Conditional Independence in Continuous Domain[1]

ETH zürich

Kartik G. Waghmare

[1] Research funded by Swiss National Science Foundation Research Grant.

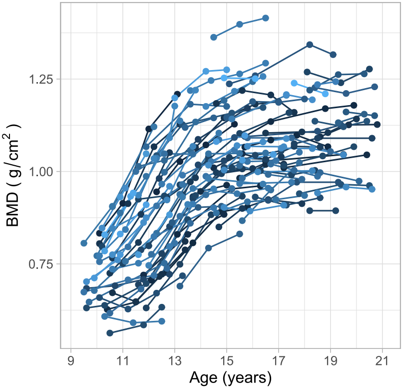
Based on joint work[2,3,4] with:

Prof. Victor Panaretos, EPFL

[2] Waghmare, K.G. and Panaretos, V.M., 2022. The Completion of Covariance Kernels.
[3] Waghmare, K.G. and Panaretos, V.M., 2024. Continuously Indexed Graphical Models.
[4] Waghmare, K.G. and Panaretos, V.M., 2023. The Positive-Definite Completion Problem.

Motivation

Spine Bone Mineral Density: A Longitudinal Study[5]

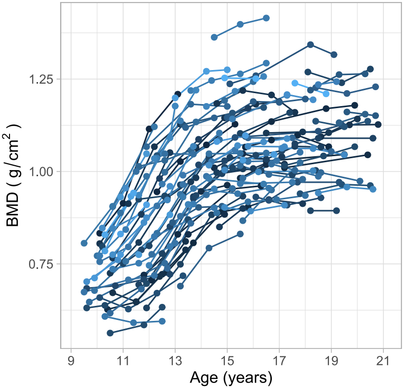


- **Figure.** The BMD measurements of ∼100 individuals taken between the ages of 9 and 21 years.

[5] Bachrach, L.K., Hastie, T., Wang, M.C., 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), pp.4702-4712.

Motivation

Spine Bone Mineral Density: A Longitudinal Study[5]

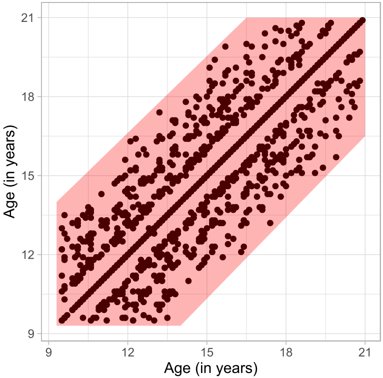


- **Figure.** The BMD measurements of ∼100 individuals taken between the ages of 9 and 21 years.
- No individual was followed for longer than 5 years.

[5] Bachrach, L.K., Hastie, T., Wang, M.C., 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), pp.4702-4712.

Motivation

