\right] \end{aligned}$$

Thus,

$$\mathbf{K}f(t) = \sum_{j:t\in I_j} \mathbf{K}_j f|_{I_j}(t) + \sum_{p:t\in S_p} \mathbf{R}_p f|_{D_p}(t) + \sum_{p:t\in D_p} \mathbf{R}_p^* f|_{S_p}(t) - \sum_{p:t\in J_p} \mathbf{J}_p f|_{J_p}(t) \text{ a.e.}$$

□

Proof of Theorem 1.7. We use induction. Consider the base case $m = 2$. Using Theorem 1.6, we know that the integral kernel K_{R_1} of $\mathbf{R_1}$ at some point (s,t) is given by the contribution due to the canonical completion which is $\langle k_{s,J_1}, k_{t,J_1}\rangle$ plus the perturbation.

By (Paulsen and Raghupathi, 2016, Theorem 11.18), we can write the first term as

$$\langle k^{\Omega}_{s,J_1}, k^{\star}_{t,J_1} \rangle = \langle \mathbf{J}_1^{-1/2} k^{\Omega}_{s,J_1}, \mathbf{J}_1^{-1/2} k^{\star}_{t,J_1} \rangle_{L^2(J_1)}$$

and therefore corresponding integral operator is given by

$$\left[\mathbf{J}_1^{-1/2}\mathbf{S}_1^*\right]^* \left[\mathbf{J}_1^{-1/2}\mathbf{D}_1\right]. \tag{A.2}$$

Due to (Baker, 1973, Theorem 2) as mentioned before, the second term has to be of the form

$$\mathbf{U}_1^{1/2}\Psi_1\mathbf{V}_1^{1/2}$$

for some bounded linear map $\Psi_1 : L^2(D_1) \to L^2(S_1)$ with $\|\Psi_1\| \leq 1$ where

$$\mathbf{U}_1 = \mathbf{K}_{I_1} - \left[\mathbf{J}_1^{-1/2}\mathbf{S}_1^*\right]^* \left[\mathbf{J}_1^{-1/2}\mathbf{S}_1^*\right], \qquad \mathbf{V}_1 = \mathbf{K}_{I_2} - \left[\mathbf{J}_1^{-1/2}\mathbf{D}_1^*\right]^* \left[\mathbf{J}_1^{-1/2}\mathbf{D}_1^*\right] \tag{A.3}$$

are simply integral operators corresponding to the Schur complements K_{I_1}/K_{J_1} and K_{I_2}/K_{J_1} found using the technique in Equation (A.2).

In the base case $m = 2$, we have from Theorem 1.6 that

$$\mathbf{R}_1 = \left[\mathbf{J}_1^{-1/2}\mathbf{S}_1^*\right]^* \left[\mathbf{J}_1^{-1/2}\mathbf{D}_1\right] + \mathbf{U}_1^{1/2}\Psi_1\mathbf{V}_1^{1/2} \tag{A.4}$$

Now for the induction case, assume that K is known over the region $(\cup_{j=1}^{p} I_j) \times (\cup_{j=1}^{p} I_j)$. Consider the 2-serrated domain given by

$$\left[(\cup_{j=1}^{p} I_j) \times (\cup_{j=1}^{p} I_j)\right] \cup (I_{p+1} \times I_{p+1}).$$

Then $(\cup_{j=1}^{p} I_j) \cap I_{p+1} = J_p$, $(\cup_{j=1}^{p} I_j) \setminus I_{p+1} = S_p$ and $I_{p+1} \setminus (\cup_{j=1}^{p} I_j) = D_p$. Repeating the above reasoning gives

$$\mathbf{R}_p = \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p\right] + \mathbf{U}_p^{1/2}\Psi_p\mathbf{V}_p^{1/2}$$

for some bounded linear map $\Psi_p : L^2(D_p) \to L^2(S_p)$ with $\|\Psi_p\| \leq 1$ where

$$\mathbf{U}_p = \mathbf{K}_{S_p} - \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{S}_p^*\right], \qquad \mathbf{V}_p = \mathbf{K}_{D_p} - \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p^*\right]^* \left[\mathbf{J}_p^{-1/2}\mathbf{D}_p^*\right]$$

The proof is this complete. □

A.2.5 Estimation of the Canonical Completion

Solving Equation (1.10) involves an interesting complication. Since K_Ω is specified inexactly, both the *operator*, which is $\mathbf{J}_p^{1/2}$, as well as the *data*, in the form of the operators $\mathbf{D}_p$ and $\mathbf{S}_p$, are inexactly specified. In essence, the problem is to estimate an

operator

$$\mathbf{W} = \sum_{j=1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j}$$

where $\{(\lambda_j, e_j)\}_{j=1}^{\infty}$ are the eigenpairs of $\mathbf{T}$, from the estimates $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$ and $\hat{\mathbf{T}}$ which converge to the operators $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$ almost surely or in L^2. A natural candidate is the estimator

$$\hat{\mathbf{W}} = \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\hat{\lambda}_j}$$

where N serves as the truncation or regularization parameter. We shall work out how fast N can grow as the estimates $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$ of $\hat{\mathbf{T}}$ converge to $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$ for $\hat{\mathbf{W}}$ to converge to $\mathbf{W}$. We have the following estimate:

Lemma A.4. *Let α_j be the eigenvalue gap given by*

$$\alpha_j = \begin{cases} (\lambda_1 - \lambda_2)/2\sqrt{2} & j = 1 \\ \Big[(\lambda_{j-1} - \lambda_j) \wedge (\lambda_j - \lambda_{j+1})\Big]/2\sqrt{2} & j > 1. \end{cases}$$

If α_j is monotonically decreasing with j, then for every N satisfying $\lambda_N > \|\mathbf{T} - \hat{\mathbf{T}}\|_2$, we have the bound

$$\begin{aligned} \|\mathbf{W} - \hat{\mathbf{W}}\|_2^2 \preceq \frac{N}{\lambda_N^2} \left[\|\mathbf{U} - \hat{\mathbf{U}}\|_2^2 + \|\mathbf{V} - \hat{\mathbf{V}}\|_2^2\right] \\ + \frac{N}{\lambda_N^2 \alpha_N^2} \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2 + \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2. \end{aligned} \tag{A.5}$$

Moreover, if α_j is not monotonically decreasing with j, the above bound still holds if we replace α_N with $\min_{j \le N} \alpha_j$,

Proof. We take a step wise approach. Define

$$\hat{\mathbf{W}}_1 = \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}e_j \otimes \hat{\mathbf{V}}e_j}{\lambda_j}, \hat{\mathbf{W}}_2 = \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\lambda_j} \text{ and } \hat{\mathbf{W}}_3 = \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\hat{\lambda}_j}.$$

Naturally,

$$\|\mathbf{W} - \hat{\mathbf{W}}\|_2^2 \le 2\|\mathbf{W} - \hat{\mathbf{W}}_1\|_2^2 + 2\|\hat{\mathbf{W}}_1 - \hat{\mathbf{W}}_2\|_2^2 + 2\|\hat{\mathbf{W}}_2 - \hat{\mathbf{W}}_3\|_2^2$$

We now proceed by working out an upper bound for every term individually:

Step 1. Using the identity $\|x + y\|^2 \le 2\|x\|^2 + 2\|y\|^2$ we can write the first term as

follows

$$\begin{aligned}
\|\mathbf{W} - \hat{\mathbf{W}}_1\|_2^2 &\leq 2 \left\| \sum_{j=1}^{N} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} - \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}e_j \otimes \hat{\mathbf{V}}e_j}{\lambda_j} \right\|_2^2 + 2 \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 \\
&= 2 \left\| \sum_{j=1}^{N} \frac{[\mathbf{U} - \hat{\mathbf{U}}]e_j \otimes \mathbf{V}e_j + \hat{\mathbf{U}}e_j \otimes [\mathbf{V} - \hat{\mathbf{V}}]e_j}{\lambda_j} \right\|_2^2 + 2 \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 \\
&\leq 2 \left\| \sum_{j=1}^{N} \frac{[\mathbf{U} - \hat{\mathbf{U}}]e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 + 2 \left\| \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}e_j \otimes [\mathbf{V} - \hat{\mathbf{V}}]e_j}{\lambda_j} \right\|_2^2 \\
&\quad + 2 \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 \\
&\leq 2 \left[\|\mathbf{V}\|_2^2 \|\mathbf{U} - \hat{\mathbf{U}}\|_2^2 + \|\hat{\mathbf{U}}\|_2^2 \|\mathbf{V} - \hat{\mathbf{V}}\|_2^2 \right] \sum_{j=1}^{N} \frac{1}{\lambda_j^2} + 2 \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 \\
&\leq 2 \frac{N}{\lambda_N^2} \left[\|\mathbf{V}\|_2^2 \|\mathbf{U} - \hat{\mathbf{U}}\|_2^2 + \|\hat{\mathbf{U}}\|_2^2 \|\mathbf{V} - \hat{\mathbf{V}}\|_2^2 \right] + 2 \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2
\end{aligned}$$

Step 2. In the same way, we can write the second term as

$$\begin{aligned}
\|\hat{\mathbf{W}}_1 - \hat{\mathbf{W}}_2\|_2^2 &= \left\| \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}e_j \otimes \hat{\mathbf{V}}e_j}{\lambda_j} - \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\lambda_j} \right\|_2^2 \\
&\leq \left\| \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}(e_j - \hat{e}_j) \otimes \hat{\mathbf{V}}e_j}{\lambda_j} \right\|_2^2 + \left\| \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}(e_j - \hat{e}_j)}{\lambda_j} \right\|_2^2 \\
&\leq 2 \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \cdot \sum_{j=1}^{N} \frac{\|e_j - \hat{e}_j\|^2}{\lambda_j^2} \\
&\preceq 2 \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2 \cdot \sum_{j=1}^{N} \frac{1}{\alpha_j^2 \lambda_j^2} \\
&\preceq \frac{N}{\alpha_N^2 \lambda_N^2} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2
\end{aligned}$$

The third inequality is a consequence of the perturbation bound for eigenfunctions which states that the perturbation $\|e_j - \hat{e}_j\|$ can be controlled by the perturbation $\|\mathbf{T} - \hat{\mathbf{T}}\|$ of $\mathbf{T}$ divided by the eigenvalue gap α_j. In the last inequality, we use the assumption that the eigenvalue gap α_j decreases with N. If this is not true we can simply replace α_N above with $\min_{j \leq N} \alpha_j$.

Step 3. And now the third term satisfies,

$$\begin{aligned}\|\hat{\mathbf{W}}_3 - \hat{\mathbf{W}}_2\|_2^2 &= \left\| \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\lambda_j} - \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\hat{\lambda}_j} \right\|_2^2 \\ &= \left\| \sum_{j=1}^{N} \hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j \hat{\lambda}_j} \right\|_2^2 \\ &\leq 2 \sum_{j=1}^{N} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \left[\frac{\hat{\lambda}_j - \lambda_j}{\lambda_j \hat{\lambda}_j} \right]^2 \\ &\leq 2 \frac{N}{\lambda_N^2 \hat{\lambda}_N^2} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2\end{aligned}$$

Here, we used the perturbation bound for eigenvalues which is given by

$$|\hat{\lambda}_j - \lambda_j| \leq \|\mathbf{T} - \hat{\mathbf{T}}\|.$$

Since, we choose N such that $\|\mathbf{T} - \hat{\mathbf{T}}\| \leq \lambda_N$, we can bound the $\hat{\lambda}_j$ in the denominator using the fact that $\hat{\lambda}_N \geq \lambda_N - \|\mathbf{T} - \hat{\mathbf{T}}\| > 0$ and write,

$$\begin{aligned}\left\| \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\lambda_j} - \sum_{j=1}^{N} \frac{\hat{\mathbf{U}}\hat{e}_j \otimes \hat{\mathbf{V}}\hat{e}_j}{\hat{\lambda}_j} \right\|_2^2 &\leq 2 \frac{N}{\lambda_N^2} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \frac{\|\mathbf{T} - \hat{\mathbf{T}}\|^2}{\left[\lambda_N - \|\mathbf{T} - \hat{\mathbf{T}}\|\right]^2} \\ &\preceq \frac{N}{\lambda_N^4} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2.\end{aligned}$$

The last inequality follows from the fact that

$$\frac{1}{\left[\lambda_N - \|\mathbf{T} - \hat{\mathbf{T}}\|\right]} = \frac{1}{\lambda_N} \left[1 + \frac{\|\mathbf{T} - \hat{\mathbf{T}}\|}{\lambda_N} + \frac{\|\mathbf{T} - \hat{\mathbf{T}}\|^2}{\lambda_N^2} + \cdots \right] \preceq \frac{1}{\lambda_N}.$$

Step 4. Putting everything together, we get

$$\begin{aligned}&\|\hat{\mathbf{W}}_1 - \mathbf{W}\|_2^2 + \|\hat{\mathbf{W}}_1 - \hat{\mathbf{W}}_2\|_2^2 + \|\hat{\mathbf{W}}_3 - \hat{\mathbf{W}}_2\|_2^2 \\ &\qquad \preceq \frac{N}{\lambda_N^2} \left[\|\mathbf{V}\|_2^2 \|\mathbf{U} - \hat{\mathbf{U}}\|_2^2 + \|\hat{\mathbf{U}}\|_2^2 \|\mathbf{V} - \hat{\mathbf{V}}\|_2^2 \right] + \frac{N}{\alpha_N^2 \lambda_N^2} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2 \\ &\qquad + \frac{N}{\lambda_N^4} \|\hat{\mathbf{U}}\|_2^2 \|\hat{\mathbf{V}}\|_2^2 \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2 + 2 \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 \\ &\qquad \preceq \frac{N}{\lambda_N^2} \left[\|\mathbf{U} - \hat{\mathbf{U}}\|_2^2 + \|\mathbf{V} - \hat{\mathbf{V}}\|_2^2 \right] + \frac{N}{\lambda_N^2 \alpha_N^2} \|\mathbf{T} - \hat{\mathbf{T}}\|_2^2 + \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2\end{aligned}$$

since $\alpha_j < \lambda_j$. Hence proved. □

Remark A.4. *The next result illustrates how the estimate from Lemma A.4 can be used to derive consistency and rates of convergence for our estimator. We begin by considering the case when the tuning parameters N_p are allowed to be random and then consider the case when they are required to be deterministic. Strictly speaking the results are independent and the impatient reader can skip them and go directly to Lemma A.6, but we believe that they are helpful in understanding the the proof of the rate of convergence result.*

Theorem A.3 (Consistency)**.** *Let K_Ω be a continuous partial covariance on a serrated domain Ω of m intervals, $\hat{K}_\Omega \in L^2(\Omega)$. If $\hat{K}_\Omega \to K_\Omega$ in $L^2(\Omega)$ and the regularization parameters $\mathbf{N} = (N_p)_{p=1}^{m-1}$ are chosen such that for $1 \leq p < m$, $\delta_p = \|\hat{\mathbf{S}}_p - \mathbf{S}_p\|_2 \vee \|\hat{\mathbf{D}}_p - \mathbf{D}_p\|_2$ and $\epsilon_p = \|\hat{\mathbf{J}}_p - \mathbf{J}_p\|_2$ we have*

1. *$N_p \to \infty$,*
2. *$\lambda_{p,N_p} > \epsilon_p$,*
3. *$\frac{N_p}{\lambda_{p,k}^2}\delta_p^2 \to 0$ and $\frac{N_p}{\lambda_{p,k}^2\alpha_{p,k}^2}\epsilon_p^2 \to 0$*

as $\delta_p, \epsilon_p \to 0$ where $\lambda_{p,k}$ denotes the kth eigenvalue of $\mathbf{J}_p$ and $\alpha_{p,k}$ is given by

$$\alpha_{p,k} = \begin{cases} (\lambda_{p,1} - \lambda_{p,2})/2\sqrt{2} & k = 1 \\ \Big[(\lambda_{p,k-1} - \lambda_{p,k}) \wedge (\lambda_{p,k} - \lambda_{p,k+1})\Big]/2\sqrt{2} & k > 1 \end{cases}$$

then $\hat{K}_\star \to K_\star$ in $L^2(I \times I)$.

Proof of Theorem A.3. We again proceed by induction on the number of intervals m. The claim is vacuously true for $m = 1$. Assume that it holds for $m = q - 1$ for some $q \geq 2$. We shall show that it holds for $m = q$.

Consider a partial covariance K_Ω on a serrated domain Ω of q intervals: $I_1, \ldots, I_q$. Let $I' = \cup_{j=1}^{q-1} I_j$ and $\Omega' = \cup_{j=1}^{q-1} I_j \times I_j$. Define $K_{\Omega'} = K_\Omega|_{\Omega'}$. Let $\epsilon = \int_\Omega [\hat{K}_\Omega(x,y) - K_\Omega(x,y)]^2 \, dxdy$. We can decompose the error of $\hat{K}$ as follows:

$$\begin{aligned} \int_{I\times I} \Big[\hat{K}_\star(x,y) - K_\star(x,y)\Big]^2 \, dx \, dy &= \int_{I'\times I'} \Big[\hat{K}_\star(x,y) - K_\star(x,y)\Big]^2 \, dx \, dy \\ &\quad + \int_{A_q} \Big[\hat{K}_\Omega(x,y) - K_\Omega(x,y)\Big]^2 \, dx \, dy \\ &\quad + 2\int_{R_q} \Big[\hat{K}_\star(x,y) - K_\star(x,y)\Big]^2 \, dx \, dy \end{aligned} \tag{A.6}$$

where $A_q = [I_q \times I_q] \setminus [J_{q-1} \times J_{q-1}]$. By construction, the estimator for the canonical extension of $K_{\Omega'}$ is the restriction $\hat{K}|_{I'\times I'}$ of the estimator $\tilde{K}$ for the canonical extension of K_Ω. Therefore, the first term in Equation A.6 converges to zero as $\epsilon \to 0$, by the

induction hypothesis. The same applies to the second term for more obvious reasons. It suffices to show that the third term

$$\int_{R_q} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dxdy = \|\hat{\mathbf{R}}_q - \mathbf{R}_q\|_2^2$$

converges to zero as $\epsilon \to 0$. Clearly, by Lemma A.4,

$$\|\hat{\mathbf{R}}_q - \mathbf{R}_q\|_2^2 \preceq \frac{N_q}{\lambda_{q,N_q}^2}\left[\|\mathbf{S}_q - \hat{\mathbf{S}}_q\|_2 \wedge \|\mathbf{D}_q - \hat{\mathbf{D}}_q\|_2\right]^2 + \frac{N_q}{\lambda_{q,N_q}^2\alpha_{q,N_q}^2}\|\mathbf{J}_q - \hat{\mathbf{J}}_q\|_2^2 + \left\|\sum_{j=N_q+1}^{\infty}\frac{\mathbf{S}_q e_{q,j}\otimes\mathbf{D}_q^* e_{q,j}}{\lambda_{q,j}}\right\|_2^2.$$

The first two terms converge to zero because N_q has been chosen such that $\frac{N_p}{\lambda_{p,N_p}^2\tilde{\alpha}_{p,N_p}^2}\|\hat{K}_\Omega - K_\Omega\|_{L^2(\Omega)}^2 \to 0$ which means that $\frac{N_q}{\lambda_{q,k}^2}\delta_q^2 \to 0$ and $\frac{N_q}{\lambda_{q,k}^2\alpha_{q,k}^2}\epsilon_q^2 \to 0$ and the last term converges to 0 as $N_q \to \infty$. The conclusion follows. □

Notice that Equation A.5 decomposes the error $\|\mathbf{W} - \hat{\mathbf{W}}\|_2^2$ into estimation and approximation terms as follows:

$$E_N = \frac{N}{\lambda_N^2}\left[\|\mathbf{U} - \hat{\mathbf{U}}\|_2^2 + \|\mathbf{V} - \hat{\mathbf{V}}\|_2^2\right] + \frac{N}{\lambda_N^2\alpha_N^2}\|\mathbf{T} - \hat{\mathbf{T}}\|_2^2$$

$$A_N = \left\|\sum_{j=N+1}^{\infty}\frac{\mathbf{U}e_j\otimes\mathbf{V}e_j}{\lambda_j}\right\|_2^2$$

Notice that A_N is a completely deterministic term which depends on $\mathbf{U}$, $\mathbf{V}$ and the spectral properties of $\mathbf{T}$. Furthermore, the error in $\mathbf{U}$ and $\mathbf{V}$ has much less weight than the error in $\mathbf{T}$.

Lemma A.5. *Under the setting of Lemma A.4, if $\lambda_N \sim N^{-\alpha}$ and $A_N \sim N^{-\beta}$, the best error is achieved for*

$$N \sim \delta^{-2/2\alpha+\beta+1} \wedge \epsilon^{-2/4\alpha+\beta+3}$$

and is given by

$$\|\mathbf{W} - \hat{\mathbf{W}}\|_2 \preceq N^{-\beta/2}$$

Proof. Since, $\lambda_N \sim N^{-\alpha}$, it follows that $\alpha_N \sim N^{-\alpha-1}$. Following the above discussion, we can write

$$\|\mathbf{W} - \hat{\mathbf{W}}\|_2^2 \preceq E_N + A_N$$

where

$$E_N = N^{2\alpha+1}\delta^2 + N^{4\alpha+3}\epsilon^2$$
$$A_N = N^{-\beta}$$

To minimize the error or in other wordsm maximize the decay rate, the estimation term should decrease at least as fast as the approximation term as N increases. Thus, $N^{2\alpha+1}\delta^2 \preceq N^{-\beta}$ and $N^{4\alpha+3}\epsilon^2 \preceq N^{-\beta}$. It follows that for the sum $E_N + A_N$ to decay at the maximum rate we need that

$$N \sim \delta^{-2/2\alpha+\beta+1} \wedge \epsilon^{-2/4\alpha+\beta+3}.$$

The total error then satisfies

$$\|\mathbf{W} - \hat{\mathbf{W}}\|_2^2 \preceq N^{-\beta}$$

and the conclusion follows. □

Theorem A.4 (Rate of Convergence)**.** *Let K_Ω be a partial covariance on a serrated domain Ω of m intervals and $\hat{K}_\Omega$ be its estimate. Let $\hat{K}$ be defined as above. Assume that for every $1 \leq p < m$, we have $\lambda_{p,k} \sim k^{-\alpha}$ and $A_{p,k} \sim k^{-\beta}$. If the truncation parameters $\mathbf{N} = (N_p)_{p=1}^{m-1}$ are chosen according to the rule*

$$N_p \sim \|\hat{K}_\Omega - K_\Omega\|_{L^2(\Omega)}^{-2\gamma_p/\beta}$$

then

$$\|\hat{K}_\star - K_\star\|_{L^2(I\times I)} \preceq \|\hat{K}_\Omega - K_\Omega\|_{L^2(\Omega)}^{\gamma_{m-1}}$$

where $\gamma_{m-1} = \frac{\beta}{4\alpha+\beta+3}\left[\frac{\beta}{2\alpha+\beta+1}\right]^{m-2}$ for $m > 1$ and 1 for $m = 1$.

Proof of Theorem A.4. We proceed by induction on the number of intervals m. The statement is vacuously true for the base case $m = 1$. Assume that it holds for $m = q - 1$ for some $q \geq 3$. We shall show that it must hold for $m = q$.

As before, consider a partial covariance K_Ω on a serrated domain Ω of q intervals: $I_1, \ldots, I_q$. Let $I' = \cup_{j=1}^{q-1} I_j$ and $\Omega' = \cup_{j=1}^{q-1} I_j \times I_j$. Define $K_{\Omega'} = K_\Omega|_{\Omega'}$. Let $\epsilon = \int_\Omega [\hat{K}_\Omega(x,y) - K_\Omega(x,y)]^2\, dxdy$. The error of $\hat{K}_\star$ can be decomposed as in Equation A.6.

By construction, the estimator for the canonical extension of $K_{\Omega'}$ is the restriction $\hat{K}_\star|_{I'\times I'}$ of the estimator $\hat{K}_\star$ for the canonical extension of K_Ω. By the induction hypothesis,

$$\int_{I'\times I'} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dxdy \lesssim \left[\int_{\Omega'} \left[\hat{K}_\Omega(x,y) - K_\Omega(x,y)\right]^2 dxdy\right]^{\gamma_{m-2}} \lesssim \epsilon^{\gamma_{m-2}}.$$

Thus the first term in Equation (A.6) can be bounded by a power of ϵ. The second term is obviously less than ϵ. We now turn our attention to the third term,

$$\int_{R_q} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dxdy = \|\hat{\mathbf{R}}_q - \mathbf{R}_q\|_2^2$$

Using Lemma A.5, we have

$$N_q \sim [\epsilon^{\gamma_{m-2}}]^{-2/2\alpha+\beta+1} \wedge \epsilon^{-2/4\alpha+\beta+1} = [\epsilon^{\gamma_{m-2}}]^{-2/2\alpha+\beta+1}$$

and,

$$\|\hat{\mathbf{R}}_q - \mathbf{R}_q\|_2^2 \preceq N_q^{-\beta} \sim [\epsilon^{\gamma_{m-2}}]^{\beta/2\alpha+\beta+1} = \epsilon^{\gamma_{m-1}}$$

From Equation A.6,

$$\int_{I\times I} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dx\, dy \preceq \epsilon^{\gamma_{m-2}} + \epsilon + 2\epsilon^{\gamma_{m-1}} \sim \epsilon^{\gamma_{m-1}} \tag{A.7}$$

and the proof is complete. □

We now present results which treat the tuning parameter as a deterministic quantity.

A.2.6 Consistency and Rates of Convergence with Nonrandom Tuning Parameter

Recall the setting of Lemma A.4 and let $\epsilon_n = \|\mathbf{T} - \hat{\mathbf{T}}_n\|_2$ and $\delta_n = \|\mathbf{U} - \hat{\mathbf{U}}_n\|_2 \vee \|\mathbf{V} - \hat{\mathbf{V}}_n\|_2$.

Lemma A.6. *Assume that $\lambda_N \sim N^{-\alpha}$, $A_N \sim N^{-\beta}$, $\epsilon_n = O_{\mathbb{P}}(n^{-\zeta})$ and $\delta_n = O_{\mathbb{P}}(n^{-\zeta'})$.*

1. *If the tuning parameter scales as $N \sim n^x$ where*

$$0 < x < \frac{\zeta}{2\alpha + 2/3} \wedge \frac{\zeta'}{\alpha + 1/2},$$

then $\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2 \to 0$ in probability as $n \to \infty$.

2. *For every $\varepsilon > 0$,*

$$\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2 = O_{\mathbb{P}}(1/n^{y_* - \varepsilon})$$

where

$$y_* = \frac{\beta\zeta}{\beta + 2\alpha + 3/2} \wedge \frac{\beta\zeta'}{\beta + \alpha + 1/2}$$

so long as the tuning parameter satisfies $N \sim n^{y_*/\beta}$.

Proof. Let $\{\eta_n\}_{n=1}^{\infty}$ be a sequence with $\eta_n > 0$. Then,

$$\begin{aligned}
&\mathbb{P}\{\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2^2 > \eta_n\} \\
&= \mathbb{P}\{\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2^2 > \eta_n \mid \lambda_N > \|\hat{\mathbf{T}}_n - \mathbf{T}\|\} \cdot \mathbb{P}\{\lambda_N > \|\hat{\mathbf{T}}_n - \mathbf{T}\|\} \\
&\qquad + \mathbb{P}\{\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2^2 > \eta_n \mid \lambda_N \leq \|\hat{\mathbf{T}}_n - \mathbf{T}\|\} \cdot \mathbb{P}\{\lambda_N \leq \|\hat{\mathbf{T}}_n - \mathbf{T}\|\} \\
&\leq \mathbb{P}\{\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2^2 > \eta_n \mid \lambda_N > \|\hat{\mathbf{T}}_n - \mathbf{T}\|\} \cdot 1 \\
&\qquad + 1 \cdot \mathbb{P}\{\lambda_N \leq \|\hat{\mathbf{T}}_n - \mathbf{T}\|\} \\
&\leq \mathbb{P}\left\{ \tfrac{N}{\lambda_N^2}\delta_n^2 + \tfrac{N}{\lambda_N^2\alpha_N^2}\epsilon_n^2 + \left\| \textstyle\sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 > C\eta_n \;\middle|\; \lambda_N > \epsilon_n \right\} + \mathbb{P}\{\epsilon_n \geq \lambda_N\} \\
&\leq \frac{\mathbb{P}\left\{ \tfrac{N}{\lambda_N^2}\delta_n^2 + \tfrac{N}{\lambda_N^2\alpha_N^2}\epsilon_n^2 + \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 > C\eta_n \right\}}{\mathbb{P}\{\lambda_N > \epsilon_n\}} + \mathbb{P}\{\epsilon_n \geq \lambda_N\} \\
&\leq \frac{\mathbb{P}\left\{ \tfrac{N}{\lambda_N^2}\delta_n^2 + \tfrac{N}{\lambda_N^2\alpha_N^2}\epsilon_n^2 + \left\| \sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 > C\eta_n \right\}}{1 - \mathbb{P}\{\epsilon_n \geq \lambda_N\}} + \mathbb{P}\{\epsilon_n \geq \lambda_N\}
\end{aligned}$$

It suffices for us to show that $\mathbb{P}\{\epsilon_n \geq \lambda_N\} \to 0$ and

$$\mathbb{P}\left\{ \tfrac{N}{\lambda_N^2}\delta_n^2 + \tfrac{N}{\lambda_N^2\alpha_N^2}\epsilon_n^2 + \left\| \textstyle\sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 > C\eta_n \right\} \to 0$$

as $n \to \infty$. Furthermore, since $\lambda_N \sim N^{-\alpha}$, we have $\alpha_N \sim N^{-\alpha-1}$ and we are given that

$$\left\| \textstyle\sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2 \sim N^{-\beta}.$$

Now,

$$\mathbb{P}\{\epsilon_n \geq \lambda_N\} \leq \mathbb{P}\{n^{\zeta}\epsilon_n \geq n^{\zeta}N^{-\alpha}\}$$

and

$$\begin{aligned}
&\mathbb{P}\left\{ \tfrac{N}{\lambda_N^2}\delta_n^2 + \tfrac{N}{\lambda_N^2\alpha_N^2}\epsilon_n^2 + \left\| \textstyle\sum_{j=N+1}^{\infty} \frac{\mathbf{U}e_j \otimes \mathbf{V}e_j}{\lambda_j} \right\|_2^2 > C\eta_n \right\} \\
&\leq \mathbb{P}\left\{N^{2\alpha+1}\delta_n^2 > C'\eta_n\right\} + \mathbb{P}\left\{N^{4\alpha+3}\epsilon_n^2 > C'\eta_n\right\} + \mathbb{P}\left\{N^{-2\beta} > C'\eta_n\right\} \\
&\leq \mathbb{P}\left\{ n^{2\zeta'}\delta_n^2 > C'\frac{n^{2\zeta'}\eta_n}{N^{2\alpha+1}} \right\} + \mathbb{P}\left\{ n^{2\zeta}\epsilon_n^2 > C'\frac{n^{2\zeta}\eta_n}{N^{4\alpha+3}} \right\} + \mathbb{P}\left\{ 1 > C'\frac{\eta_n}{N^{-2\beta}} \right\}
\end{aligned}$$

where $C' = C/3$.

Let $\eta_n \sim n^{-2y}$ and $N \sim n^x$. We need to show that there exists $y, x > 0$ such that the following terms increase with n:

$$\begin{aligned}
\frac{n^{\zeta}}{N^{\alpha}} &\sim n^{\zeta - \alpha x}, & \frac{n^{2\zeta'}\eta_n}{N^{2\alpha+1}} &\sim n^{2'\zeta' - 2y - (2\alpha+1)x}, \\
\frac{n^{2\zeta}\eta_n}{N^{4\alpha+3}} &\sim n^{2\zeta - 2y - (4\alpha+3)x}, & \frac{\eta_n}{N^{-2\beta}} &\sim n^{-2y + 2\beta x}.
\end{aligned}$$

It follows that:

$$\begin{aligned} \zeta - \alpha x &> 0 \\ 2\zeta' - 2y - (2\alpha + 1)x &> 0 \\ 2\zeta - 2y - (4\alpha + 3)x &> 0 \\ -2y + 2\beta x &> 0 \end{aligned} \tag{A.8}$$

Any pair (x, y) with $x, y > 0$ satisfying the inequalities (A.8), indicated by the yellow

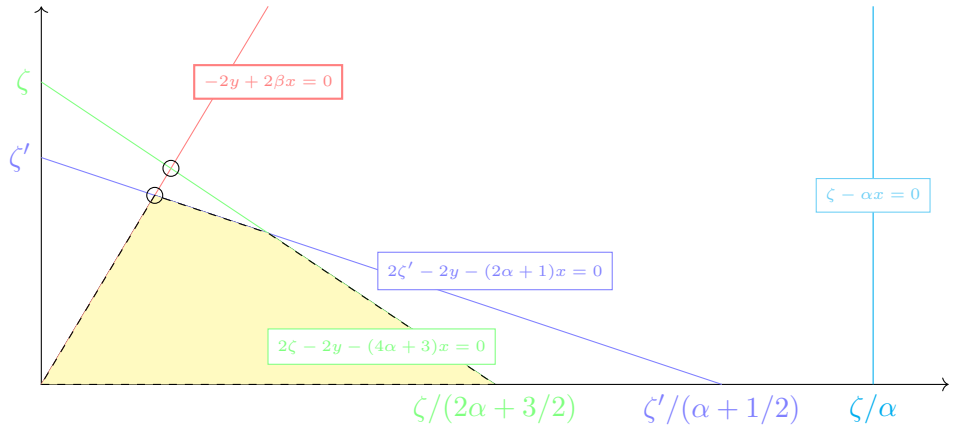


Figure A.8: The yellow region indicates the solutions (x, y) of the inequalities (A.8).

region in Figure A.8, corresponds to a consistent estimator for which the tuning parameter scales according to $N \sim n^x$ and the error decreases at least as fast as n^{-y} in probability. It thus follows that so long as x satisfies,

$$x < \frac{\zeta}{2\alpha + 2/3} \wedge \frac{\zeta'}{\alpha + 1/2}$$

we have $\|\mathbf{W} - \hat{\mathbf{W}}_n\| \to 0$ in probability as $n \to \infty$.

Of course, nothing prevents us from choosing an x for which we can have the highest possible value of y. From Figure A.8, it is clear that the supremum of y is given by

$$y_* = \frac{\beta\zeta}{\beta + 2\alpha + 3/2} \wedge \frac{\beta\zeta'}{\beta + \alpha + 1/2}$$

depending on which line among the blue and green ones intersects with the red line first and thus at a higher value of y. The corresponding value of x is given by $x_* = y_*/\beta$. Thus by choosing x to be x_*, we have that $\|\mathbf{W} - \hat{\mathbf{W}}_n\|_2 = O_{\mathbb{P}}(1/n^{y_* - \varepsilon})$ for every $\varepsilon > 0$. □

Proof of Theorem 1.8. In essence, we shall merely apply Lemma A.6 repeatedly. We proceed by induction on the number of intervals m. The statement is true for the base

case $m = 2$ due to Lemma A.6. Assume that it holds for $m = q - 1$ for some $q > 3$. We shall show that it must hold for $m = q$.

As before, consider a partial covariance K_Ω on a serrated domain Ω of q intervals: $I_1, \ldots, I_q$. Let $I' = \cup_{j=1}^{q-1} I_j$ and $\Omega' = \cup_{j=1}^{q-1} I_j \times I_j$. Define $K_{\Omega'} = K_\Omega|_{\Omega'}$. Let $\epsilon = \int_\Omega [\hat{K}_\Omega(x,y) - K_\Omega(x,y)]^2 \, dxdy$. The error of $\hat{K}_\star$ can be decomposed as in Equation A.6.

By construction, the estimator for the canonical extension of $K_{\Omega'}$ is the restriction $\hat{K}_\star|_{I' \times I'}$ of the estimator $\hat{K}_\star$ for the canonical extension of K_Ω. By the induction hypothesis,

$$\int_{\Omega'} \left[\hat{K}_\Omega(x,y) - K_\Omega(x,y)\right]^2 dxdy = O_{\mathbb{P}}(n^{-\zeta})$$

implies that for every $\varepsilon > 0$

$$\int_{I' \times I'} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dxdy = O_{\mathbb{P}}(1/n^{\zeta\gamma_{m-2}-\varepsilon}).$$

Thus the first term in Equation (A.6) can be bounded in probability by a power of $n^{-\zeta}$. This means that in the language of Lemma A.6, $\delta_n = O_{\mathbb{P}}(1/n^{\zeta\gamma_{m-2}-\varepsilon})$ and thus $\zeta' = \zeta\gamma_{m-2} - \varepsilon$. It is given that the second term is $O_{\mathbb{P}}(n^{-\zeta})$. We now turn our attention to the third term,

$$\int_{R_q} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dxdy \; = \|\hat{\mathbf{R}}_q - \mathbf{R}_q\|_2^2$$

Using Lemma A.6, if the tuning parameter satisfies $N_q \sim n^{\gamma_{m-1}/\beta}$, we have

$$\begin{aligned}
\|\hat{\mathbf{R}}_q - \mathbf{R}_q\|_2 &= O_{\mathbb{P}}\left(1/n^{\frac{\beta\zeta}{\beta+2\alpha+3/2} \wedge \frac{\beta\zeta'}{\beta+\alpha+1/2}}\right) \\
&= O_{\mathbb{P}}\left(1/n^{\frac{\beta\zeta}{\beta+2\alpha+3/2} \wedge \frac{\beta(\zeta\gamma_{m-2}-\varepsilon)}{\beta+\alpha+1/2}}\right) \\
&= O_{\mathbb{P}}\left(1/n^{\zeta\gamma_{m-1}-\varepsilon'}\right)
\end{aligned}$$

where $\varepsilon' > 0$ can be arbitrarily small. From Equation A.6,

$$\begin{aligned}
&\int_{I \times I} \left[\hat{K}_\star(x,y) - K_\star(x,y)\right]^2 dx \; dy \\
&= O_{\mathbb{P}}(1/n^{\zeta\gamma_{m-2}-\varepsilon}) + O_{\mathbb{P}}(1/n^{\zeta}) + O_{\mathbb{P}}(1/n^{\zeta\gamma_{m-1}-\varepsilon'}) = O_{\mathbb{P}}(1/n^{\zeta\gamma_{m-1}-\varepsilon'})
\end{aligned} \tag{A.9}$$

and the proof is complete. □

Remark A.5. *It is indeed possible to give a general consistency result like Lemma A.6(i) but for the m-serrated domain with $m > 2$ using Lemma A.6 as before, however this proves to be a tedious exercise which doesn't tell us significantly more than what we already know from Figure A.8 and Theorem 1.8. Hence, we shall skip it.*

A.2.7 Beyond Serrated Domains

Proof of Theorem 1.9. Let K_1 and K_2 be completions of $K_{\widetilde{\Omega}}$ and assume that $K_{\widetilde{\Omega}}|_\Omega$ admits a unique completion. Then K_1 and K_2 are completions of $K_{\widetilde{\Omega}}|_\Omega$, implying that $K_1 = K_2$. □

Proof of Theorem 1.10. Let $s, t \in \widetilde{\Omega}^c$ separated by $J \subset I$ in $(I, \widetilde{\Omega})$. Let $J_- = \{u \in I : u \leq v \text{ for some } v \in J\}$ and $J_+ = \{u \in I : u \geq v \text{ for some } v \in J\}$. Define

$$\bar{\Omega} = (J_- \times J_-) \cup (J_+ \times J_+)$$

If K is the unique completion of $K_{\widetilde{\Omega}}$ then it is a unique completion of the partial covariance $K|_{\bar{\Omega}}$ on the serrated domain Ω. The canonical completion of $K|_{\bar{\Omega}}$ would thus have to be same as the unique completion K and therefore,

$$K(s,t) = \langle k_{s,J}, k_{t,J} \rangle$$

It follows that K is the canonical completion of $K_{\widetilde{\Omega}}$.

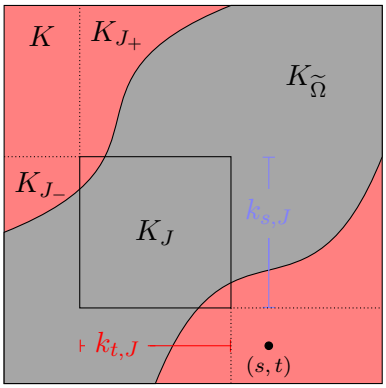


Figure A.9: The partial covariance $K_{\widetilde{\Omega}}$

□

Proof of Theorem 1.11. Let $\widetilde{\Omega} \subset \Omega$. Then, every pair $(s,t) \in \Omega^c$ separated by $J \subset I$ in Ω is also separated by $J \subset I$ in $\widetilde{\Omega}$. Therefore the canonical completion of $K_\star|_\Omega$ is equal to $K_\star(s,t)$. Since this is true for every $(s,t) \in \Omega^c$, it follows that the canonical completion of $K_\star|_\Omega$ is $K_\star$.

For the converse, if the canonical completion of $K_\star|_\Omega$ is $K_\star$ then $K_\star$ is the completion of a partial covariance on the nearly serrated domain Ω which is a improper subset of Ω. □

A.3 Some Additional Remarks on Uniqueness

Theorem 1.5 provides a condition that is both sufficient and necessary for a partial covariance on a serrated domain to admit a unique extension. It is interesting to briefly discuss this condition and contrast it to analyticity and/or low-rank conditions, that have previously been employed as means to guarantee uniqueness.

Intuitively, the condition in Theorem 1.5 is a "rigidity" condition: it specifies that that the whole process can be generated by some deterministic transformation of a certain part of it. Analyticity is a special case of "rigidity", where this deterministic transformation is manifested as analytic continuation. But it is a very special case, and the condition in Theorem 1.5 is substantially weaker. Indeed, it makes no reference to smoothness and can be satisfied by non-differentiable covariances. Such examples can be generated readily, based on the theorem's perfect linear prediction interpretation. The following example illustrates the essence of the general case. Let $I_1 = [0, 2/3]$, $I_2 = [1/3, 1]$ and $\Omega = I_1^2 \cup I_2^2$ be a 2-serrated domain. Let $\{B(t) : t \in [1/3, 2/3]\}$ be a standard Brownian motion on $I_1 \cap I_2$. Define

$$X(t) = \begin{cases} B(2/3 - t), & t \in I_1 \setminus I_2 \\ B(t), & t \in I_1 \cap I_2, \\ B(4/3 - t), & t \in I_2 \setminus I_1. \end{cases} \tag{A.10}$$

to be a process on $[0, 1]$ (Figure A.10 depicts some sample paths thereof). Since $X|_{I_1 \setminus I_2}$ and $X|_{I_2 \setminus I_1}$ can be perfectly linearly predicted from $X|_{I_1 \cap I_2}$, the covariance K of X is the unique extension of its restriction K_Ω to the domain Ω. The process X is "rigid", in the sense that its global fluctuations are a deterministic propagation of the local fluctuations on $I_1 \cap I_2$. Yet its covariance is far from analytic – indeed it is not even differentiable.

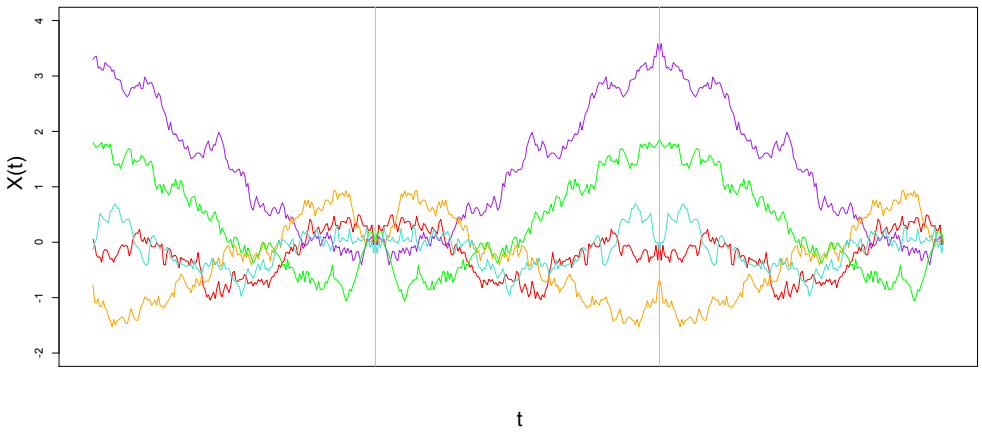


Figure A.10: Five sample paths corresponding to the process defined in Equation (A.10).

Another way to induce rigidity is via rank constraints, e.g. assuming that the partial covariance K_Ω is the restriction of a finite rank covariance $K(s,t) = \sum_{i=1}^{d} \lambda_i \varphi(s)\varphi(t)$ on $[0,1]$ to the serrated domain $\Omega = \cup_{j=1}^{m} I_j \times I_j$. Equivalently, one can assume that each $K_\Omega|_{I_j \times I_j}$ has finite rank, i.e. place the assumption directly on K_Ω, because having finite rank *locally* is equivalent to having finite rank *globally*[1]. No additional smoothness assumptions are made. In this context, Theorem 1.5 can be reduced to a statement purely about ranks. To do so, for $S \subset I$ such that $S \times S \subset \Omega$, let d_S denote the dimension of $\mathcal{H}(K_S)$ or equivalently, the rank of the covariance $K_S = K_\Omega|_{S \times S}$. Then we have:

Theorem A.5 (The finite rank case)**.** *Let K_Ω be a partial covariance over a serrated domain $\Omega = \cup_{j=1}^{m} I_p \times I_p \subset [0,1]^2$ such that $d_{I_p} < \infty$, $1 \leq p \leq m$. Then, K_Ω extends uniquely to a covariance on $[0,1]^2$ if and only if there exists an $r \in \{1, ..., m\}$ such that $d_{I_p} = d_{J_p}$ for $1 \leq p < r$ and $d_{I_{q+1}} = d_{J_q}$ for $r \leq q < m$.*

Proof. Notice that $K_{I_p}/K_{J_p} = 0$ if and only if H_{J_p} is dense in H_{I_p} which is if and only if $d_{I_p} = \dim H_{I_p} = \dim H_{J_p} = d_{J_p}$. The statement then follows from Theorem 1.5. □

Remark A.6 (Finite ranks and nearly serrated domains)**.** *In fact, it is straightforward to see that the last proof can establish that the same condition remains* sufficient *for unique extension from a* nearly *serrated domain, in the following sense: if the condition is satisfied for some serrated subdomain Ω of a nearly serrated domain $\widetilde{\Omega}$, then the partial covariance $K_{\widetilde{\Omega}}$ extends uniquely. This is weaker than the condition in Delaigle et al. (2021), where it was assumed that $[0,1]$ can be partitioned into intervals $\{J'_k\}_{k=1}^{M}$ such that $J'_k \times J'_k \subset \widetilde{\Omega}$ and the corresponding restrictions of the eigenfunctions $\{\varphi_j|_{J'_k}\}_{j=1}^{d}$ are linearly independent for every k (or more simply that $K_{J'_k}$ is a rank-d covariance for every k).*

[1] If K is an extension of K_Ω, then every $f \in \mathcal{H}(K)$ can be written as a sum of elements $f_j \in H_{I_j}$ for $1 \leq j \leq m$. But H_{I_j} are all finite dimensional, so f_j can be written as a sum of finite number of generators k_u for $u \in I_j$. Therefore, f can be represented as a finite sum of the generators. These generators span $\mathcal{H}(K)$, so it follows that $\mathcal{H}(K)$ is finite dimensional and hence K must be a finite rank covariance.

B Appendix B. Positive-Definite Completions

This appendix collects the the proofs of the formal statements in Chapter 2.

Proof of Lemma 2.1. Let $H : X \times X \to \mathbb{R}$ be such that $K + H, K - H \geq O$. Thus, $K + H, K - H \geq O$ which implies that $H(x, y) = H(y, x)$ for $x, y \in X$ and

$$-\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \leq \sum_{i,j=1}^{n} \alpha_i \alpha_j H(x_i, x_j) \leq \sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j).$$

for $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$ and $\{x_i\}_{i=1}^n \subset X$. Let $\mathcal{H}$ be the RKHS of K and $\mathcal{H}_0 = \text{Span}\{k_x : x \in X\}\mathcal{H}$ where $k_x : X \to \mathbb{R}$ is defined by $k_x(y) = K(x, y)$ for $x, y \in X$. Let $B : \mathcal{H}_0 \times \mathcal{H}_0 \to \mathbb{R}$ be the symmetric bilinear linear functional given by $B(k_x, k_y) = H(x, y)$. B is well-defined because of the above equation. Moreover, $|B(f, f)| \leq \|f\|^2$ for every $f \in \mathcal{H}_0$. So,

$$\|f\|^2 + B(f, f), \|f\|^2 - B(f, f) \geq 0$$

Notice that $\|f - g\|^2 + B(f - g, f - g) \geq 0$ implies that

$$\begin{aligned} B(f, g) &\leq \tfrac{1}{2}\left[\|f - g\|^2 + B(f, f) + B(g, g)\right] \\ &\leq \|f\|^2 + \|g\|^2 - \langle f, g\rangle \end{aligned}$$

Replacing f by $\sqrt{c}f$ and g by $g/\sqrt{c}$ for some $c > 0$ gives

$$\begin{aligned} B(f, g) &\leq c\|f\|^2 + \|g\|^2/c - \langle f, g\rangle \\ &\leq 2\|f\|\|g\| + \|f\|\|g\| = 3\|f\|\|g\| \end{aligned}$$

by choosing $c = \|g\|/\|f\|$ and applying the Cauchy-Schwarz inequality. By replacing g by $-g$, we can derive $B(f,g) \geq -3\|f\|\|g\|$. It follows that $|B(f,g)| \leq 3\|f\|\|g\|$ and therefore B is continuous. It uniquely extends by continuity to $\mathcal{H} \times \mathcal{H}$ and admits a Riesz representation (Kreyszig, 1978, Theorem 3.8-4) of the form $B(f,g) = \langle \Phi_H f, g \rangle$, where $\Phi_H \in \mathcal{L}(\mathcal{H})$. Moreover, Φ_H is self-adjoint since

$$\langle \Phi_H k_x, k_y \rangle = H(x,y) = H(y,x) = \langle \Phi_H k_y, k_x \rangle$$

for $x, y \in X$. By (Conway, 2019, 2.13 Proposition), it follows that

$$\|\Phi_H\|^2 = \sup_{f \in \mathcal{H}_0 \setminus \{0\}} \frac{|\langle \Phi_H f, f \rangle|}{\|f\|^2} \leq 1$$

because $|\langle \Phi_H f, f \rangle| = |B(f,f)| \leq \|f\|^2$ for $f \in \mathcal{H}_0$. Thus, Φ_H is a self-adjoint contraction. □

C Appendix C. Continuously Indexed Graphical Models

This appendix collects the the proofs of the formal statements in Chapter 3.

C.1 Graphical Regularization

C.1.1 Approximate Inverse Zero Characterization

Proof of Theorem 3.2. By Theorem 2.2.3 of Bakonyi and Woerdeman (2011), $\mathbf{P}_{ij} = \mathbf{0}$ is equivalent to saying that

$$\mathbf{R}_{ij} = [\mathbf{R}_{ik}]_{k\in S}^{\top}[\mathbf{R}_{kl}]_{k,l\in S}^{-1}[\mathbf{R}_{lj}]_{l\in S} \tag{C.1}$$

for $S = \{m : m \neq i, j\}$. Through appropriate manipulations, this can be used to show that

$$\mathbf{K}_{ij} = \left([\mathbf{K}_{kl}]_{k,l\in S}^{-1/2}[\mathbf{K}_{ki}]_{k\in S}\right)^{\top}\left([\mathbf{K}_{kl}]_{k,l\in S}^{-1/2}[\mathbf{K}_{lj}]_{l\in S}\right). \tag{C.2}$$

By Theorem 11.18. of Paulsen and Raghupathi (2016), the above equality can be rewritten as

$$K(s,t) = \langle K(s,\cdot), K(\cdot,t)\rangle_{\mathcal{H}(V)} \tag{C.3}$$

for $s \in U_i$, $t \in U_j$ and $V = \cup_{k\in S} U_k$. It follows that $\Omega_X \subset (U_i \cup V)^2 \cup (V \cup U_j)^2$ or more simply, that Ω_X and $U_i \times U_j$ are disjoint. Thus implying that $U_i \times U_j$ and $\tilde{\Omega}_X^{\pi}$ are disjoint.

The converse requires more work. Assume that $U_i \times U_j$ and $\tilde{\Omega}_X^{\pi}$ are disjoint. Now, if $x = (s,t)$ is in the closure of $U_i \times U_j$, there exists some closed $\Omega \supset \Omega_X$ for which (C.3) holds and $x \in \Omega^c$. It follows that there is an open ball B_x centered at x such that

$B_x \subset \Omega_c$. The closure of $U_i \times U_j$ is contained in $\cup_x B_x$, and by compactness there exists a finite subcover $\cup_{i=1}^{q} B_{x_i}$. We now show that there exists a partition π' of U such that every pixel associated with π' lies in one of the balls B_{x_i}.

Define the function $d : U_i \times U_j \to \mathbb{R}_+$ as

$$d(x) = \max\{d(x, B_{x_i}^c) : x \in B_{x_i}\}.$$

Alternatvely, d maps every x to the maximum of its distance from the set $B_{x_i}^c$ for every i such that $x \in B_{x_i}$. Observe that $R = \inf_x d(x) > 0$. So long as we partition U such that every pixel $U'_k \times U'_l$ satisfies that the maximum distance between two points in it is less than $R/2$, every pixel will be contained entirely in one of the balls B_{x_i}.

The precision operator $\mathbf{P}' = \mathbf{P}_{\pi'}$ corresponding to this new partition π' satisfies $\mathbf{P}'_{i'j'} = \mathbf{0}$ for every i', j' corresponding to a pixel contained in in the closure of $U_i \times U_j$. Since such operators $\mathbf{P}'_{i'j'}$ can be considered together as an operator, we can write the π'-analogue of (C.1) and work our way to (C.3) using appropriate manipulations. But (C.3) is partition independent, we can work our way backwards, this time for π instead of π' and derive that $\mathbf{P}_{ij} = \mathbf{0}$. Hence proved. □

C.1.2 Identifiability

Proof of Corollary 3.1. The first part is a tautology. For the second part, notice that for some $\epsilon_\pi > 0$, we can write with a slight abuse of notation that the set $\cap_{\epsilon>0}(\Omega_X + \mathbb{B}_\epsilon)^\pi$ is equal to $(\Omega_X + \mathbb{B}_\epsilon)^\pi$ if $\epsilon < \epsilon_\pi$. Thus for $\epsilon < \epsilon_{\pi_1} \wedge \epsilon_{\pi_2}$ we have

$$\begin{aligned}\left[\cap_{\epsilon>0}(\Omega_X + \mathbb{B}_\epsilon)^{\pi_1}\right] \cap \left[\cap_{\epsilon>0}(\Omega_X + \mathbb{B}_\epsilon)^{\pi_2}\right] &= (\Omega_X + \mathbb{B}_\epsilon)^{\pi_1} \cap (\Omega_X + \mathbb{B}_\epsilon)^{\pi_2} \\ &= (\Omega_X + \mathbb{B}_\epsilon)^{\pi_1 \wedge \pi_2} \\ &= \cap_{\epsilon>0}(\Omega_X + \mathbb{B}_\epsilon)^{\pi_1 \wedge \pi_2}.\end{aligned}$$

It follows that $\cap_{j=1}^{\infty} \tilde{\Omega}_X^{\pi_j} = \lim_{k\to\infty} \tilde{\Omega}_X^{\wedge_{j=1}^{k} \pi_j}$. If $(u, v) \in U \times U$ is not contained in the closure of Ω_X, then for a small enough $\delta > 0$ the δ-ball $(u, v) + \mathbb{B}_\delta$ does not intersect with the closure of Ω. For a sufficiently large k, there will be a pixel induced by $\wedge_{j=1}^{k} \pi_j$ containing (u, v) and which is itself contained in the δ-ball, for otherwise this would imply that the partitions do not separate points. For a small enough $\epsilon > 0$, this pixel will not be included in $(\Omega_X + \mathbb{B}_\epsilon)^{\wedge_{j=1}^{k} \pi_j}$. It can be worked out from the zero entries of the operator matrices $\mathbf{P}_{\pi_j}$ for $1 \le j \le k$ that this pixel and hence the point is indeed not contained in the closure of Ω_X. Similarly, if (u, v) is in the closure of Ω_X we can show that no pixel containing it will ever be rejected by a finite number of precision operator matrices $\mathbf{P}_j$. This establishes the claim. □

C.2 Estimation of the Precision Operator Matrix

C.2.1 Correlation Operator Matrix

Proof of Theorem 3.2. We decompose the difference $\hat{\mathbf{R}} - \mathbf{R}$ into approximation and estimation terms as follows

$$\hat{\mathbf{R}} - \mathbf{R} = \hat{\mathbf{R}} - \mathbf{R}_e + \mathbf{R}_e - \mathbf{R}$$

where $\mathbf{R}_e = \mathbf{I} + [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}\mathbf{K}_0[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}$. By Lemma C.1 and C.3 it follows that

$$\|\hat{\mathbf{R}} - \mathbf{R}\| \leq 5\|\mathbf{R}\| \left[\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|^2}{\epsilon^2} + \frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} \right] + 2\epsilon^\beta \cdot \|\Phi_0\| \cdot \|\mathbf{K}\|^\beta$$

Choosing $\epsilon = \|\hat{\mathbf{K}} - \mathbf{K}\|^{\frac{1}{\beta+1}}$ gives

$$\|\hat{\mathbf{R}} - \mathbf{R}\| \leq 10(\|\mathbf{R}\| \vee \|\Phi_0\|\|\mathbf{K}\|^\beta) \cdot \|\hat{\mathbf{K}} - \mathbf{K}\|^{\frac{\beta}{\beta+1}}$$

Similarly, for the case $\beta > 1$, we can choose $\epsilon = \|\hat{\mathbf{K}} - \mathbf{K}\|^{\frac{1}{2}}$ and argue likewise to conclude that

$$\|\hat{\mathbf{R}} - \mathbf{R}\| \leq 10(\|\mathbf{R}\| \vee \|\Phi_0\|\|\mathbf{K}\|^{2\beta-1}) \cdot \|\hat{\mathbf{K}} - \mathbf{K}\|^{\frac{1}{2}}.$$

□

Lemma C.1. *We have*

$$\|\hat{\mathbf{R}} - \mathbf{R}_e\| \leq 5\|\mathbf{R}\| \left[\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|^2}{\epsilon^2} + \frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} \right]$$

Proof. The following equation can be verified with some calculation.

$$\begin{aligned}
\hat{\mathbf{R}} - \mathbf{R}_e &= \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \right] [\hat{\mathbf{K}}_0 - \mathbf{K}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} \\
&+ \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \right] \mathbf{K}_0 \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \right] \\
&+ \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \right] \mathbf{K}_0 [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \\
&+ [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} [\hat{\mathbf{K}}_0 - \mathbf{K}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} \\
&+ [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \mathbf{K}_0 \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \right]
\end{aligned}$$

Using $\mathbf{K} = [\operatorname{dg}\mathbf{K}]^{1/2}\mathbf{R}[\operatorname{dg}\mathbf{K}]^{1/2}$ we can write this expansion as

$$\begin{aligned}
&= \mathbf{D}[\hat{\mathbf{K}}_0 - \mathbf{K}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} + \mathbf{A}\mathbf{R}_0\mathbf{A}^* + \mathbf{A}\mathbf{R}_0[\operatorname{dg}\mathbf{K}]^{1/2}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \\
&+ [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}[\hat{\mathbf{K}}_0 - \mathbf{K}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} + [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}[\operatorname{dg}\mathbf{K}]^{1/2}\mathbf{R}_0\mathbf{A}^*
\end{aligned}$$

where

$$\begin{aligned}\mathbf{R}_0 &= \mathbf{R} - \mathbf{I} \\ \mathbf{D} &= [\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \\ \mathbf{A} &= \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{K}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}\right][\operatorname{dg}\mathbf{K}]^{1/2}.\end{aligned}$$

So,

$$\begin{aligned}\|\hat{\mathbf{R}} - \mathbf{R}_e\| &\leq \|\mathbf{D}\| \cdot \|\hat{\mathbf{K}}_0 - \mathbf{K}_0\| \cdot \frac{1}{\sqrt{\epsilon}} + \|\mathbf{A}\| \cdot \|\mathbf{R}_0\| \cdot \|\mathbf{A}\| + \|\mathbf{A}\| \cdot \|\mathbf{R}_0\| \cdot 1 \\ &+ \frac{1}{\sqrt{\epsilon}} \cdot \|\hat{\mathbf{K}}_0 - \mathbf{K}_0\| \cdot \frac{1}{\sqrt{\epsilon}} + 1 \cdot \|\mathbf{R}_0\| \cdot \|\mathbf{A}\|.\end{aligned}$$

Applying Lemma C.2 to $\hat{\mathbf{A}} = \operatorname{dg}\hat{\mathbf{K}}$ and $\mathbf{A} = \operatorname{dg}\mathbf{K}$, we derive

$$\|\mathbf{D}\| \leq \|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|/\epsilon^{3/2} \qquad \text{and} \qquad \|\mathbf{A}\| \leq \|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|/\epsilon.$$

Using the simple observation that

$$\begin{aligned}\|\operatorname{dg}\mathbf{A}\| &= \max_i \|\mathbf{A}_{ii}\| \leq \|\mathbf{A}\| \\ \|\mathbf{A}_0\| &= \|\mathbf{A} - \operatorname{dg}\mathbf{A}\| \leq \|\mathbf{A}\| + \|\operatorname{dg}\mathbf{A}\| \leq 2\|\mathbf{A}\|\end{aligned}$$

we can write

$$\begin{aligned}\|\hat{\mathbf{R}} - \mathbf{R}_e\| &\leq \frac{\|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|\|\hat{\mathbf{K}}_0 - \mathbf{K}_0\|}{\epsilon^2} + \|\mathbf{R}_0\|\frac{\|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|^2}{\epsilon^2} \\ &+ \|\mathbf{R}_0\|\frac{\|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|}{\epsilon} + \frac{\|\hat{\mathbf{K}}_0 - \mathbf{K}_0\|}{\epsilon} + \|\mathbf{R}_0\|\frac{\|\operatorname{dg}\hat{\mathbf{K}} - \operatorname{dg}\mathbf{K}\|}{\epsilon} \\ &\leq \frac{\|\hat{\mathbf{K}} - \mathbf{K}\|^2}{\epsilon^2} + \|\mathbf{R}_0\|\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|^2}{\epsilon^2} \\ &+ \|\mathbf{R}_0\|\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} + \frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} + \|\mathbf{R}_0\|\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} \\ &\leq (2\|\mathbf{R}_0\| + 1)\left[\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} + \frac{\|\hat{\mathbf{K}} - \mathbf{K}\|^2}{\epsilon^2}\right] \\ &\leq 5\|\mathbf{R}\|\left[\frac{\|\hat{\mathbf{K}} - \mathbf{K}\|}{\epsilon} + \frac{\|\hat{\mathbf{K}} - \mathbf{K}\|^2}{\epsilon^2}\right]\end{aligned}$$

since $\|\mathbf{R}_0\| = \|\mathbf{R} - \mathbf{I}\| \leq \|\mathbf{R}\| + 1$ and $\|\mathbf{R}\| \geq 1$. Hence proved. □

Lemma C.2. *If* $\hat{\mathbf{A}}$ *is positive, then*

$$\begin{aligned}\|[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2}\| &\leq \|\hat{\mathbf{A}} - \mathbf{A}\|/\epsilon^{3/2} \\ \left\|\left[[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2}\right]\mathbf{A}^{1/2}\right\| &\leq \|\hat{\mathbf{A}} - \mathbf{A}\|/\epsilon\end{aligned}$$

Proof. Notice that

$$\begin{aligned}
&[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{-1/2}-[\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\\
&=[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{-1/2}\left[[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{1/2}-[\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\right][\epsilon\mathbf{I}+\mathbf{A}]^{1/2}\\
&=[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{-1/2}\left[[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{1/2}+[\epsilon\mathbf{I}+\mathbf{A}]^{1/2}\right]^{-1}\left[[\epsilon\mathbf{I}+\hat{\mathbf{A}}]-[\epsilon\mathbf{I}+\mathbf{A}]\right][\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\\
&=\left[\epsilon\mathbf{I}+\hat{\mathbf{A}}+[\epsilon\mathbf{I}+\mathbf{A}]^{1/2}[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{1/2}\right]^{-1}[\hat{\mathbf{A}}-\mathbf{A}][\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}
\end{aligned}$$

Since $\hat{\mathbf{A}}+[\epsilon\mathbf{I}+\mathbf{A}]^{1/2}[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{1/2}$ is positive, we can write

$$\begin{aligned}
&\|[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{-1/2}-[\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\|\\
&\leq\left\|\left[\epsilon\mathbf{I}+\hat{\mathbf{A}}+[\epsilon\mathbf{I}+\mathbf{A}]^{1/2}[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{1/2}\right]^{-1}\right\|\cdot\|\hat{\mathbf{A}}-\mathbf{A}\|\cdot\|[\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\|\\
&\leq\frac{1}{\epsilon}\cdot\|\hat{\mathbf{A}}-\mathbf{A}\|\cdot\frac{1}{\epsilon^{1/2}}
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\left\|\left[[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{-1/2}-[\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\right]\mathbf{A}^{1/2}\right\|\\
&\leq\left\|\left[\epsilon\mathbf{I}+\hat{\mathbf{A}}+[\epsilon\mathbf{I}+\mathbf{A}]^{1/2}[\epsilon\mathbf{I}+\hat{\mathbf{A}}]^{1/2}\right]^{-1}\right\|\cdot\|\hat{\mathbf{A}}-\mathbf{A}\|\cdot\|[\epsilon\mathbf{I}+\mathbf{A}]^{-1/2}\mathbf{A}^{1/2}\|\\
&\leq\frac{1}{\epsilon}\cdot\|\hat{\mathbf{A}}-\mathbf{A}\|\cdot 1.
\end{aligned}$$

Hence proved. □

Now, we shall find an upper bound for the approximation error under a regularity condition.

Lemma C.3. *If* $\mathbf{R}_0=[\mathrm{dg}\,\mathbf{K}]^\beta\Phi_0[\mathrm{dg}\,\mathbf{K}]^\beta$ *for some bounded operator matrix* Φ_0 *with the diagonal entries all zero and* $\beta>0$*, then*

$$\|\mathbf{R}_e-\mathbf{R}\|\leq\begin{cases}2\epsilon^\beta\cdot\|\Phi_0\|\cdot\|\mathbf{K}\|^\beta & 0<\beta\leq 1\\ 2\epsilon\ \cdot\|\Phi_0\|\cdot\|\mathbf{K}\|^{2\beta-1} & 1<\beta<\infty\end{cases}$$

Proof. We decompose the difference as follows:

$$\begin{aligned}
\mathbf{R}_e - \mathbf{R} &= [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}[\operatorname{dg}\mathbf{K}]^{1/2}\mathbf{R}_0[\operatorname{dg}\mathbf{K}]^{1/2}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - \mathbf{R}_0 \\
&= \Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big][\operatorname{dg}\mathbf{K}]^{1/2}\mathbf{R}_0[\operatorname{dg}\mathbf{K}]^{1/2}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \\
&\quad + \quad \mathbf{R}_0[\operatorname{dg}\mathbf{K}]^{1/2}\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big] \\
&= \Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big][\operatorname{dg}\mathbf{K}]^{1/2+\beta}\Phi_0[\operatorname{dg}\mathbf{K}]^{1/2+\beta}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} \\
&\quad + \quad [\operatorname{dg}\mathbf{K}]^{\beta}\Phi_0[\operatorname{dg}\mathbf{K}]^{1/2+\beta}\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big]
\end{aligned}$$

Using $\|[\operatorname{dg}\mathbf{K}]^{1/2+\beta}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2}\| \leq \|\operatorname{dg}\mathbf{K}\|^{\beta} \leq \|\mathbf{K}\|^{\beta}$, it follows that

$$\begin{aligned}
\|\mathbf{R} - \mathbf{R}_e\| &\leq \left\|\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big][\operatorname{dg}\mathbf{K}]^{1/2+\beta}\right\|\|\Phi_0\|\|\operatorname{dg}\mathbf{K}\|^{\beta} \\
&\quad + \quad \|\operatorname{dg}\mathbf{K}\|^{\beta}\|\Phi_0\|\left\|[\operatorname{dg}\mathbf{K}]^{1/2+\beta}\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big]\right\|
\end{aligned}$$

The conclusion is now an obvious consequence of Lemma C.4. □

Lemma C.4. *We have*

$$\left\|\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big][\operatorname{dg}\mathbf{K}]^{1/2+\beta}\right\| \leq \begin{cases} \epsilon^{\beta} & 0 < \beta \leq 1 \\ \epsilon \cdot \|\operatorname{dg}\mathbf{K}\|^{\beta-1} & 1 < \beta < \infty \end{cases}$$

Proof. By the spectral mapping theorem,

$$\left\|\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{K}]^{-1/2} - [\operatorname{dg}\mathbf{K}]^{-1/2}\Big][\operatorname{dg}\mathbf{K}]^{1/2+\beta}\right\| \leq \sup_{0\leq\lambda\leq\|\operatorname{dg}\mathbf{K}\|}\left\{\left|\frac{1}{\sqrt{\epsilon+\lambda}} - \frac{1}{\sqrt{\lambda}}\right| \cdot \lambda^{1/2+\beta}\right\}$$

It can be shown using some elementary calculations that

$$\left|\frac{1}{\sqrt{\epsilon+\lambda}} - \frac{1}{\sqrt{\lambda}}\right| \cdot \lambda^{1/2+\beta} = \frac{\epsilon\lambda^{\beta}}{\sqrt{\epsilon+\lambda}(\sqrt{\lambda} + \sqrt{\epsilon+\lambda})} \leq \begin{cases} \epsilon\left[\frac{\lambda^{\beta}}{\epsilon+\lambda}\right] & 0 < \beta < 1/2 \\ \epsilon\left[\frac{\lambda^{2\beta-1}}{\epsilon+\lambda}\right]^{1/2} & 1/2 \leq \beta < 1 \\ \epsilon\lambda^{\beta-1} & 1 \leq \beta < \infty \end{cases}$$

The conclusion follows from Lemma C.5. □

Lemma C.5. *For $0 < x < 1$ and $\lambda \geq 0$, we have*

$$\frac{\lambda^x}{\epsilon+\lambda} \leq \frac{\epsilon^{x-1}}{2}$$

Proof. Consider the reciprocal expression. It follows from elementary differential calculus

that the minimum of the reciprocal occurs at $\lambda_* = x\epsilon/(1-x)$. Therefore,

$$\frac{\epsilon}{\lambda^x} + \lambda^{1-x} \geq \frac{\epsilon}{\lambda_*^x} + \lambda_*^{1-x} = \frac{\epsilon^{1-x}}{x^x(1-x)^{1-x}} \geq \frac{\epsilon^{1-x}}{\max_{0<x<1}[x^x(1-x)^{1-x}]} = 2\epsilon^{1-x}$$

□

C.2.2 Concentration Inequalities

Proof of Theorem 3.4. Apply Theorem 9 from Koltchinskii and Lounici (2017) and replace t with $nt^2/\|\mathbf{K}\|^2$, simplify and restate the conditions accordingly. □

We now prove a concentration inequality for the correlation operator.

Proof of Theorem 3.5.

1. This is a straightforward consequence of Theorem 3.4 and 3.3.

$$\mathbb{P}[\|\hat{\mathbf{R}} - \mathbf{R}\| > \rho] \leq \mathbb{P}[\|\hat{\mathbf{K}} - \mathbf{K}\| > (\rho/M_R)^{1+1/\beta\wedge 1}] \leq \exp\left[-c_R n \rho^{2+2/\beta\wedge 1}\right].$$

2. Under Assumption 3∗, $r = 1 + \inf_k \lambda_k(\mathbf{R}_0) > 0$. Thus, $\mathbf{R} \geq r\mathbf{I}$. By the spectral mapping theorem, $\|\mathbf{P}\| \leq 1/r$. For $\|f\| = 1$, we have

$$\langle f, [\hat{\mathbf{R}} - (r-\rho)\mathbf{I}]f\rangle = \rho + \langle f, [\hat{\mathbf{R}} - \mathbf{R}]f\rangle + \langle f, [\mathbf{R} - r\mathbf{I}]f\rangle$$

and so,

$$\begin{aligned}\inf_f \langle f, [\hat{\mathbf{R}} - (r-\rho)\mathbf{I}]f\rangle &\geq \rho + \inf_f \langle f, [\hat{\mathbf{R}} - \mathbf{R}]f\rangle + \inf_f \langle f, [\mathbf{R} - r\mathbf{I}]f\rangle \\ &\geq \rho - \|\hat{\mathbf{R}} - \mathbf{R}\|.\end{aligned}$$

The result follows by the spectral mapping theorem from the following observation

$$\mathbb{P}[\|\hat{\mathbf{P}}\| > (r-\rho)^{-1}] \leq \mathbb{P}[\|\hat{\mathbf{R}} - \mathbf{R}\| > \rho].$$

3. Using a union bound, we have

$$\begin{aligned}\mathbb{P}[\|\hat{\mathbf{P}} - \mathbf{P}\| > \rho] &\leq \mathbb{P}[\|\hat{\mathbf{P}}\| > (r-\rho)^{-1}] + \mathbb{P}[\|\hat{\mathbf{R}} - \mathbf{R}\| > \rho(r-\rho)/\|\mathbf{P}\|] \\ &\leq \exp\left\{-c_R n \rho^{2+2/(\beta\wedge 1)}\right\} + \exp\left\{-c_R n \left[\rho(r-\rho)/\|\mathbf{P}\|\right]^{2+2/(\beta\wedge 1)}\right\}\end{aligned}$$

Now we need only notice that since $0 < r \leq 1$ and $\|\mathbf{P}\| = 1/r$, we must have $\rho > \rho(r-\rho)/\|\mathbf{P}\|$. If we require that $\rho \leq r/2$, then $\rho(r-\rho)/\|\mathbf{P}\| \geq \rho r^2/2$ and the conclusion follows.

□

C.3 Model Selection Consistency

Proof of Theorem 3.6. Notice that $\hat{\Omega}_\pi \neq \Omega_\pi$ if and only if for some $1 \leq i, j \leq p$ we have

1. $\|\mathbf{P}_{ij}\| \neq 0$ and $\|\hat{\mathbf{P}}_{ij}\| < \rho$, or
2. $\|\mathbf{P}_{ij}\| = 0$ and $\|\hat{\mathbf{P}}_{ij}\| \geq \rho$.

If we require that $\rho < \frac{1}{2} \min_{i,j} \|\mathbf{P}_{ij}\|$, then this implies that for some (i, j) we must have

$$\|\hat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}\| > \rho.$$

Therefore,

$$\begin{aligned} \mathbb{P}[\hat{\Omega}_\pi \neq \Omega_\pi] &= \mathbb{P} \cup_{i,j} [\|\hat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}\| > \rho] \\ &\leq \sum_{i,j=1}^{p} \mathbb{P}[\|\hat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}\| > \rho] \\ &\leq p^2 \cdot \mathbb{P}[\|\hat{\mathbf{P}} - \mathbf{P}\| > \rho]. \end{aligned}$$

Now we apply Theorem 3.5 (3). □

Proof of Theorem 3.7. The proof is a straightforward application of the Borel-Cantelli lemma. Since,

$$\sum_{j=1}^{\infty} \mathbb{P}[\hat{\Omega}_j \neq \tilde{\Omega}_X^{\pi_j}] \leq \sum_{j=1}^{\infty} \alpha_j < \infty$$

it follows that $\mathbb{P}[\hat{\Omega}_j \neq \tilde{\Omega}_X^{\pi_j} \ i.o.] = 0$. With probability 1, there exists some $j_0 \geq 1$ such that for all $j \geq j_0$ we have $\hat{\Omega}_j = \tilde{\Omega}_X^{\pi_j}$. The conclusion follows from observing that $\cap_{j \geq j_0} \tilde{\Omega}_X^{\pi_j} = \Omega_X$. □

D Appendix D. Functional Graphical Lasso

D.1 Background and Notation

Lemma D.1. *The functionals* $|||\cdot|||_{2,\infty}$ *and* $|||\cdot|||_{2,1}$ *are sub-multiplicative norms.*

Proof. For $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1}^{p}$, $\mathbf{B} = [\mathbf{B}_{ij}]_{i,j=1}^{p}$, $A = [\|\mathbf{A}_{ij}\|_2]_{i,j=1}^{p}$ and $B = [\|\mathbf{B}_{ij}\|_2]_{i,j=1}^{p}$, we have

$$|||\mathbf{A}+\mathbf{B}|||_{2,\infty} \leq |||A+B|||_{\infty} \leq |||A|||_{\infty} + |||B|||_{\infty} = |||\mathbf{A}|||_{2,\infty} + |||\mathbf{B}|||_{2,\infty},$$

and similarly,

$$|||\mathbf{AB}|||_{2,\infty} \leq |||AB|||_{\infty} \leq |||A|||_{\infty} \cdot |||B|||_{\infty} = |||\mathbf{A}|||_{2,\infty} \cdot |||\mathbf{B}|||_{2,\infty}.$$

with the same conclusion following for $|||\cdot|||_{2,1}$ from $|||\mathbf{A}|||_{2,1} = \left|\left|\left|\mathbf{A}^{\top}\right|\right|\right|_{2,\infty}$. The first inequalities in both the cases follow from the sub-additivity and sub-multiplicativity of the Hilbert-Schmidt norm. □

D.2 Conditional Independence for Random Elements

In the following proof we will use "sub-setted" matrix $\mathbf{A}_{ss}$ to mean the matrix $\mathbf{A}$ with ijth entries with neither i nor j in s being equal to $\mathbf{0}$. The symbols $\mathbf{A}_{si}$ and $\mathbf{A}_{js}$ are defined accordingly.

Proof of Theorem 4.2. Pick $i \neq j$ such that $\mathbf{P}_{ij} = \mathbf{0}$. Let $s = \{k : 1 \leq k \leq p \text{ for } k \neq i, j\}$.

By Theorem 2.2.3 of Bakonyi and Woerdeman (2011), this is equivalent to

$$\mathbf{R}_{ij} = \mathbf{R}_{is}\mathbf{R}_{ss}^{-1}\mathbf{R}_{sj}.$$

We will show that this is in turn equivalent to saying that for every $f_i \in \mathcal{H}_i$ and $f_j \in \mathcal{H}_j$, we have

$$\langle f_i, X_i\rangle \perp\!\!\!\perp \langle f_j, X_j\rangle \mid \{\langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j\}.$$

Because of Gaussianity, this is equivalent to saying that

$$\begin{aligned}
&\mathrm{Cov}[\langle f_i, X_i\rangle, \langle f_j, X_j\rangle | \langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j] \\
&= \mathbb{E}[\langle f_i, X_i\rangle\langle f_j, X_j\rangle | \langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j] \\
&\quad -\mathbb{E}[\langle f_i, X_i\rangle | \langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j] \cdot \mathbb{E}[\langle f_j, X_j\rangle | \langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j] \\
&= 0.
\end{aligned}$$

Taking the expectation gives that $\mathbb{E}[\langle f_i, X_i\rangle\langle f_j, X_j\rangle]$ is equal to

$$\mathbb{E}\left[\mathbb{E}[\langle f_i, X_i\rangle | \langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j]\mathbb{E}[\langle f_j, X_j\rangle | \langle f_k, X_k\rangle : f_k \in \mathcal{H}_k \text{ where } k \neq i, j]\right] \tag{D.1}$$

Define $\mathbf{f} = (f_1, \ldots, f_p)^\top$, $\mathbf{f}_i = (\mathbf{0}, \ldots, \mathbf{0}, f_i, \mathbf{0}, \ldots, \mathbf{0})^\top$, $\mathbf{f}_j = (\mathbf{0}, \ldots, \mathbf{0}, f_j, \mathbf{0}, \ldots, \mathbf{0})^\top$ and

$$\mathbf{f}_{ij} = (f_1, \ldots, f_{i-1}, \mathbf{0}, f_{i+1}, \ldots, f_{j-1}, \mathbf{0}, f_{j+1}, \ldots, f_p)^\top.$$

Then $\mathbf{f}$, $\mathbf{f}_i$, $\mathbf{f}_j$ and $\mathbf{f}_{ij}$ can be thought of as elements of the product space $\mathcal{H}$ and the random variables $\langle f_i, X_i\rangle$, $\langle f_j, X_j\rangle$ and $\langle f_k, X_k\rangle$ can be written as $\langle \mathbf{f}_i, X\rangle$, $\langle \mathbf{f}_j, X\rangle$ and $\langle \mathbf{f}_{ij}, X\rangle$ respectively.

Notice that the space of random variables $\langle \mathbf{f}_{ij}, X\rangle$ under the inner product $(\langle \mathbf{f}_{ij}, X\rangle, \langle \mathbf{g}_{ij}, X\rangle) \mapsto \mathbb{E}[\langle \mathbf{f}_{ij}, X\rangle\langle \mathbf{g}_{ij}, X\rangle]$ is isomorphic to the reproducing kernel Hilbert space $\mathfrak{H}$ generated by the kernel $K(\mathbf{f}_{ij}, \mathbf{g}_{ij}) = \langle \mathbf{f}_{ij}, \mathbf{C}\mathbf{g}_{ij}\rangle = \langle \mathbf{f}_{ij}, \mathbf{C}_{ss}\mathbf{g}_{ij}\rangle$. By Loève isometry, the expression (D.1) can be rewritten as the inner product in $\mathfrak{H}$ of the elements $\mathbf{f}_{ij} \mapsto \langle \mathbf{C}\mathbf{f}_i, \mathbf{f}_{ij}\rangle = \langle \mathbf{C}_{si}\mathbf{f}_i, \mathbf{f}_{ij}\rangle, \mathbf{f}_{ij} \mapsto \langle \mathbf{C}\mathbf{f}_j, \mathbf{f}_{ij}\rangle = \langle \mathbf{C}_{sj}\mathbf{f}_j, \mathbf{f}_{ij}\rangle \in \mathfrak{H}$ which can be expressed as

$$\langle \mathbf{C}_{ss}^{-1/2}\mathbf{C}_{si}\mathbf{f}_i, \mathbf{C}_{ss}^{-1/2}\mathbf{C}_{sj}\mathbf{f}_j\rangle = \langle \mathbf{f}_i, [\mathbf{C}_{ss}^{-1/2}\mathbf{C}_{si}]^*[\mathbf{C}_{ss}^{-1/2}\mathbf{C}_{sj}]\mathbf{f}_j\rangle$$

which means that we can write $\mathbb{E}[\langle f_i, X_i\rangle\langle f_j, X_j\rangle] = \langle \mathbf{f}_i, \mathbf{C}\mathbf{f}_j\rangle$ as

$$\langle \mathbf{f}_i, \mathbf{C}\mathbf{f}_j\rangle\langle \mathbf{f}_i, \mathbf{C}\mathbf{f}_j\rangle = \langle \mathbf{f}_i, \mathbf{C}_{ij}\mathbf{f}_j\rangle = \langle \mathbf{f}_i, [\mathbf{C}_{ss}^{-1/2}\mathbf{C}_{si}]^*[\mathbf{C}_{ss}^{-1/2}\mathbf{C}_{sj}]\mathbf{f}_j\rangle$$

or equivalently, $\mathbf{C}_{ij} = [\mathbf{C}_{ss}^{-1/2}\mathbf{C}_{si}]^*[\mathbf{C}_{ss}^{-1/2}\mathbf{C}_{sj}]$. This can be written as $\mathbf{C}_{ij} = \mathbf{C}_{is}[\mathbf{C}_{ss}]^{-1}\mathbf{C}_{sj}$ if $\mathbf{C}_{sj}$ is invertible with respect to $\mathbf{C}_{ss}$ which can be easily shown to be equivalent to $\mathbf{R}_{ij} = \mathbf{R}_{is}[\mathbf{R}_{ss}]^{-1}\mathbf{R}_{sj}$. The conclusion in the general case follows from a density argument, namely that the space $\mathbf{C}_{ss}\mathcal{L}(\mathcal{H})$ is dense in $\mathbf{C}_{ss}^{1/2}\mathcal{L}(\mathcal{H})$ and the function

$\mathbf{R}_{sj} \mapsto (\operatorname{dg} \mathbf{C}_{ss})^{1/2} \mathbf{R}_{sj} (\mathbf{C}_{jj})^{1/2} = \mathbf{C}_{sj}$ is continuous. □

Proof of Lemma 4.2. By Corollary 6.4.11 of Bogachev (1998), we can write the log-likelihood $\log \left[\frac{d\mathbb{P}}{d\mathbb{Q}}\right]$ evaluated at X as

$$\log \frac{d\mathbb{P}}{d\mathbb{Q}}(\mathrm{X}) = \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} \left[\sum_{j=1}^{\infty} \frac{1}{\sqrt{\mu_j}} \langle \varphi_i, \psi_j \rangle \langle \mathrm{X}, \psi_j \rangle \right]^2 - \log(1+\lambda_i) \right]$$

where (λ_j, φ_j) and (μ_j, ψ_j) are the eigenpairs of $\mathbf{R}_0$ and $\mathbf{C}_{\mathbb{Q}}$ respectively. Let $\tilde{\lambda}_j$ be such that $1 + \tilde{\lambda}_j = (1+\lambda_j)^{-1}$. Then we have

$$\begin{aligned}
&\int \log \frac{d\mathbb{P}}{d\mathbb{Q}}(\mathrm{X}) d\tilde{\mathbb{P}}(\mathrm{X}) \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} \int \left[\sum_{j=1}^{\infty} \frac{1}{\sqrt{\mu_j}} \langle \varphi_i, \psi_j \rangle \langle \mathrm{X}, \psi_j \rangle \right]^2 d\tilde{\mathbb{P}}(\mathrm{X}) - \log(1+\lambda_i) \right] \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} \int \left[\sum_{j,j'=1}^{\infty} \frac{1}{\sqrt{\mu_j \mu_{j'}}} \langle \varphi_i, \psi_j \rangle \langle \varphi_i, \psi_{j'} \rangle \langle \mathrm{X}, \psi_j \rangle \langle \mathrm{X}, \psi_{j'} \rangle \right] d\tilde{\mathbb{P}}(\mathrm{X}) - \log(1+\lambda_i) \right] \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} \left[\sum_{j,j'=1}^{\infty} \frac{1}{\sqrt{\mu_j \mu_{j'}}} \langle \varphi_i, \psi_j \rangle \langle \varphi_i, \psi_{j'} \rangle \int \langle \mathrm{X}, \psi_j \rangle \langle \mathrm{X}, \psi_{j'} \rangle d\tilde{\mathbb{P}}(\mathrm{X}) \right] - \log(1+\lambda_i) \right] \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} \left[\sum_{j,j'=1}^{\infty} \frac{1}{\sqrt{\mu_j \mu_{j'}}} \langle \varphi_i, \psi_j \rangle \langle \varphi_i, \psi_{j'} \rangle \langle \psi_j, \mathbf{C}_{\tilde{\mathbb{P}}} \psi_{j'} \rangle \right] - \log(1+\lambda_i) \right] \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} \left\langle \varphi_i, \mathbf{C}_{\mathbb{Q}}^{-1/2} \mathbf{C}_{\tilde{\mathbb{P}}} \mathbf{C}_{\mathbb{Q}}^{-1/2} \varphi_i \right\rangle - \log(1+\lambda_i) \right] \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[\frac{\lambda_i}{1+\lambda_i} (1 + \langle \varphi_i, \tilde{\mathbf{R}}_0 \varphi_i \rangle) - \log(1+\lambda_i) \right] \\
&= \frac{1}{2} \sum_{i=1}^{\infty} \left[-\tilde{\lambda}_i \langle \varphi_i, \tilde{\mathbf{R}}_0 \varphi_i \rangle + \log(1+\tilde{\lambda}_i) - \tilde{\lambda}_i \right] \\
&= \frac{1}{2} \left[- \left\langle \sum_{i=1}^{\infty} \tilde{\lambda}_i \varphi_i \otimes \varphi_i, \tilde{\mathbf{R}}_0 \right\rangle + \sum_{i=1}^{\infty} \log(1+\tilde{\lambda}_i) - \tilde{\lambda}_i \right] \\
&= \frac{1}{2} \left[- \operatorname{tr}(\mathbf{H} \tilde{\mathbf{R}}_0) + \log \det{}_2(\mathbf{I} + \mathbf{H}) \right],
\end{aligned}$$

where $\mathbf{H} = (\mathbf{I} + \mathbf{R}_0)^{-1} - \mathbf{I}$. This establishes the claim. □

D.3 Assumptions

Proof of Remark 4.2. Assume that the regression operator $\mathbf{C}_{-i,-i}^{\dagger}\mathbf{C}_{-i,i}$ is Hilbert-Schmidt. This implies that $\mathbf{C}_{-i,i} = \mathbf{C}_{-i,-i}\mathbf{A}$ for some Hilbert-Schmidt operator $\mathbf{A}$. Because $p(\operatorname{dg}\mathbf{C}_{-i,-i}) \geq \mathbf{C}_{-i,-i}$ and consequently $p^2(\operatorname{dg}\mathbf{C}_{-i,-i})^2 \geq \mathbf{C}_{-i,-i}^2$, we can write using Douglas majorization (see Douglas, 1966, Theorem 1) that $\mathbf{C}_{-i,-i} = (\operatorname{dg}\mathbf{C}_{-i,-i})\mathbf{B}$ for some bounded operator $\mathbf{B}$. This implies that $\mathbf{C}_{-i,i} = (\operatorname{dg}\mathbf{C}_{-i,-i})\mathbf{B}\mathbf{A}$ for some Hilbert-Schmidt operator $\mathbf{B}\mathbf{A}$. It follows that $\mathbf{C}_{ii}^{-1}\mathbf{C}_{ij}$ is Hilbert-Schmidt for every $i \neq j$. Now, let $\{(\mu_k, e_k)\}_{k=1}^{\infty}$ $\{(\lambda_l, f_l)\}_{l=1}^{\infty}$ be the eigenpairs of $\mathbf{C}_{ii}$ and $\mathbf{C}_{jj}$ respectively. Then $\mathbf{C}_{ii}^{-1}\mathbf{C}_{ij}$ and $\mathbf{C}_{jj}^{-1}\mathbf{C}_{ji}$ being Hilbert-Schmidt implies $\sum_{k,l=1}^{\infty}\langle e_k, \mathbf{C}_{ij}f_l\rangle^2/\mu_k^2 < \infty$ and $\sum_{k,l=1}^{\infty}\langle e_k, \mathbf{C}_{ij}f_l\rangle^2/\lambda_l^2 < \infty$. By Cauchy-Schwarz inequality, it follows that $\sum_{k,l=1}^{\infty}\langle e_k, \mathbf{C}_{ij}f_l\rangle^2/\mu_k\lambda_l < \infty$ and thus, $\mathbf{R}_{ij} = \mathbf{C}_{ii}^{-1/2}\mathbf{C}_{ij}\mathbf{C}_{jj}^{-1/2}$ is Hilbert-Schmidt for $i \neq j$ which is Assumption 3/1*. It follows that Hilbert-Schmidtness of regression operators $\mathbf{C}_{-i,-i}^{\dagger}\mathbf{C}_{-i,i}$ implies that of the off-diagonal entries $\mathbf{R}_{ij}$ of the correlation operator matrix $\mathbf{R}$.

To see why this is strict, consider the case $p = 2$ with $i = 1$ and $j = 2$, $\mathbf{C}_{12} = \sum_{k=1}^{\infty}\alpha_k e_k \otimes f_k$ with $\lambda_k \sim 1/k^{1/2+\epsilon_\lambda}, \mu_k \sim 1/k^{1/2+\epsilon_\mu}$ and $\alpha_k \sim 1/k^{1/2+\epsilon_\alpha}$ such that $\epsilon_\lambda > \epsilon_\mu > 0$ and $2\epsilon_\alpha - 1 \in (\epsilon_\mu + \epsilon_\lambda, 2\epsilon_\lambda)$. This ensures that $\sum_{k,l=1}^{\infty}\langle e_k, \mathbf{C}_{ij}f_l\rangle^2/\mu_k\lambda_l < \infty$ while $\sum_{k,l=1}^{\infty}\langle e_k, \mathbf{C}_{ij}f_l\rangle^2/\lambda_l^2 = \infty$, implying that $\mathbf{R}_{12}$ is Hilbert-Schmidt while the regression operators $\mathbf{C}_{2,2}^{\dagger}\mathbf{C}_{2,1}$ isn't. □

D.4 Dual Problem

Proof of Theorem 4.1. Let $\mathcal{G}_1[\mathbf{A}] = -\log\det_2(\mathbf{I}+\mathbf{A})$ and $\mathcal{G}_2[\mathbf{A}] = \lambda_n\|\mathbf{A}\|_{2,1}$. It can be shown that

$$\begin{aligned}
\mathcal{G}_1^*[\mathbf{B}] &= \max_{\mathbf{A}}\left[\operatorname{tr}(\mathbf{A}\mathbf{B}) + \log\det_2(\mathbf{I}+\mathbf{A})\right] \\
&= \operatorname{tr}([(\mathbf{I}-\mathbf{B})^{-1}-\mathbf{I}]\mathbf{B}) + \log\det_2(\mathbf{I}+[(\mathbf{I}-\mathbf{B})^{-1}-\mathbf{I}]) \\
&= -\left[\operatorname{tr}([(\mathbf{I}-\mathbf{B})^{-1}-\mathbf{I}][-\mathbf{B}]) - \log\det_2(\mathbf{I}+[(\mathbf{I}-\mathbf{B})^{-1}-\mathbf{I}])\right].
\end{aligned}$$

Note that if we replace $\mathbf{B}$ with $-\mathbf{B}_0$ such that $\operatorname{dg}\mathbf{B}_0 = \mathbf{0}$, the above expression is equal to twice the Kullback-Leibler divergence $\mathcal{D}(\mathbf{B}_0)$ of the Gaussian measure with the correlation operator $\mathbf{I}+\mathbf{B}_0$ with respect to its product measure since $\mathcal{D}(\mathbf{B}_0) = -\frac{1}{2}\left[\operatorname{tr}([(\mathbf{I}+\mathbf{B}_0)^{-1}-\mathbf{I}][\mathbf{B}_0]) - \log\det_2(\mathbf{I}+\mathbf{B}_0)\right]$.

If $\mathbf{R}_0$ is trace-class, so is $\tilde{\mathbf{R}} = (\mathbf{I} + \mathbf{R}_0)^{-1} - \mathbf{I}$, and we can write

$$\begin{aligned}
\mathcal{D}(\mathbf{R}_0) &= -\frac{1}{2}\left[\operatorname{tr}(\tilde{\mathbf{R}}\mathbf{R}_0) - \log\det(\mathbf{I} + \tilde{\mathbf{R}}) + \operatorname{tr}(\tilde{\mathbf{R}})\right] \\
&= -\frac{1}{2}\left[\operatorname{tr}(\tilde{\mathbf{R}}[\mathbf{R}_0 + \mathbf{I}]) - \log\det(\mathbf{I} + \tilde{\mathbf{R}})\right] \\
&= -\frac{1}{2}\left[\operatorname{tr}(\mathbf{I} - [\mathbf{R}_0 + \mathbf{I}]) - \log\det(\mathbf{I} + \tilde{\mathbf{R}})\right] \\
&= -\frac{1}{2}\log\det(\mathbf{I} + \mathbf{R}_0) = -\frac{1}{2}\log\det{}_2(\mathbf{I} + \mathbf{R}_0).
\end{aligned}$$

Since the expression is continuous in the Hilbert-Schmidt norm, the result holds even for Hilbert-Schmidt $\mathbf{R}_0$. Finally,

$$\mathcal{G}_2^*[\mathbf{B}] = \begin{cases} 0 & \text{if } \operatorname{dg}\mathbf{B} = \mathbf{0} \text{ and } \|\mathbf{B}_0\|_{2,\infty} \leq \lambda_n \\ \infty & \text{otherwise.} \end{cases}$$

By combining these two using infimal convolution, we get

$$\begin{aligned}
\mathcal{G}^*(-\hat{\mathbf{R}}_0) &= \inf_{\mathbf{B}}\left[\mathcal{G}_1^*[\mathbf{B}] + \mathcal{G}_2^*(-\hat{\mathbf{R}}_0 - \mathbf{B})\right] \\
&= \inf\{2\mathcal{D}(-\mathbf{B}_0) : \| - \hat{\mathbf{R}}_0 - \mathbf{B}_0\|_{2,\infty} \leq \lambda_n\} \\
&= 2\inf\{\mathcal{D}(\mathbf{B}_0) : \|\mathbf{B}_0 - \hat{\mathbf{R}}_0\|_{2,\infty} \leq \lambda_n\}
\end{aligned}$$

□

D.5 Existence and Uniqueness of Minimizer

The proof below relies on some basic results in convex analysis in Hilbert spaces that can be consulted in Bauschke and Combettes (2011) or Ekeland and Temam (1976).

Lemma D.2. *The functional* $\mathbf{A} \mapsto \log\det{}_2(\mathbf{I} + \mathbf{A})$ *is twice differentiable in the Gâteux sense, with the first and second Gâteux derivatives at* $\mathbf{A}$ *given by* $(\mathbf{I} + \mathbf{A})^{-1} - \mathbf{I}$ *and* $[(\mathbf{I} + \mathbf{A}) \otimes (\mathbf{I} + \mathbf{A})]^{-1}$*, respectively.*

Proof. We simply evaluate the derivative of f at $t = 0$ by looking at its Taylor expansion.

For $\mathbf{A}, \mathbf{B} \in \mathcal{L}_1$:

$$\begin{aligned} f(t) - f(0) &= \log\det{}_2(\mathbf{I} + \mathbf{A} + t\mathbf{B}) - \log\det{}_2(\mathbf{I} + \mathbf{A}) \\ &= \log\det(\mathbf{I} + \mathbf{A} + t\mathbf{B}) - \log\det(\mathbf{I} + \mathbf{A}) - \operatorname{tr}(\mathbf{A} + t\mathbf{B}) + \operatorname{tr}\mathbf{A} \\ &= \log\det\left[\mathbf{I} + t(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B})\right] - t\operatorname{tr}(\mathbf{B}) \\ &= \left[t\operatorname{tr}\left[(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}\right] - \frac{1}{2}t^2\operatorname{tr}\left\{\left[(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}\right]^2\right\} + o(t^3)\right] - t\operatorname{tr}(\mathbf{B}) \\ &= t\operatorname{tr}\left\{[(\mathbf{I} + \mathbf{A})^{-1} - \mathbf{I}]\mathbf{B}\right\} - \frac{1}{2!}t^2\operatorname{tr}\left\{\left[(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}\right]^2\right\} + o(t^3) \end{aligned}$$

The result follows from the continuity of expressions in $\|\cdot\|_2$ norm and the fact that $\mathcal{L}_1$ is dense in $\mathcal{L}_2$. □

Lemma D.3. *The optimization problem (4.11) admits a unique solution* $\hat{\mathbf{H}}$ *for every* $\lambda_n > 0$ *and correlation operator* $\hat{\mathbf{R}}$ *which satisfies*

$$\hat{\mathbf{R}} - (\mathbf{I} + \hat{\mathbf{H}})^{-1} + \lambda_n\hat{\mathbf{Z}} = \mathbf{0} \tag{D.2}$$

for some $\hat{\mathbf{Z}} \in \partial\|\hat{\mathbf{H}}_0\|_1$.

Proof of Lemma D.3. The Carleman-Fredholm determinant is known to be strictly log-concave (see Lemma 2.1 of Bakonyi and Woerdeman (1998)). From Theorem 6.5 of Simon (1977),

$$|\det{}_2(\mathbf{I} + \mathbf{A}) - \det{}_2(\mathbf{I} + \mathbf{B})| \quad \leq \|\mathbf{A} - \mathbf{B}\|_2 \exp\left[\tfrac{1}{2}(\|\mathbf{A}\|_2 + \|\mathbf{B}\|_2 + 1)^2\right].$$

Thus the function $\mathbf{A} \mapsto -\log\det{}_2(\mathbf{I} + \mathbf{A})$ is strictly convex and continuous in $\|\cdot\|_2$. Because the functions $\mathbf{H} \mapsto \operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0)$ are $\mathbf{H} \mapsto \|\mathbf{H}_0\|_1 = \sum_{i\neq j}\|\mathbf{H}_{ij}\|_2$ are also convex, it follows that the function $\mathbf{H} \mapsto F(\mathbf{H})$ is strictly convex and continuous.

Using the method of Lagrange multipliers, we can rewrite the optimization problem (4.11) in a constrained form as

$$\inf_{\|\mathbf{H}_0\|_1 \leq r}\left[\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det{}_2(\mathbf{I} + \mathbf{H})\right]$$

for some $r = r(\lambda_n)$. Notice that $\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) = \sum_{i\neq j}\operatorname{tr}_i(\mathbf{H}_{ij}\hat{\mathbf{R}}_{ji})$ depends only on $\mathbf{H}_0$ and is

hence bounded. Let $\lambda_\infty = \max_j \lambda_j(\mathbf{H})$. Using $\lambda_j(\mathbf{H}) > -1$ we can write

$$\begin{aligned}
[\det{}_2(\mathbf{I}+\mathbf{H})]^{-1} &= \prod_{j=1}^{\infty} e^{\lambda_j(\mathbf{H})}(1+\lambda_j(\mathbf{H}))^{-1} \\
&= \prod_{j=1}^{\infty}\left[1+\frac{1}{(1+\lambda_j(\mathbf{H}))}\sum_{k=2}^{\infty}\frac{\lambda_j^k(\mathbf{H})}{k!}\right] \\
&= \prod_{j=1}^{\infty}\left[1+\frac{1}{(1+\lambda_j(\mathbf{H}))}\sum_{m=1}^{\infty}\frac{\lambda_j^{2m}(\mathbf{H})}{(2m)!}\left(1+\frac{\lambda_j(\mathbf{H})}{2m+1}\right)\right] \\
&\geq \prod_{j=1}^{\infty}\left[1+\frac{1}{(1+\lambda_\infty)}\frac{\lambda_j^2(\mathbf{H})}{2!}\left(1-\frac{1}{3}\right)\right] \\
&\geq 1+\frac{1}{3(1+\lambda_\infty)}\sum_{j=1}^{\infty}\lambda_j^2(\mathbf{H}) \\
&= 1+\frac{1}{3(1+\lambda_\infty)}\|\mathbf{H}\|_2^2
\end{aligned}$$

In the first inequality, we used the fact that $1+\lambda_j(\mathbf{H})/(2m+1) > 0$ and we retained only the first term of the infinite sum. Thus the Carleman-Fredholm determinant $-\log\det{}_2(\mathbf{I}+\mathbf{H}) \to \infty$ as $\|\mathbf{H}\|_2 \to \infty$ and is therefore coercive. It immediately follows that $\mathcal{F}$ admits a unique minimum, say at $\hat{\mathbf{H}}$ (Propostion 1.2, Ekeland and Temam (1976)). Consequently, it must satisfy the stationary condition (Theorem 16.3, Bauschke and Combettes (2011)) at $\hat{\mathbf{H}}$ given by

$$\mathbf{0} \in \partial\mathcal{F}(\hat{\mathbf{H}}).$$

Because $\mathbf{H} \mapsto \mathrm{tr}(\mathbf{H}\hat{\mathbf{R}}_0)$ and $\mathbf{H} \mapsto \log\det{}_2(\mathbf{I}+\mathbf{H})$ are Gâteaux differentiable with the Gâteaux derivatives at $\hat{\mathbf{H}}$ given by $\hat{\mathbf{R}}_0$ and $(\mathbf{I}+\hat{\mathbf{H}})^{-1}-\mathbf{I}$ respectively, this is equivalent to saying that there exists $\hat{\mathbf{Z}} \in \partial\|\hat{\mathbf{H}}_0\|_1$ such that

$$\hat{\mathbf{R}}_0 - \left[(\mathbf{I}+\hat{\mathbf{H}})^{-1}-\mathbf{I}\right] + \lambda_n\hat{\mathbf{Z}} = \hat{\mathbf{R}} - (\mathbf{I}+\hat{\mathbf{H}})^{-1} + \lambda_n\hat{\mathbf{Z}} = \mathbf{0}.$$

Hence proved. □

D.6 Proof of Main Result

Our proof is a nontrivial adaptation of the multivariate proof in Ravikumar et al. (2011) to the more general setting of elements and operators in Hilbert spaces. We closely follow the steps in the original proof and adapt them so as to not rely on topological properties such as the compactness of closed and bounded sets which are absent in infinite-dimensional spaces. We will use the new norms that we have developed to mimic the topology of Euclidean space in the product Hilbert space $\mathcal{H}$. Crucially, controlling the difference between the oracle solution and the true solution (Lemma D.6) is achieved

using the Banach fixed-point theorem instead of Brouwer's fixed-point theorem.

Let $\tilde{\mathbf{H}}$ be the solution of the original optimization problem (4.11) assuming that the support $S = \{(i,j) : \mathbf{H}^*_{ij} \neq \mathbf{0}\}$ of $\mathbf{H}^*$ is known. Thus,

$$\tilde{\mathbf{H}} = \underset{\mathbf{H}_{S^c}=\mathbf{0}}{\arg\min} \left[\operatorname{tr}(\mathbf{H}\hat{\mathbf{R}}_0) - \log\det_2(\mathbf{I}+\mathbf{H}) + \lambda_n \|\mathbf{H}_0\|_{2,1} \right]. \tag{D.3}$$

The essence of the proof is to show that the *oracle* solution $\tilde{\mathbf{H}}$ assuming the support is known, is equal to the solution $\hat{\mathbf{H}}$ with high probability. Because the oracle solution $\tilde{\mathbf{H}}$ possesses nice properties, the same applies to $\hat{\mathbf{H}}$ with high probability. To show that $\tilde{\mathbf{H}}$ solves the original problem we will show that it satisfies the stationary condition (4.17) for some $\tilde{\mathbf{Z}} \in \partial\|\tilde{\mathbf{H}}_0\|_{2,1}$. We will therefore construct a *witness* $\tilde{\mathbf{Z}} \in \partial\|\tilde{\mathbf{H}}_0\|_{2,1}$ such that

$$\hat{\mathbf{R}} - (\mathbf{I}+\tilde{\mathbf{H}})^{-1} + \lambda_n \tilde{\mathbf{Z}} = \mathbf{0} \tag{D.4}$$

holds.

Notice that because $\tilde{\mathbf{H}}$ solves (D.3), it satisfies the corresponding stationary condition, which can be written as

$$\hat{\mathbf{R}}_S - (\mathbf{I}+\tilde{\mathbf{H}})^{-1}_S + \lambda_n \tilde{\mathbf{Z}}_S = \mathbf{0}$$

for some $\mathbf{Z}_S \in \partial \left[\sum_{(i,j)\in S, i\neq j} \|\tilde{\mathbf{H}}_{ij}\|_2 \right]$ (note that this implies that $\|\mathbf{Z}_S\|_{2,\infty} \leq 1$). Thus (D.4) is already satisfied for the entries $(i,j) \in S$. To ensure that (D.4) is satisfied for all entries, we simply define $\mathbf{Z}_{S^c}$ as follows:

$$\tilde{\mathbf{Z}}_{S^c} = \frac{1}{\lambda_n} \left[-\hat{\mathbf{R}}_{S^c} + [(\mathbf{I}+\tilde{\mathbf{H}})^{-1}]_{S^c} \right].$$

Now that (D.4) is satisfied, we need to show that $\tilde{\mathbf{Z}} \in \partial\|\tilde{\mathbf{H}}_0\|_{2,1}$, which obviously holds for the entries $(i,j) \in S$ by construction. For $(i,j) \in S^c$, notice that $\tilde{\mathbf{H}}_{S^c} = \mathbf{0}$ and therefore its subdifferential $\partial\|\tilde{\mathbf{H}}_{S^c}\|_{2,1}$ has a particularly simple form given by $\partial\|\tilde{\mathbf{H}}_{S^c}\|_{2,1} = \{\mathbf{Z}_{S^c} : \|\mathbf{Z}_{S^c}\|_{2,\infty} < 1\}$. The proof thus reduces to showing that $\tilde{\mathbf{Z}}_{S^c}$ as defined above satisfies

$$\|\tilde{\mathbf{Z}}_{S^c}\|_{2,\infty} < 1. \tag{D.5}$$

with high probability. The condition (D.5) is known as *strict dual feasibility*.

We begin by showing in Lemma D.4 that strict dual feasibility is satisfied if the *sampling noise* $\mathbf{W}$ and the *remainder term* $\mathcal{E}(\mathbf{D})$ are small enough compared to λ_n where

$$\mathbf{W} = \hat{\mathbf{R}} - (\mathbf{I}+\mathbf{H}^*)^{-1} = \hat{\mathbf{R}} - \mathbf{R}^*, \text{ and} \tag{D.6}$$

$$\mathcal{E}(\mathbf{D}) = (\mathbf{I}+\tilde{\mathbf{H}})^{-1} - (\mathbf{I}+\mathbf{H}^*)^{-1} + (\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I}+\mathbf{H}^*)^{-1} \text{ with } \mathbf{D} = \tilde{\mathbf{H}} - \mathbf{H}^*. \tag{D.7}$$

Then we show that $\mathcal{E}(\mathbf{D})$ can be controlled by controlling $\mathbf{D}$ in Lemma D.5. And finally, we prove in Lemma D.6 that by controlling $\mathbf{W}$ and λ_n we can control $\mathbf{D}$. In summary, choosing λ_n appropriately and having enough samples to keep $\mathbf{W}$ is small, ensures that $\tilde{\mathbf{H}} = \hat{\mathbf{H}}$, thus transferring the nice oracle properties that $\tilde{\mathbf{H}}$ possesses to $\hat{\mathbf{H}}$.

Proof of Theorem 4.5. Notice that the deviations of $\mathbf{W} = \hat{\mathbf{R}} - \mathbf{R}$ satisfy

$$\mathbb{P}[\|\mathbf{W}_{ij}\|_2 \geq \delta] \leq 1/g(n, \delta)$$

where $g(n, \delta) = f(n, [\delta/\kappa]^{1+1/\beta \wedge 1})$.

According to Lemma D.4, we need to choose the tuning parameter λ_n such that for a large enough sample size n we would have strict dual feasibility, which boils down to:

$$\begin{aligned} \|\mathbf{W}\|_{2,\infty} &\leq \frac{\alpha\lambda_n}{8}, \\ \|\mathcal{E}(\mathbf{D})\|_{2,\infty} &\leq \frac{\alpha\lambda_n}{8}. \end{aligned}$$

The first of the above conditions is satisfied with probability greater than $1 - 1/p^\tau$ by requiring $\lambda_n = 8\bar{\delta}_g(n, p^\tau)/\alpha$ since $\|\mathbf{W}\|_{2,\infty} \leq \|\mathbf{W}\|_{2,2} \leq \bar{\delta}_g(n, p^\tau)$. It turns out that the second condition is also satisfied if we require that n be large enough such that

$$2\gamma \left(1 + \frac{8}{\alpha}\right)^2 \bar{\delta}_g(n, p^\tau) \leq \frac{1}{3\rho d \vee 6\rho^3\gamma d}. \tag{D.8}$$

Indeed, by Lemma D.6, we have $\|\mathbf{D}\|_{2,\infty} \leq 2\gamma(\|\mathbf{W}\|_{2,\infty} + \lambda_n)$ since

$$2\gamma(\|\mathbf{W}\|_{2,\infty}+\lambda_n) \leq 2\gamma \left(1 + \frac{8}{\alpha}\right) \bar{\delta}_g(n, p^\tau) \leq 2\gamma \left(1 + \frac{8}{\alpha}\right)^2 \bar{\delta}_g(n, p^\tau) \leq \frac{1}{3\rho d \vee 6\rho^3\gamma d} \leq \frac{1}{3\rho d} \tag{D.9}$$

and therefore, by Lemma D.5, we have from the bound (D.9) on $\|\mathbf{D}\|_{2,\infty}$,

$$\begin{aligned} \|\mathcal{E}(\mathbf{D})\|_{2,\infty} &\leq \frac{3}{2} d \|\mathbf{D}\|_{2,\infty}^2 \rho^3 \\ &\leq 6\rho^3\gamma^2 d \cdot \left(1 + \frac{8}{\alpha}\right)^2 \bar{\delta}_g(n, p^\tau)^2 \\ &\leq 6\rho^3\gamma^2 d \cdot \left(1 + \frac{8}{\alpha}\right)^2 \bar{\delta}_g(n, p^\tau) \cdot \bar{\delta}_g(n, p^\tau) \\ &\leq 6\rho^3\gamma^2 d \cdot \frac{1}{6\rho\gamma d \vee 12\rho^3\gamma^2 d} \cdot \bar{\delta}_g(n, p^\tau) \\ &\leq \frac{1}{2} \cdot \bar{\delta}_g(n, p^\tau) \leq \bar{\delta}_g(n, p^\tau) = \frac{\alpha\lambda_n}{8}. \end{aligned}$$

Now, (D.8) actually follows from the given condition on the sample size n because it

implies

$$\bar{\delta}_g(n, p^\tau) = \kappa \bar{\delta}_f(n, p^\tau)^{\beta \wedge 1/1+\beta \wedge 1} \leq \left[12(1+8/\alpha)^2 d(\rho\gamma \vee \rho^3\gamma^2)\right]^{-1}.$$

Therefore, we have $\hat{\mathbf{H}} = \mathbf{H}$ with probability $1 - p^\tau$ and when this is indeed true, we can write that $\mathbf{D} = \|\hat{\mathbf{H}} - \mathbf{H}\|_{2,\infty} \leq 2\gamma(1+8/\alpha)\bar{\delta}_g(n, p^\tau)$. It follows that if $\|\mathbf{H}_{ij}\|_2 > 2\gamma(1+8/\alpha)\bar{\delta}_g(n, p^\tau)$, then $\hat{\mathbf{H}}_{ij} \neq \mathbf{0}$. The result follows from rewriting $\bar{\delta}_g(n, p^\tau)$ in terms of $\bar{\delta}_f(n, p^\tau)$. □

D.6.1 Strict Dual Feasibility

For an operator $\mathbf{A}$, let $\overline{\mathbf{A}}$ denote the column vector $[\mathbf{A}_{ij} : (i,j) = (1,1), (1,2), \ldots, (p, p-1), (p,p)]^\top$ indexed by the pairs (i,j), instead of i and j separately as in the original matrix form. In other words, $\overline{\mathbf{A}}$ is the vectorized version of the operator matrix $\mathbf{A}$. Denote the *subvectors* $[\mathbf{A}_{ij} : (i,j) \in S]^\top$ and $[\mathbf{A}_{ij} : (i,j) \in S^c]^\top$ as $\overline{\mathbf{A}}_S$ and $\overline{\mathbf{A}}_{S^c}$ respectively.

Lemma D.4 (Strict Dual Feasibility)**.** *Under the following condition, we have* $\|\tilde{\mathbf{Z}}_{S^c}\|_{2,\infty} < 1$ *and hence,* $\tilde{\mathbf{H}} = \hat{\mathbf{H}}$*.*

$$\|\mathbf{W}\|_{2,\infty} \vee \|\mathcal{E}(\mathbf{D})\|_{2,\infty} \leq \frac{\alpha\lambda_n}{8} \tag{D.10}$$

Proof. We rewrite Equation (4.17) using Equations (D.6) and (D.7) as

$$(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1} + \mathbf{W} - \mathcal{E}(\mathbf{D}) + \lambda_n \tilde{\mathbf{Z}} = \mathbf{0}. \tag{D.11}$$

Let Γ denote the outer product of $(\mathbf{I} + \mathbf{H}^*)^{-1}$ with itself. Then,

$$\overline{(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1}} = \left[(\mathbf{I} + \mathbf{H}^*)^{-1} \otimes (\mathbf{I} + \mathbf{H}^*)^{-1}\right] \overline{\mathbf{D}} = \Gamma\overline{\mathbf{D}}.$$

By vectorizing Equation (D.11), subsetting on S and S^c, and using $\overline{\mathbf{D}}_{S^c} = \mathbf{0}$, we can write

$$\begin{aligned} \Gamma_{SS}\overline{\mathbf{D}}_S + \overline{\mathbf{W}}_S - \overline{\mathbf{E}}_S + \lambda_n \overline{\tilde{\mathbf{Z}}}_S &= \mathbf{0}, & \text{(D.12)} \\ \Gamma_{S^cS}\overline{\mathbf{D}}_S + \overline{\mathbf{W}}_{S^c} - \overline{\mathbf{E}}_{S^c} + \lambda_n \overline{\tilde{\mathbf{Z}}}_{S^c} &= \mathbf{0}. & \text{(D.13)} \end{aligned}$$

for $\mathbf{E} = \mathcal{E}(\mathbf{D})$. Notice that Γ_{SS} is invertible and therefore we can solve the above system of equations for $\overline{\mathbf{D}}_S$ and $\overline{\tilde{\mathbf{Z}}}_{S^c}$ as follows:

$$\begin{aligned} \overline{\mathbf{D}}_S &= \Gamma_{SS}^{-1}\left[-\overline{\mathbf{W}}_S + \overline{\mathbf{E}}_S - \lambda_n \overline{\tilde{\mathbf{Z}}}_S\right] & \text{(D.14)} \\ \overline{\tilde{\mathbf{Z}}}_{S^c} &= -\frac{1}{\lambda_n}\Gamma_{S^cS}\Gamma_{SS}^{-1}(\overline{\mathbf{W}}_S - \overline{\mathbf{E}}_S) - \frac{1}{\lambda_n}(\overline{\mathbf{W}}_{S^c} - \overline{\mathbf{E}}_{S^c}) + \Gamma_{S^cS}\Gamma_{SS}^{-1}\overline{\tilde{\mathbf{Z}}}_S & \text{(D.15)} \end{aligned}$$

Then by taking the $\|\cdot\|_{2,\infty}$-norm,

$$\|\overline{\tilde{\mathbf{Z}}}_{S^c}\|_{2,\infty} \leq \frac{1}{\lambda_n}\left|\left|\left|\Gamma_{S^cS}\Gamma_{SS}^{-1}\right|\right|\right|_{2,\infty}\left(\|\overline{\mathbf{W}}_S\|_{2,\infty} + \|\overline{\mathbf{E}}_S\|_{2,\infty}\right) \tag{D.16}$$

$$+ \frac{1}{\lambda_n}\left(\|\overline{\mathbf{W}}_{S^c}\|_{2,\infty} + \|\overline{\mathbf{E}}_{S^c}\|_{2,\infty}\right) + \|\Gamma_{S^cS}\Gamma_{SS}^{-1}\overline{\tilde{\mathbf{Z}}}_S\|_{2,\infty}. \tag{D.17}$$

By Assumption 4, we have $\left|\left|\left|\Gamma_{S^cS}\Gamma_{SS}^{-1}\right|\right|\right|_{2,\infty} = \max_{e\in S^c}\|\Gamma_{eS}\Gamma_{SS}^{-1}\|_{2,1} \leq 1-\alpha$ and using $\|\overline{\tilde{\mathbf{Z}}}_S\|_{2,\infty} \leq 1$ which follows by construction, we get

$$\|\Gamma_{S^cS}\Gamma_{SS}^{-1}\overline{\tilde{\mathbf{Z}}}_S\|_{2,\infty} = \max_{e\in S^c}\|\Gamma_{eS}\Gamma_{SS}^{-1}\overline{\tilde{\mathbf{Z}}}_S\|_2 \tag{D.18}$$

$$\leq \max_{e\in S^c}\|\Gamma_{eS}\Gamma_{SS}^{-1}\|_{2,1}\|\overline{\tilde{\mathbf{Z}}}_S\|_{2,\infty} \leq (1-\alpha). \tag{D.19}$$

The above bounds together with the inequality (D.10) imply that

$$\|\overline{\tilde{\mathbf{Z}}}_{S^c}\|_{2,\infty} \leq \frac{1-\alpha}{\lambda_n}\left(\frac{\alpha\lambda_n}{4}\right) + \frac{1}{\lambda_n}\left(\frac{\alpha\lambda_n}{4}\right) + (1-\alpha) = 1 - \alpha/2 - \alpha^2/4 < 1 \tag{D.20}$$

□

Thus, controlling the noise level $\mathbf{W}$ and the remainder term $\mathbf{E} = \mathcal{E}(\mathbf{D})$ enables us to enforce strict dual feasibility. We now show how the remainder term itself can be controlled by controlling $\mathbf{D} = \tilde{\mathbf{H}} - \mathbf{H}^*$.

D.6.2 Control of Remainder

Lemma D.5 (Control of Remainder)**.** *Let* $\mathbf{J} = \sum_{k=0}^{\infty}(-1)^k\left[(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}\right]^k$. *Then,*

$$\mathcal{E}(\mathbf{D}) = (\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1}.$$

If $\|\mathbf{D}\|_{2,\infty} \leq 1/3\rho d$, *then* $\left|\left|\left|\mathbf{J}^\top\right|\right|\right|_{2,\infty} \leq 3/2$ *and* $\|\mathcal{E}(\mathbf{D})\|_{2,\infty} \leq \frac{3}{2}d\|\mathbf{D}\|_{2,\infty}^2\rho^3$.

Proof. Recall that

$$\mathcal{E}(\mathbf{D}) = (\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1} - (\mathbf{I}+\mathbf{H}^*)^{-1} + (\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I}+\mathbf{H}^*)^{-1}.$$

Because $\mathbf{D}$ has no more than d non-zero entries in every row or column, we have $|||\mathbf{D}|||_{2,\infty} \leq d\|\mathbf{D}\|_{2,\infty}$. Using sub-additivity and sub-multiplicativity of $|||\cdot|||_{2,\infty}$ we can write

$$\left|\left|\left|(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}\right|\right|\right|_{2,\infty} = |||\mathbf{D}+\mathbf{R}_0^*\mathbf{D}|||_{2,\infty} \leq |||\mathbf{D}|||_{2,\infty} + |||\mathbf{R}_0^*|||_{2,\infty}|||\mathbf{D}|||_{2,\infty} \leq \rho d\|\mathbf{D}\|_{2,\infty} < 1/3.$$

By expanding $(\mathbf{I} + (\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D})^{-1}$ into a geometric series we can write

$$\begin{aligned}(\mathbf{I} + \mathbf{H}^* + \mathbf{D})^{-1} &= (\mathbf{I} + (\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D})^{-1}(\mathbf{I} + \mathbf{H}^*)^{-1} \\ &= (\mathbf{I} + \mathbf{H}^*)^{-1} - (\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1} \\ &+ \sum_{k=2}^{\infty}(-1)^k \left[(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\right]^k (\mathbf{I} + \mathbf{H}^*)^{-1} \\ &= (\mathbf{I} + \mathbf{H}^*)^{-1} - (\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1} \\ &+ (\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}\end{aligned}$$

where $\mathbf{J} = \sum_{k=0}^{\infty}(-1)^k \left[(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\right]^k$ which satisfies,

$$|||\mathbf{J}|||_{2,1} = \left|\left|\left|\mathbf{J}^\top\right|\right|\right|_{2,\infty} \leq \left[1 - |||\mathbf{D}|||_{2,\infty}\left(1 + |||\mathbf{R}_0^*|||_{2,\infty}\right)\right]^{-1} \leq 3/2.$$

It follows from the above expansion that

$$\mathcal{E}(\mathbf{D}) = (\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}$$

and therefore, we can control the remainder term as follows:

$$\begin{aligned}\|\mathcal{E}(\mathbf{D})\|_{2,\infty} &= \|(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}\|_{2,\infty} \\ &= \max_{i,j} \|e_i^\top(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}e_j\|_2 \\ &\leq \max_i \|e_i^\top(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\|_{2,\infty} \cdot \max_j \|(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}e_j\|_{2,1}\end{aligned}$$

where $\{e_i\}_{i=1}^p$ are "unit vectors" given by $e_i = (\delta_{ij}\mathbf{I}_j)_{j=1}^p$ where δ_{ij} is the Kronecker delta. The first term can be bounded as

$$\begin{aligned}\max_i \|e_i^\top(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\|_{2,\infty} &= \max_i \|e_i^\top(\mathbf{D} + \mathbf{R}_0^*\mathbf{D})\|_{2,\infty} \\ &\leq \max_i \|e_i^\top\mathbf{D}\|_{2,\infty} + \max_i \|e_i^\top\mathbf{R}_0^*\mathbf{D}\|_{2,\infty} \\ &\leq \|\mathbf{D}\|_{2,\infty} + \max_i \|e_i^\top\mathbf{R}_0^*\|_{2,1}\|\mathbf{D}\|_{2,\infty} \\ &= (1 + |||\mathbf{R}_0^*|||_{2,\infty})\|\mathbf{D}\|_{2,\infty} = \rho\|\mathbf{D}\|_{2,\infty}\end{aligned}$$

and the second term as

$$\begin{aligned}\max_j \|(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}e_j\|_{2,1} &= \left|\left|\left|(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{D}\mathbf{J}(\mathbf{I} + \mathbf{H}^*)^{-1}\right|\right|\right|_{2,1} \\ &= \left|\left|\left|(\mathbf{I} + \mathbf{H}^*)^{-1}\mathbf{J}^\top\mathbf{D}(\mathbf{I} + \mathbf{H}^*)^{-1}\right|\right|\right|_{2,\infty}.\end{aligned}$$

By substituting $(\mathbf{I} + \mathbf{H}^*)^{-1} = \mathbf{I} + \mathbf{R}_0^*$ and using sub-additivity and sub-multiplicativity

of $|||\cdot|||_{2,\infty}$ we can bound the above term by

$$\begin{aligned} &\leq \left|\left|\left|\mathbf{J}^\top\mathbf{D}\right|\right|\right|_{2,\infty} + \left|\left|\left|\mathbf{R}_0^*\mathbf{J}^\top\mathbf{D}\right|\right|\right|_{2,\infty} + \left|\left|\left|\mathbf{J}^\top\mathbf{D}\mathbf{R}_0^*\right|\right|\right|_{2,\infty} + \left|\left|\left|\mathbf{R}_0^*\mathbf{J}^\top\mathbf{D}\mathbf{R}_0^*\right|\right|\right|_{2,\infty} \\ &\leq \frac{3}{2}|||\mathbf{D}|||_{2,\infty}(1+|||\mathbf{R}_0^*|||_{2,\infty})^2 \\ &\leq \frac{3}{2}d\|\mathbf{D}\|_{2,\infty}\rho^2 \end{aligned}$$

It follows by combining all the above estimates that

$$\|\mathcal{E}(\mathbf{D})\|_{2,\infty} \leq \frac{3}{2}d\|\mathbf{D}\|_{2,\infty}^2\rho^3$$

□

Now that we know how to control the remainder with the error term $\mathbf{D}$, we will see how $\mathbf{D}$ can be controlled using the noise level $\mathbf{W}$ and the tuning parameter λ_n.

D.6.3 The Fixed Point Argument

The proof uses the Banach fixed-point theorem instead of Brouwer's fixed-point theorem as in Ravikumar et al. (2011) to make up for the lack of compactness in $\mathcal{L}_2$.

Lemma D.6 (Control of $\mathbf{D}$)**.** *Let* $r_0 = \min\{1/3\rho d, 1/6\rho^3\gamma d\}$. *If* $r = 2\gamma(\|\mathbf{W}\|_{2,\infty}+\lambda_n) \leq r_0$ *then* $\|\mathbf{D}\|_{2,\infty} = \|\tilde{\mathbf{H}} - \mathbf{H}^*\|_{2,\infty} \leq r$.

Proof. Notice that $\tilde{\mathbf{H}}_{S^c} = \mathbf{H}_S^* = \mathbf{0}$, so it suffices to bound $\mathbf{D}_S = \tilde{\mathbf{H}}_S - \mathbf{H}_S^*$. Notice further that $\tilde{\mathbf{H}}_S$ is the unique solution to the stationary condition (D.3) and hence by subsetting on S we can write that $\tilde{\mathbf{H}}_S$ is the unique solution to

$$\mathcal{G}(\mathbf{H}_S) = -[(\mathbf{I}+\mathbf{H})^{-1}]_S + \hat{\mathbf{R}}_S + \lambda_n\tilde{\mathbf{Z}}_S$$

since we can assume $\mathbf{H}_{S^c} = \mathbf{0}$. Define $\mathbb{B}_r = \{\mathbf{A} \in \mathcal{L}_2 : \|\mathbf{A}_S\|_{2,\infty} \leq r \text{ and } \mathbf{A}_{S^c} = \mathbf{0}\}$. It follows that showing that $\|\mathbf{D}_S\|_{2,\infty} \leq r$ is equivalent to showing that the equation $\mathcal{G}(\mathbf{H}_S^* + \mathbf{D}_S) = \mathbf{0}$ admits a (unique) solution $\mathbf{D}_S \in \mathbb{B}_r$.

Note that a fixed point of the function $\mathcal{F} : \mathcal{L}_2 \to \mathcal{L}_2$ given in the vectorized form by

$$\bar{\mathcal{F}}(\overline{\mathbf{D}}_S) = -\Gamma_{SS}^{-1}\left[\bar{\mathcal{G}}(\mathbf{H}_S^* + \mathbf{D}_S)\right] + \overline{\mathbf{D}}_S. \tag{D.21}$$

in $\mathbb{B}_r$ corresponds to a solution of $\mathcal{G}(\mathbf{H}_S^* + \mathbf{D}_S) = \mathbf{0}$. It suffices for us to show that $\mathcal{F}$ admits a unique fixed point in $\mathbb{B}_r$. We begin by simplifying the expression for $\mathcal{F}$. By

definition,

$$\begin{aligned}\mathcal{G}(\mathbf{H}^*_S + \mathbf{D}_S) &= -\left[(\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1}\right]_S + \hat{\mathbf{R}}_S + \lambda_n \tilde{\mathbf{Z}}_S \\ &= -\left[(\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1}\right]_S + \left[(\mathbf{I}+\mathbf{H}^*)^{-1}\right]_S + \hat{\mathbf{R}}_S - \left[(\mathbf{I}+\mathbf{H}^*)^{-1}\right]_S + \lambda_n \tilde{\mathbf{Z}}_S \\ &= -\left[(\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1}\right]_S + \left[(\mathbf{I}+\mathbf{H}^*)^{-1}\right]_S + \mathbf{W}_S + \lambda_n \tilde{\mathbf{Z}}_S.\end{aligned}$$

By Lemma D.5,

$$(\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1} - (\mathbf{I}+\mathbf{H}^*)^{-1} + (\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}(\mathbf{I}+\mathbf{H}^*)^{-1} = \left[(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}\right]^2 \mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1}.$$

and this can be vectorized as

$$\overline{(\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1} - (\mathbf{I}+\mathbf{H}^*)^{-1}} + \Gamma\overline{\mathbf{D}} = \overline{\left[(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}\right]^2 \mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1}}.$$

By subsetting the above equation on S we can rewrite $\mathcal{F}$ as

$$\begin{aligned}\bar{\mathcal{F}}(\overline{\mathbf{D}}_S) &= -\Gamma^{-1}_{SS}\left[\bar{\mathcal{G}}(\mathbf{H}^*_S + \mathbf{D}_S)\right] + \overline{\mathbf{D}}_S \\ &= \Gamma^{-1}_{SS}\left[\overline{(\mathbf{I}+\mathbf{H}^*+\mathbf{D})^{-1} - (\mathbf{I}+\mathbf{H}^*)^{-1}}\right]_S - \Gamma^{-1}_{SS}\left(\overline{\mathbf{W}_S + \lambda_n \tilde{\mathbf{Z}}_S}\right) + \overline{\mathbf{D}}_S \\ &= \Gamma^{-1}_{SS}\left[\left[(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{D}\right]^2 \mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1}\right]_S - \Gamma^{-1}_{SS}\left(\overline{\mathbf{W}}_S + \lambda_n \overline{\tilde{\mathbf{Z}}}_S\right)\end{aligned}$$

Using essentially the same technique as in Lemma D.5, we can show that for $\mathbf{A}, \mathbf{B} \in \mathbb{B}_r$, we can write

$$\begin{aligned}\|\mathcal{F}(\mathbf{A}) - \mathcal{F}(\mathbf{B})\|_{2,\infty} &\le \gamma \| \left[(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{A}\right]^2 \mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1} - \left[(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{B}\right]^2 \mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1} \|_{2,\infty} \\ &\le \gamma \|(\mathbf{I}+\mathbf{H}^*)^{-1}(\mathbf{A}-\mathbf{B})(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{A}\mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1}\|_{2,\infty} \\ &+ \gamma \|(\mathbf{I}+\mathbf{H}^*)^{-1}\mathbf{B}(\mathbf{I}+\mathbf{H}^*)^{-1}(\mathbf{A}-\mathbf{B})\mathbf{J}(\mathbf{I}+\mathbf{H}^*)^{-1}\|_{2,\infty} \\ &\le \frac{3}{2} d\rho^3 \gamma \left(\|\mathbf{A}\|_{2,\infty} + \|\mathbf{B}\|_{2,\infty}\right) \|\mathbf{A}-\mathbf{B}\|_{2,\infty}.\end{aligned}$$

Therefore, $\|\mathcal{F}(\mathbf{A}) - \mathcal{F}(\mathbf{B})\|_{2,\infty} \le \frac{3}{2} d\rho^3 \gamma \left(\|\mathbf{A}\|_{2,\infty} + \|\mathbf{B}\|_{2,\infty}\right) \|\mathbf{A}-\mathbf{B}\|_{2,\infty}$. It follows from $\|\mathbf{A}\|_{2,\infty}, \|\mathbf{B}\|_{2,\infty} \le r$, that $\|\mathcal{F}(\mathbf{A}) - \mathcal{F}(\mathbf{B})\|_{2,\infty} \le \frac{1}{2}\|\mathbf{A}-\mathbf{B}\|_{2,\infty}$. Thus $\mathcal{F}$ is contractive in $\mathbb{B}_r$. Moreover, $\mathcal{F}$ maps $\mathbb{B}_r$ into itself since

$$\begin{aligned}\|\mathcal{F}(\mathbf{A})\|_{2,\infty} &\le \frac{1}{2}\|\mathbf{A}\|_{2,\infty} + \|\mathcal{F}(\mathbf{0})\|_{2,\infty} \\ &\le \frac{1}{2}\|\mathbf{A}\|_{2,\infty} + \gamma(\|\mathbf{W}\|_{2,\infty} + \lambda_n) \\ &\le \frac{1}{2}r + \frac{1}{2}r = r.\end{aligned}$$

By Banach's fixed-point theorem, it follows that $\mathcal{F}$ has a unique fixed point in $\mathbb{B}_r$. □

D.7 Estimation of the Correlation Operator

Theorem 4.4 is a consequences of a slightly stronger result, in the form of Lemma D.7.

Proof of Theorem 4.4. Apply Lemma D.7 to $X' = (\mathbf{0}, \ldots, \mathbf{0}, X_i, \mathbf{0}, \ldots, \mathbf{0}, X_j, \mathbf{0}, \ldots, \mathbf{0})$ and make the necessary simplifications for κ. □

Lemma D.7 is a variant of Theorem 6.2 from Waghmare and Panaretos (2023), and can be proved similarly with the only important change being the use of the Hilbert-Schmidt norm for bounding the involved quantities. The difference is that the lemma concerns concentration in the Hilbert-Schmidt norm while Theorem 6.2 concerns the same in operator norm. Because the proof is somewhat complicated, we include an outline of the proof along with the most tedious calculations here. We then give the statement of the lemma, and break down its proof into further lemmas.

Recall that the empirical correlation matrix $\hat{\mathbf{R}}$ is given by $\hat{\mathbf{R}} = \mathbf{I} + [\epsilon_n \mathbf{I} + \operatorname{dg} \hat{\mathbf{C}}]^{-1/2} \hat{\mathbf{C}}_0 [\epsilon_n \mathbf{I} + \operatorname{dg} \mathbf{C}]^{-1/2}$ and we are interested in quantifying how well it estimates the correlation operator matrix $\mathbf{R}$. To this end, we define $\mathbf{R}_e$ as

$$\mathbf{R}_e = \mathbf{I} + [\epsilon \mathbf{I} + \operatorname{dg} \mathbf{C}]^{-1/2} \mathbf{C}_0 [\epsilon \mathbf{I} + \operatorname{dg} \mathbf{C}]^{-1/2}$$

where it is implicit that $\epsilon \equiv \epsilon_n$. The operator matrix $\mathbf{R}_e$ is essentially $\hat{\mathbf{R}}$ assuming that $\mathbf{C}$ is known and hence it can be thought of as an oracle estimator of $\mathbf{R}$. The error of estimating $\mathbf{R}$ with $\hat{\mathbf{R}}$ can now be bounded above by estimation and approximation terms as follows:

$$\|\hat{\mathbf{R}} - \mathbf{R}\|_2 \leq \|\hat{\mathbf{R}} - \mathbf{R}_e\|_2 + \|\mathbf{R}_e - \mathbf{R}\|_2.$$

We will now bound the terms on the right hand side in terms of the regularization parameter ϵ, the error $\|\hat{\mathbf{C}} - \mathbf{C}\|_2$ and a few other quantities which depend only on $\mathbf{C}$. Finally, we will choose ϵ so as to minimize the bound and this will give us a rate of convergence in the form of the following result.

Lemma D.7. *If* $\operatorname{dg}(\hat{\mathbf{C}})$ *is positive, then*

$$\|\hat{\mathbf{R}} - \mathbf{R}\|_2 \leq 2(1 + \|\mathbf{R}_0\|_2) \left[\frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2^2}{\epsilon_n^2} + \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon_n} \right] + 2\epsilon_n^{\beta \wedge 1} \cdot \|\Phi\|_2 \cdot \|\operatorname{dg} \mathbf{C}\|^{2\beta - \beta \wedge 1}.$$

In particular, if $\|\hat{\mathbf{C}} - \mathbf{C}\|_2 \leq \delta$ *then for* $\epsilon_n = \delta^{\frac{1}{1+\beta \wedge 1}}$ *we have* $\|\hat{\mathbf{R}} - \mathbf{R}\|_2 \leq \kappa \delta^{\frac{\beta \wedge 1}{1+\beta \wedge 1}}$ *where*

$$\kappa = 8 \left[(1 + \|\mathbf{R}_0\|_2) \vee \|\Phi\|_2 \|\operatorname{dg} \mathbf{C}\|^{2\beta - \beta \wedge 1} \right].$$

Proof. The proof follows from Lemma D.8 and D.10. □

In the following discussion, Lemmas D.7, D.8, D.9, D.10, D.11 are merely Hilbert-Schmidt counterparts of almost identical results (with almost identical proofs) in Waghmare and Panaretos (2023), while Lemma D.12 is exactly identical and is included here for convenience. We begin by treating the estimation error.

Lemma D.8. *If* $\operatorname{dg}\hat{\mathbf{C}}$ *is positive then*

$$\|\hat{\mathbf{R}} - \mathbf{R}_e\|_2 \leq 2(1 + \|\mathbf{R}_0\|_2)\left[\frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2^2}{\epsilon^2} + \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon}\right]$$

Proof. It can be verified with some simple algebraic manipulation that

$$\begin{aligned}
\hat{\mathbf{R}} - \mathbf{R}_e &= \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\right][\hat{\mathbf{C}}_0 - \mathbf{C}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} \\
&+ \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\right]\mathbf{C}_0\left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\right] \\
&+ \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\right]\mathbf{C}_0[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} \\
&+ [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}[\hat{\mathbf{C}}_0 - \mathbf{C}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} \\
&+ [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\mathbf{C}_0\left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\right].
\end{aligned}$$

Using $\mathbf{C}_0 = [\operatorname{dg}\mathbf{C}]^{1/2}\mathbf{R}_0[\operatorname{dg}\mathbf{C}]^{1/2}$, we can rewrite this expansion in terms of $\mathbf{D}$ and $\mathbf{A}$ given by

$$\begin{aligned}
\mathbf{D} &= [\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} \\
\mathbf{A} &= \left[[\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} - [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\right][\operatorname{dg}\mathbf{C}]^{1/2}
\end{aligned}$$

as

$$\begin{aligned}
&= \mathbf{D}[\hat{\mathbf{C}}_0 - \mathbf{C}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} + \mathbf{A}\mathbf{R}_0\mathbf{A}^* + \mathbf{A}\mathbf{R}_0[\operatorname{dg}\mathbf{C}]^{1/2}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} \\
&+ [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}[\hat{\mathbf{C}}_0 - \mathbf{C}_0][\epsilon\mathbf{I} + \operatorname{dg}\hat{\mathbf{C}}]^{-1/2} + [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}[\operatorname{dg}\mathbf{C}]^{1/2}\mathbf{R}_0\mathbf{A}^*.
\end{aligned}$$

Thus,

$$\begin{aligned}
\|\hat{\mathbf{R}} - \mathbf{R}_e\|_2 &\leq \|\mathbf{D}\| \cdot \|\hat{\mathbf{C}} - \mathbf{C}\|_2 \cdot \frac{1}{\sqrt{\epsilon}} + \|\mathbf{A}\| \cdot \|\mathbf{R}_0\|_2 \cdot \|\mathbf{A}\| + \|\mathbf{A}\| \cdot \|\mathbf{R}_0\|_2 \cdot 1 \\
&+ \frac{1}{\sqrt{\epsilon}} \cdot \|\hat{\mathbf{C}} - \mathbf{C}\|_2 \cdot \frac{1}{\sqrt{\epsilon}} + 1 \cdot \|\mathbf{R}_0\|_2 \cdot \|\mathbf{A}\|.
\end{aligned}$$

Using Lemma D.9 (immediately below) for $\hat{\mathbf{A}} = \operatorname{dg}\hat{\mathbf{C}}$ and $\mathbf{A} = \operatorname{dg}\mathbf{C}$, and using $\|\operatorname{dg}\hat{\mathbf{C}} - \operatorname{dg}\mathbf{C}\|_2 \leq \|\hat{\mathbf{C}} - \mathbf{C}\|_2$, we derive $\|\mathbf{D}\| \leq \|\hat{\mathbf{C}} - \mathbf{C}\|_2/\epsilon^{3/2}$ and $\|\mathbf{A}\| \leq \|\hat{\mathbf{C}} - \mathbf{C}\|_2/\epsilon$. It follows

that

$$\begin{aligned}\|\hat{\mathbf{R}} - \mathbf{R}_\epsilon\|_2 &\leq \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2^2}{\epsilon^2} + \|\mathbf{R}_0\|_2 \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2^2}{\epsilon^2} + \|\mathbf{R}_0\|_2 \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon} \\ &\quad + \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon} + \|\mathbf{R}_0\|_2 \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon} \\ &= (1 + \|\mathbf{R}_0\|_2) \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2^2}{\epsilon^2} + (1 + 2\|\mathbf{R}_0\|_2) \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon} \\ &\leq 2(1 + \|\mathbf{R}_0\|_2) \left[\frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2^2}{\epsilon^2} + \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_2}{\epsilon} \right].\end{aligned}$$

□

Lemma D.9. *If* $\hat{\mathbf{A}}$ *is positive, then*

$$\begin{aligned}\|[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2}\| &\leq \|\hat{\mathbf{A}} - \mathbf{A}\|_2 / \epsilon^{3/2}, \\ \left\| \left[[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2} \right] \mathbf{A}^{1/2} \right\| &\leq \|\hat{\mathbf{A}} - \mathbf{A}\|_2 / \epsilon.\end{aligned}$$

Proof. Notice that

$$\begin{aligned}&[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2} \\ &= [\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} \left[[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2} \right] [\epsilon\mathbf{I} + \mathbf{A}]^{1/2} \\ &= [\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} \left[[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{1/2} + [\epsilon\mathbf{I} + \mathbf{A}]^{1/2} \right]^{-1} \left[[\epsilon\mathbf{I} + \hat{\mathbf{A}}] - [\epsilon\mathbf{I} + \mathbf{A}] \right] [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2} \\ &= \left[\epsilon\mathbf{I} + \hat{\mathbf{A}} + [\epsilon\mathbf{I} + \mathbf{A}]^{1/2} [\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{1/2} \right]^{-1} [\hat{\mathbf{A}} - \mathbf{A}] [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2}.\end{aligned}$$

Since $\hat{\mathbf{A}} + [\epsilon\mathbf{I} + \mathbf{A}]^{1/2} [\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{1/2}$ is positive, we can write

$$\begin{aligned}&\|[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2}\| \\ &\leq \left\| \left[\epsilon\mathbf{I} + \hat{\mathbf{A}} + [\epsilon\mathbf{I} + \mathbf{A}]^{1/2} [\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{1/2} \right]^{-1} \right\| \|\hat{\mathbf{A}} - \mathbf{A}\| \|[\epsilon\mathbf{I} + \mathbf{A}]^{-1/2}\| \\ &\leq \frac{1}{\epsilon} \cdot \|\hat{\mathbf{A}} - \mathbf{A}\|_2 \cdot \frac{1}{\epsilon^{1/2}}\end{aligned}$$

and similarly,

$$\begin{aligned}&\left\| \left[[\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{-1/2} - [\epsilon\mathbf{I} + \mathbf{A}]^{-1/2} \right] \mathbf{A}^{1/2} \right\| \\ &\leq \left\| \left[\epsilon\mathbf{I} + \hat{\mathbf{A}} + [\epsilon\mathbf{I} + \mathbf{A}]^{1/2} [\epsilon\mathbf{I} + \hat{\mathbf{A}}]^{1/2} \right]^{-1} \right\| \|\hat{\mathbf{A}} - \mathbf{A}\| \|[\epsilon\mathbf{I} + \mathbf{A}]^{-1/2} \mathbf{A}^{1/2}\| \\ &\leq \frac{1}{\epsilon} \cdot \|\hat{\mathbf{A}} - \mathbf{A}\|_2 \cdot 1.\end{aligned}$$

□

Now, we will find an upper bound for the approximation error under a regularity condition.

Lemma D.10. *If* $\mathbf{R}_0 = [\operatorname{dg}\mathbf{C}]^\beta \Phi [\operatorname{dg}\mathbf{C}]^\beta$ *for some* $\Phi \in \mathcal{S}_2$ *and* $\beta > 0$*, then*

$$\|\mathbf{R}_e - \mathbf{R}\|_2 \leq \begin{cases} 2\epsilon^\beta \cdot \|\Phi\|_2 \cdot \|\operatorname{dg}\mathbf{C}\|^\beta & 0 < \beta \leq 1 \\ 2\epsilon \ \cdot \|\Phi\|_2 \cdot \|\operatorname{dg}\mathbf{C}\|^{2\beta-1} & 1 < \beta < \infty. \end{cases}$$

Proof. We decompose the difference as follows:

$$\begin{aligned} \mathbf{R} - \mathbf{R}_e &= [\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}[\operatorname{dg}\mathbf{C}]^{1/2}\mathbf{R}[\operatorname{dg}\mathbf{C}]^{1/2}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - \mathbf{R} \\ &= \Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big][\operatorname{dg}\mathbf{C}]^{1/2}\mathbf{R}[\operatorname{dg}\mathbf{C}]^{1/2}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} \\ &\quad + \quad \mathbf{R}[\operatorname{dg}\mathbf{C}]^{1/2}\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big] \\ &= \Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big][\operatorname{dg}\mathbf{C}]^{1/2+\beta}\Phi[\operatorname{dg}\mathbf{C}]^{1/2+\beta}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} \\ &\quad + \quad [\operatorname{dg}\mathbf{C}]^{\beta}\Phi[\operatorname{dg}\mathbf{C}]^{1/2+\beta}\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big]. \end{aligned}$$

Using $\|[\operatorname{dg}\mathbf{C}]^{1/2+\beta}[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2}\| \leq \|\mathbf{C}\|^\beta$, it follows that

$$\begin{aligned} \|\mathbf{R} - \mathbf{R}_e\|_2 &\leq \left\|\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big][\operatorname{dg}\mathbf{C}]^{1/2+\beta}\right\| \|\Phi\|_2 \| \operatorname{dg}\mathbf{C}\|^\beta \\ &\quad + \quad \| \operatorname{dg}\mathbf{C}\|^\beta \|\Phi\|_2 \left\|[\operatorname{dg}\mathbf{C}]^{1/2+\beta}\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big]\right\|. \end{aligned}$$

The conclusion is now a direct consequence of Lemma D.11, stated immediately below. □

Lemma D.11. *We have*

$$\left\|\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big][\operatorname{dg}\mathbf{C}]^{1/2+\beta}\right\| \leq \begin{cases} \epsilon^\beta & 0 < \beta \leq 1 \\ \epsilon \cdot \| \operatorname{dg}\mathbf{C}\|^{\beta-1} & 1 < \beta < \infty. \end{cases}$$

Proof. By the spectral mapping theorem,

$$\left\|\Big[[\epsilon\mathbf{I} + \operatorname{dg}\mathbf{C}]^{-1/2} - [\operatorname{dg}\mathbf{C}]^{-1/2}\Big][\operatorname{dg}\mathbf{C}]^{1/2+\beta}\right\| \leq \sup_{0 \leq \lambda \leq \| \operatorname{dg}\mathbf{C}\|} \left\{\left|\frac{1}{\sqrt{\epsilon+\lambda}} - \frac{1}{\sqrt{\lambda}}\right| \cdot \lambda^{1/2+\beta}\right\}.$$

It can be shown using some elementary calculations that

$$\left|\frac{1}{\sqrt{\epsilon+\lambda}} - \frac{1}{\sqrt{\lambda}}\right| \cdot \lambda^{1/2+\beta} = \frac{\epsilon\lambda^\beta}{\sqrt{\epsilon+\lambda}(\sqrt{\lambda} + \sqrt{\epsilon+\lambda})} \leq \begin{cases} \epsilon\left[\frac{\lambda^\beta}{\epsilon+\lambda}\right] & 0 < \beta < 1/2 \\ \epsilon\left[\frac{\lambda^{2\beta-1}}{\epsilon+\lambda}\right]^{1/2} & 1/2 \leq \beta < 1 \\ \epsilon\lambda^{\beta-1} & 1 \leq \beta < \infty. \end{cases}$$

The conclusion now follows from Lemma D.12, stated immediately below. □

Lemma D.12. *For $0 < x < 1$ and $\lambda \geq 0$, we have*

$$\frac{\lambda^x}{\epsilon + \lambda} \leq \frac{\epsilon^{x-1}}{2}$$

Proof. Consider the reciprocal expression. It follows from elementary differential calculus that the minimum of the reciprocal occurs at $\lambda_* = x\epsilon/(1-x)$. Therefore,

$$\frac{\epsilon}{\lambda^x} + \lambda^{1-x} \geq \frac{\epsilon}{\lambda_*^x} + \lambda_*^{1-x} = \frac{\epsilon^{1-x}}{x^x(1-x)^{1-x}} \geq \frac{\epsilon^{1-x}}{\max_{0<x<1}[x^x(1-x)^{1-x}]} = 2\epsilon^{1-x}.$$

□

D.8 Concentration of Sub-Gaussian Random Elements

Lemma D.13. *Let X be a random element in a Hilbert space $\mathcal{H}$ such that $\mathbb{E}[X] = \mathbf{0}$ and $\|X\|$ is sub-Gaussian (in the sense defined in the main paper). Then for $\mathbf{C}_{ij} = \mathbb{E}[X_i \otimes X_j] - \mathbb{E}[X_i] \otimes \mathbb{E}[X_j]$ and $\hat{\mathbf{C}}_{ij} = \frac{1}{n}\sum_{k=1}^n X_i^k \otimes X_j^k - \bar{X}_i \otimes \bar{X}_j$ we have*

$$\mathbb{P}\{\|\hat{\mathbf{C}}_{ij} - \mathbf{C}_{ij}\|_2 \geq t\} \leq 2\exp\left[-\frac{cnt^2}{\|\|X_i\|\|_{\psi_2}^2 \|\|X_j\|\|_{\psi_2}^2}\right]$$

for $0 \leq t \leq \|\|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] - \bar{X}_i \otimes \bar{X}_j + \mathbb{E}[X_i] \otimes \mathbb{E}[X_j]\|\|_{\psi_1}$ where c is a universal constant. In particular,

$$\mathbb{P}\{\|\hat{\mathbf{C}}_{ij} - \mathbf{C}_{ij}\|_2 \geq t\} \leq 2\exp\left[-\frac{cnt^2}{\|X\|_\infty^4}\right]$$

for $0 \leq t \leq t_X$, where $\|X\|_\infty = \max_i \|\|X_i\|\|_{\psi_2}$ and $t_X = \min_{ij} \|\|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] - \bar{X}_i \otimes \bar{X}_j + \mathbb{E}[X_i] \otimes \mathbb{E}[X_j]\|\|_{\psi_1}$.

Proof. Let $Y_k = X_i^k \otimes X_j^k - \mathbb{E}\left[X_i \otimes X_j\right] - \bar{X}_i \otimes \bar{X}_j + \mathbb{E}[X_i] \otimes \mathbb{E}[X_j]$ for $1 \leq k \leq n$. Then, the Y_k are sub-exponential random elements in the space of Hilbert-Schmidt operators on $\mathcal{H}$. Indeed,

$$\begin{aligned}
\|\|Y\|\|_{\psi_1} &= \|\|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] - \bar{X}_i \otimes \bar{X}_j + \mathbb{E}[X_i] \otimes \mathbb{E}[X_j]\|\|_{\psi_1} \\
&\leq \|\|X_i \otimes X_j\| + \|\mathbb{E}\left[X_i \otimes X_j\right]\| + \|\bar{X}_i \otimes \bar{X}_j\| + \|\mathbb{E}[X_i] \otimes \mathbb{E}[X_j]\|\|_{\psi_1} \\
&\leq \|\|X_i \otimes X_j\|\|_{\psi_1} + \|\|\mathbb{E}\left[X_i \otimes X_j\right]\|\|_{\psi_1} + \|\|\bar{X}_i\|\|\bar{X}_j\|\|_{\psi_1} + \|\|\mathbb{E}[X_i]\|\|\mathbb{E}[X_j]\|\|_{\psi_1} \\
&\leq \|\|X_i\|\|X_j\|\|_{\psi_1} + \|\mathbb{E}\left[\|X_i \otimes X_j\|\right]\|_{\psi_1} + \|\|X_i\|\|X_j\|\|_{\psi_1} + \|\mathbb{E}[\|X_i\|]\mathbb{E}[\|X_j\|]\|_{\psi_1} \\
&\leq \|\|X_i\|\|X_j\|\|_{\psi_1} + \|\mathbb{E}\left[\|X_i \otimes X_j\|\right]\|_{\psi_1} + \|\|X_i\|\|X_j\|\|_{\psi_1} + \|\mathbb{E}[\|X_i\|]\|_{\psi_2}\|\mathbb{E}[\|X_j\|]\|_{\psi_2}.
\end{aligned}$$

Using the centering property of the sub-exponential and sub-Gaussian norms and the fact that product of two sub-Gaussian random variables is sub-exponential (see Exercise 2.7.10 and Lemma 2.7.7 of Vershynin (2018)), we get

$$\begin{aligned}
\|\|\|Y\|\|\|_{\psi_1} &\lesssim \|\|\|X_i\|\|\|X_j\|\|\|_{\psi_1} + \|\|\|X_i \otimes X_j\|\|\|_{\psi_1} + \|\|\|X_i\|\|\|X_j\|\|\|_{\psi_1} + \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2} \\
&= \|\|\|X_i\|\|\|X_j\|\|\|_{\psi_1} + \|\|\|X_i\|\|\|X_j\|\|\|_{\psi_1} + \|\|\|X_i\|\|\|X_j\|\|\|_{\psi_1} + \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2} \\
&\lesssim \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2} + \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2} + \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2} + \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2} \\
&\lesssim \|\|\|X_i\|\|\|_{\psi_2}\|\|\|X_j\|\|\|_{\psi_2}.
\end{aligned}$$

By Theorem 2.8.1 (Bernstein's Inequality) of Vershynin (2018), we have

$$\begin{aligned}
\mathbb{P}\left\{\left\|\tfrac{1}{n}\textstyle\sum_{k=1}^n Y_k\right\| \geq t\right\} &= \mathbb{P}\left\{\|\textstyle\sum_{k=1}^n Y_k\| \geq nt\right\} \\
&\leq \mathbb{P}\left\{\textstyle\sum_{k=1}^n \|Y_k\| \geq nt\right\} \\
&\leq 2\exp\left[-c\left(\frac{n^2t^2}{n\|\|\|Y\|\|\|_{\psi_1}^2} \wedge \frac{nt}{\|\|\|Y\|\|\|_{\psi_1}}\right)\right] \\
&\leq 2\exp\left[-\frac{cnt^2}{\|\|\|Y\|\|\|_{\psi_1}^2}\right]
\end{aligned}$$

for $0 < t < \|\|\|Y\|\|\|_{\psi_1}$ where c is an absolute constant. It follows that

$$\mathbb{P}\left\{\|\hat{\mathbf{C}}_{ij} - \mathbf{C}_{ij}\|_2 \geq t\right\} \leq 2\exp\left[-\frac{cnt^2}{\|\|\|X_i\|\|\|_{\psi_2}^2\|\|\|X_j\|\|\|_{\psi_2}^2}\right]$$

for $0 < t < \|\|\|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right] - \bar{X}_i \otimes \bar{X}_j + \mathbb{E}[X_i] \otimes \mathbb{E}[X_j]\|\|\|_{\psi_1}$ for some absolute constant $c > 0$. □

Remark D.1. *If X has mean zero and we take $\hat{\mathbf{C}} = \frac{1}{n}\sum_{k=1}^n X^k \otimes X^k$, then the above result is still true for $t_X = \min_{ij} \|\|\|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right]\|\|\|_{\psi_1}$. Regardless, even if the mean is not zero, $\bar{X}_i \approx \mathbb{E}[X_i]$ for large enough n and therefore,*

$$t_X \approx \min_{ij} \|\|\|X_i \otimes X_j - \mathbb{E}\left[X_i \otimes X_j\right]\|\|\|_{\psi_1}.$$

D.9 Setups for the Simulation Study

Below we describe how we perform a single draw of a multivariate functional datum in the three setups considered in our simulation study. In all cases, the sample size is chosen as $n = 100$, so this process is repeated independently one hundred times. From the resulting sample, the empirical covariance operator is calculated, which is subsequently transformed into the correlation operator using generalized inverses. All the simulations were run on a computer cluster with the total runtime of about six hundred CPU hours.

D.9.1 Setup 1

This is the only setup where a precision matrix $\mathbf{Q} \in \mathbb{R}^{pr \times pr}$ is explicitly formed, and then inverted to obtain $\mathbf{C} \in \mathbb{R}^{pr \times pr}$. The scores $\delta = (\delta_1, \ldots, \delta_p)^\top \in \mathbb{R}^{pr}$ for the first $r = 10$ Fourier eigenfunctions in all $p = 100$ nodes are then drawn jointly from the multivariate Gaussian distribution with mean 0 and covariance $\mathbf{C}$. Finally, the multivariate functional datum is formed as

$$X = (X_1, \ldots, X_p)^\top = \left(\sum_{l=1}^{r} \delta_{1,l} e_l(t) \ , \ \ldots \ , \ \sum_{l=1}^{r} \delta_{p,l} e_l(t) \right)^\top$$

where $e_l(t)$ is the l-th Fourier basis function. In practice, these are naturally evaluated on an equidistant grid (we use the grid size $K = 30$ throughout this simulation study).

Note that the conditional dependencies are thus directly governed by the block sparsity pattern of the finite-dimensional precision matrix $\mathbf{Q} \in \mathbb{R}^{pr \times pr}$. This simulation setup is thus not truly functional.

It remains to specify the choice of $\mathbf{Q} = (\mathbf{Q}_{i,j})_{j,j=1}^{p}$, where $\mathbf{Q}_{i,j} \in \mathbb{R}^{r \times r}$ for $i, j = 1, \ldots, p$ are the respective blocks. For $i = 1, \ldots, p$, we take

$$\begin{aligned} \mathbf{Q}_{i,i} &= \operatorname{diag}(1, \ldots, 10)/10, \\ \mathbf{Q}_{i,i-1} &= 0.4 \operatorname{diag}(0, \ldots, 0, 6, \ldots, 10)/10, \\ \mathbf{Q}_{i,i-2} &= 0.2 \operatorname{diag}(0, \ldots, 0, 6, \ldots, 10)/10, \end{aligned}$$

and $\mathbf{Q}_{i,i} = 0$ for $|i - j| > 2$. This choice constitutes an AR(2) process. Note that due to the zeros at the end of the diagonals of $\mathbf{Q}_{i,i-1}$ and $\mathbf{Q}_{i,i-2}$ the dependencies are only created between eigenspaces spanned by $e_6(t), \ldots, e_{10}(t)$. Those actually have lower corresponding eigenvalues than $e_1(t), \ldots, e_5(t)$, since $\mathbf{Q}$ is the precision matrix and the relative importance in the spectrum gets reversed when inverting to obtain $\mathbf{C}$.

D.9.2 Setup 2

Here we create a process by applying linear operators to an initial multivariate functional datum $(Z_1, \ldots, Z_p)^\top$ with independent nodes. Firstly, we draw Z_j for $j = 1, \ldots, p$ (again with $p = 100$) independently from a Gaussian distribution with mean zero and a covariance Σ that has the Fourier basis eigenfunctions and quadratically decaying eigenvalues, i.e. $\lambda_l = 1/l^2$ for $l = 1, 2, \ldots$. Then, we create $X = (X_1, \ldots, X_p)^\top$ as

$$\begin{aligned} X_1 &= Z_1, \\ X_2 &= \frac{2}{5} \mathcal{A}_1(X_1) + Z_2, \\ X_j &= \frac{2}{5} \mathcal{A}_1(X_{j-1}) + \frac{1}{5} \mathcal{A}_2(X_{j-2}) + Z_j, \quad j = 3, \ldots, p, \end{aligned}$$

where $\mathcal{A}_1$ and $\mathcal{A}_2$ are zero-extended restriction operators on the first and last tenth of the functional domain. Specifically, if $\Delta_E(t) = 1$ for $t \in E$ and 0 otherwise, where $E \subset [0,1]$, we define $\mathcal{A}_1(f)(t) = \Delta_{[0,1/10]}(t)f(t)$ and $\mathcal{A}_2(f)(t) = \Delta_{[9/10,1]}(t)f(t)$ for any f. Since clearly $\|\mathcal{A}_k(f)\| \leq \|f\|$ for $k = 1, 2$, we have that $\|\mathcal{A}_k\| \leq 1$. Thus, the formulas above define a mean-reverting process that has the Markov property of order 2

The goal of this construction is to have AR(2)-type dependencies, that are however created only locally in the time domain, as opposed to global dependencies in the spectral domain, like in Setup 1. We believe this setup constitutes a more realistic scenario, e.g. from the perspective of neuroimaging applications.

D.9.3 Setup 3

In this final set of simulations, we generate Z_j, $j = 1, \ldots, p$ (with $p = 99$) similarly to the previous setup as independent Gaussian processes with mean zero and the covariance Σ_Z being rank-5 with Fourier basis eigenfunctions and the five non-zero eigenvalues all equal to one. For $j = 3k - 1$ where $k = 1, \ldots, 33$, we then generate W_j as independent zero-mean Gaussian processes (also independent of Z_j's) with the covariance Σ_W and the corresponding kernel $k(t,s) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$, where $H = 0.2$. For $j = 3k$ or $j = 3k - 2$ where $k = 1, \ldots, 33$, we set $W_j := W_{3k-1}$. That is, the W_j's are fractional Brownian motions with relatively rough sample paths (less smooth than those of standard Brownian motion, which corresponds to $H = 0.5$), and they are dependent (in fact, identical) in subsequent triplets. A single multivariate functional datum $X = (X_1, \ldots, X_p)^\top$ is then composed as

$$X_j = 3Z_j + W_j \tag{D.22}$$

and the actual measurements on the equidistant grid are also superposed with additional Gaussian white noise with variance $1/5$.

First, it is easy to verify by calculating conditional covariances that the graphical model is (V, E) with the vertex set $V = \{1, \ldots, p\}$ and the edge set $E = \{(i,j) \mid i, j = 1, \ldots, 99 : \lfloor i/3 \rfloor = \lfloor j/3 \rfloor\}$ in this case, i.e. subsequent triplets of nodes are connected. Next, note that this is the only setup where we add measurement error to the generated values. The reason why we did not do this in the two previous setups (where the signal is smoothly varying) is that this would heavily favor the competing methods, which use denoising as the first step. Of course, we could also use some form of denoising, but we wish to avoid any specific approaches to estimate the covariance, since we view the methodology developed in this paper as one of a plug-in type. In this setup, however, the signal is relatively rough and so the competing approaches working with low-rank projections of the data are at a disadvantage. It is thus reasonable to add measurement error here, which also exemplifies that our method can naturally cope with it. Finally, the constant 3 in formula (D.22) and the white noise variance $1/5$ is chosen such that the total variability

(as captured by the trace of the respective covariance operator) of the smooth component Z_j, the rough signal W_j and the additional white noise, respectively, are in proportions 3:1:2 with each other. Thus while the form of the dependency is particularly simple in this setup, the signal is not very strong in the data.

Bibliography

Antonini, R. G. (1997). Subgaussian random variables in hilbert spaces. *Rendiconti del Seminario Matematico della Università di Padova*, 98:89–99.

Artjomenko, A. (1941a). *Hermitian positive functions and positive functionals*. Dissertation, Odessa State University, Odessa, Ukraine.

Artjomenko, A. P. (1941b). *Hermitian positive functions and positive functionals*. PhD thesis, Odessa State University. Published in Teor. Funkcii, Funkcional. Anal. i Prilozen. 41, (1983) 1–16; 42 (1984), 1–21.

Aston, J. A., Pigoli, D., and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics*, 45(4):1431–1461.

Bachrach, L., Hastie, T., Wang, M., Narasimhan, B., and Marcus, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, black, and Caucasian youth: a longitudinal study. *The Journal of Clinical Endocrinology & Metabolism*, 84(12):4702–4712.

Baker, C. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289.

Bakonyi, M. and Woerdeman, H. J. (1998). Extensions of positive definite fredholm operators with partially zero inverse. *Mathematische Nachrichten*, 193(1):5–18.

Bakonyi, M. and Woerdeman, H. J. (2011). Matrix completions, moments, and sums of Hermitian squares. In *Matrix Completions, Moments, and Sums of Hermitian Squares*. Princeton University Press.

Bauschke, H. H. and Combettes, P. L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York. With a foreword by Hédy Attouch.

Bogachev, V. I. (1998). *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

Braides, A. (2002). Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford.

Burg, J. P. (1975). *Maximum entropy spectral analysis.* Stanford University.

Calderón, A. and Pepinsky, R. (1952). *On the phases of Fourier coefficients for positive real periodic functions*. The X-Ray Crystal Analysis Laboratory, Department of Physics, The Pennsylvania State College.

Carathéodory, C. (1907). Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen. *Math. Ann.*, 64:95–115.

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Mueller, H.-G., and Wang, J.-L. (2021). *fdapace: Functional Data Analysis and Empirical Dynamics*. R package version 0.5.6.

Carvalho, C. M., Massam, H., and West, M. (2007). Simulation of hyper-inverse wishart distributions in graphical models. *Biometrika*, 94(3):647–659.

Chay, S. C. (1972). On quasi-Markov random fields. *Journal of Multivariate Analysis*, 2(1):14–76.

Chen, X. and Yang, Y. (2021). Hanson–wright inequality in hilbert spaces with application to k-means clustering for non-euclidean data. *Bernoulli*, 27(1):586–614.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 24(4):1571–1596.

Codazzi, L., Colombi, A., Gianella, M., Argiento, R., Paci, L., and Pini, A. (2022). Gaussian graphical modeling for spectrometric data analysis. *Computational Statistics & Data Analysis*, page 107416.

Conway, J. B. (2019). *A course in functional analysis*, volume 96. Springer.

Dal Maso, G. (1993). *An introduction to* Γ*-convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA.

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397.

Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, pages 522–539.

Davies, E. B. (1980). *One-parameter semigroups*, volume 15. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York.

Dawid, A. P. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Annals of Mathematics and Artificial Intelligence*, 32(1):335–372.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika*, 103(4):779–799.

Delaigle, A., Hall, P., Huang, W., and Kneip, A. (2021). Estimating the covariance of fragmented and other related types of functional data. *Journal of the American Statistical Association*, 116(535):1383–1401.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28(1):157–175.

Descary, M. and Panaretos, V. (2019a). Functional data analysis by matrix completion. *The Annals of Statistics*, 47(1):1–38.

Descary, M. and Panaretos, V. (2019b). Recovering covariance from functional fragments. *Biometrika*, 106(1):145–160.

Diestel, R. (2010). *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Heidelberg, fourth edition.

Douglas, R. G. (1966). On majorization, factorization, and range inclusion of operators on hilbert space. *Proceedings of the American Mathematical Society*, 17(2):413–415.

Drton, M. and Maathuis, M. H. (2017). Structure learning in graphical modeling. *Annual Review of Statistics and Its Application*, 4:365–393.

Dym, H. and Gohberg, I. (1981). Extensions of band matrices with band inverses. *Linear algebra and its applications*, 36:1–24.

Edwards, R. E. (2012). *Functional analysis: theory and applications*. Courier Corporation.

Eidelman, Y., Milman, V., and Tsolomitis, A. (2004). *Functional analysis*, volume 66 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI. An introduction.

Ekeland, I. and Temam, R. (1976). *Convex analysis and variational problems*, volume Vol. 1. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York. Translated from the French.

Engel, K.-J. and Nagel, R. (2000). *One-parameter semigroups for linear evolution equations*, volume 194 of *Graduate Texts in Mathematics*. Springer-Verlag, New York. With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736*.

Gneiting, T. and Sasvári, Z. (1999). The characterization problem for isotropic covariance functions. *Math. Geology*, 31(1):105–111.

Gohberg, I., Goldberg, S., and Krupnik, N. (2000). *Traces and determinants of linear operators*, volume 116 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel.

Gohberg, I., Kaashoek, M., and Woerdeman, H. (1989). The band method for positive and contractive extension problems. *Journal of operator theory*, pages 109–155.

Grone, R., Johnson, C., Sá, E., and Wolkowicz, H. (1984). Positive definite completions of partial Hermitian matrices. *Linear algebra and its applications*, 58:109–124.

Grone, R., Pierce, S., and Watkins, W. (1990). Extremal correlation matrices. *Linear Algebra and its Applications*, 134:63–70.

Hanke, M. (2017). *A Taste of Inverse Problems: Basic Theory and Examples*. Society for Industrial and Applied Mathematics.

Hida, T. and Hitsuda, M. (1993). *Gaussian processes*, volume 120 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the 1976 Japanese original by the authors.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, volume 997. John Wiley & Sons.

Johnson, C. R. (1990). *Matrix theory and applications*, volume 40. American Mathematical Soc.

Jouzdani, N. and Panaretos, V. (2021). Functional data analysis with rough sampled paths? *arXiv preprint arXiv:2105.12035*.

Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media.

Keich, U. (1999). Krein's strings, the symmetric moment problem, and extending a real positive definite function. *Communications on Pure and Applied Mathematics*, 52(10):1315–1334.

Kneip, A. and Liebl, D. (2020). On the optimal reconstruction of partially observed functional data. *Annals of Statistics*, 48(3):1692–1717.

Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133.

Kovács, S., Ruckstuhl, T., Obrist, H., and Bühlmann, P. (2021). Graphical elastic net and target matrices: Fast algorithms and software for sparse precision matrix estimation. *arXiv preprint arXiv:2101.02148*.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 77(4):777–801.

Krein, M. (1940). Sur le probleme du prolongement des fonctions hermitiennes positives et continues. *CR (Doklady) Acad. Sci. URSS (NS)*, 26(1):17–22.

Krein, M. and Langer, H. (2014). Continuation of Hermitian positive definite functions and related questions. *Integral Equations and Operator Theory*, 78(1):1–69.

Kreyszig, E. (1978). *Introductory functional analysis with applications*, volume 1. Wiley New York.

Kumar, H. (2022). Indian stock market (NSE) 1-Minute Data from Jan 2017. https://www.kaggle.com/datasets/hk7797/stock-market-india.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

Lee, K.-Y., Ji, D., Li, L., Constable, T., and Zhao, H. (2023). Conditional functional graphical models. *Journal of the American Statistical Association*, 118(541):257–271.

Lee, K.-Y., Li, L., Li, B., and Zhao, H. (2022). Nonparametric functional graphical modeling through functional additive regression operator. *Journal of the American Statistical Association*, 0(0):1–15.

Li, B. and Solea, E. (2018a). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *Journal of the American Statistical Association*, 113(524):1637–1655.

Li, B. and Solea, E. (2018b). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *Journal of the American Statistical Association*, 113(524):1637–1655.

Li, C.-K. and Tam, B.-S. (1994). A note on extreme correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 15(3):903–908.

Lin, M. and SenGupta, I. (2021). Analysis of optimal portfolio on finite and small-time horizons for a stochastic volatility market model. *SIAM Journal on Financial Mathematics*, 12(4):1596–1624.

Lin, Z., Wang, J., and Zhong, Q. (2021). Basis expansions for functional snippets. *Biometrika*, 108(3):709–726.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Loeve, M. (2017). *Probability Theory: Third Edition*. Dover Books on Mathematics. Dover Publications.

Masak, T. and Panaretos, V. M. (2022). Random surface covariance estimation by shifted partial tracing. *Journal of the American Statistical Association*, 0(0):1–13.

McKean, Jr, H. (1963). Brownian motion with a several-dimensional time. *Theory of Probability & Its Applications*, 8(4):335–354.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473.

Mogensen, S. W. and Hansen, N. R. (2022). Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, 28(4):3023–3050.

Montague, D. and Rajaratnam, B. (2018). Graphical Markov models for infinitely many variables. *Transactions of the American Mathematical Society*, 370(11):7557–7603.

Narici, L. and Beckenstein, E. (2010). *Topological vector spaces*. Chapman and Hall/CRC.

Parthasarathy, K. (2002). On extremal correlations. *Journal of statistical planning and inference*, 103(1-2):173–180.

Parthasarathy, K. R. and Varadhan, S. S. (1964). Extension of stationary stochastic processes. *Theory of Probability & Its Applications*, 9(1):65–71.

Paulsen, V., Power, S., and Smith, R. (1989). Schur products and matrix completions. *Journal of Functional Analysis*, 85(1):151–178.

Paulsen, V. and Raghupathi, M. (2016). *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152. Cambridge University Press.

Pearl, J. and Paz, A. (1985). Graphoids: A graph-based logic for reasoning about relevance relations. *Technical Report*.

Qiao, X., Guo, S., and James, G. M. (2019). Functional graphical models. *Journal of the American Statistical Association*, 114(525):211–222.

Qiao, X., Qian, C., James, G. M., and Guo, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika*, 107(2):415–431.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer, New York.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515.

Rozanov, Y. A. (1982). Markov random fields. In *Markov Random Fields*, pages 55–102. Springer.

Rudin, W. (1963). The extension problem for positive-definite functions. *Illinois J. Math.*, 7:532–539.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC.

Ryan, R. A. (2002). *Introduction to tensor products of Banach spaces*, volume 73. Springer.

Saitoh, S. and Sawano, Y. (2016). *Theory of reproducing kernels and applications*. Springer Singapore.

Sasvári, Z. (2006). The extension problem for positive definite functions. a short historical survey. In *Operator Theory and Indefinite Inner Product Spaces: Presented on the occasion of the retirement of Heinz Langer in the Colloquium on Operator Theory, Vienna, March 2004*, pages 365–379. Springer.

Shu, H., Tang, H., Zhang, H., and Zheng, W. (2019). Data for: Spectra data classification with kernel extreme learning.

Simon, B. (1977). Notes on infinite determinants of hilbert space operators. *Advances in Mathematics*, 24(3):244–273.

Simon, B. (2005). *Trace ideals and their applications*, volume 120 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition.

Solea, E. and Dette, H. (2022). Nonparametric and high-dimensional functional graphical models. *Electronic Journal of Statistics*, 16(2):6175–6231.

Stewart, J. (1976). Positive definite functions and generalizations, an historical survey. *The Rocky Mountain Journal of Mathematics*, 6(3):409–434.

Stone, M. (1987). *Coordinate-free multivariable statistics*, volume 2 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York. An illustrated geometric progression from Halmos to Gauss and Bayes, Oxford Science Publications.

Treves, F. (2016). *Topological Vector Spaces, Distributions and Kernels: Pure and Applied Mathematics, Vol. 25*, volume 25. Elsevier.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

Waghmare, K. (2022). covcomp: Covariance completion. https://github.com/kgwstat/covcomp.

Waghmare, K. G. and Panaretos, V. M. (2022). The completion of covariance kernels. *The Annals of Statistics*, 50(6):3281 – 3306.

Waghmare, K. G. and Panaretos, V. M. (2023). Continuously indexed graphical models. *arXiv preprint arXiv:2302.02482*.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35.

Zapata, J., Oh, S.-Y., and Petersen, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika*, 109(3):665–681.

Zhao, B., Zhai, P. S., Wang, Y. S., and Kolar, M. (2021). High-dimensional functional graphical model structure learning via neighborhood selection approach. *arXiv preprint arXiv:2105.02487*.

List of Figures

List of Tables

Kartik G. Waghmare

• kartik.waghmare@epfl.ch •
• EPFL SB MATH SMAT, MA B1 513 (Bâtiment MA), Station 8, CH-1015 Lausanne •

Current Position

I am an assistant-doctorant (doctoral student) at the Chair of Mathematical Statistics (SMAT), EPFL working under the supervision of Prof. Victor M. Panaretos, Professor of Mathematical Statistics, Institute of Mathematics, EPFL

Research Interests

Theory of Reproducing Kernels, Functional Data Analysis, Statistical Machine Learning.

Education

- **Doctor of Philosophy (PhD) in Mathematics** **2018 - Present**
  Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.
- **Master of Science (MSc) in Mathematics** **2017 - 2018**
  Indian Institute of Science (IISc), Bangalore, India. *GPA 7.4/8.*
- **Bachelor of Science (BSc) in Mathematics** **2013 - 2017**
  Indian Institute of Science (IISc), Bangalore, India. *First Class.*

Publications and Preprints

- Waghmare, K.G. and Panaretos, V.M., 2022. The Completion of Covariance Kernels. *The Annals of Statistics*, 50(6), pp.3281-3306.
- Waghmare, K.G. and Panaretos, V.M., 2023. Continuously Indexed Graphical Models. *arXiv preprint arXiv:2302.02482.*

Teaching

I have served variously as a teaching and principal teaching assistant by tutoring students, conducting exercise classes, coordinating teaching activities with other assistants, preparing exercises, and organizing exams for the following courses:

1. **MATH-412: Statistical Machine Learning** by Dr. Guillaume Obozinski, as the principal teaching assistant in Autumn 2021 and Autumn 2022.
2. **MATH-240: Statistics** by Prof. Victor Panaretos, as the principal teaching assistant in Spring 2022 and 2020 and as a teaching assistant in Spring 2019.
3. **MATH-240: Statistics** by Dr. Mária Süveges, as the principal teaching assistant in Spring 2021.
4. **MATH-413: Statistics for Data Science** by Prof. Sofia Olhede, as the principal teaching assistant in Autumn 2019.

5. **MATH-413: Statistics for Data Science** by Prof. Victor Panaretos, as a teaching assistant in Autumn 2018.

Furthermore, I am supervising or have supervised the following semester projects:

1. *Statistical Reinforcement Learning* by Yann Eberhard, currently,
2. *Kernel methods in Machine learning and Statistics* by Yerkin Yesbay, Autumn 2022,
3. *Reproducing Kernels and Applications to Statistics* by Qayis Beji, Spring 2022,
4. *Entropy Rate Estimation with Applications to Financial Machine Learning* by Yann Eberhard, Autumn 2021.

Languages

I can speak English (C2), French (B2), German (B1), Hindi (mother tongue) and Marathi (mother tongue).

Programming

I have programmed in R, Python, Julia, MATLAB, Agda, and C.

Achievements

- From 2013 to 2018, I received the prestigious **KVPY Fellowship (SX)** funded by the Department of Science and Technology, Government of India.
- In 2013, I scored among the top 0.1 percentile in two of India's most competitive university entrance examinations (**JEE Main and Advanced)**.

Talks and Conferences

- *Continuously Indexed Graphical Models*, Poster Session at the Swiss Statistical Seminar (2023), Bern.
- *Continuously Indexed Graphical Models*, Contributed Talk at ICSDS 2022, Florence.
- *The Completion of Covariance Kernels*, Invited Talk at ISNPS 2022, Paphos.
- *The Completion of Covariance Kernels*, Poster Session at the Swiss Statistical Seminar (2022), Bern.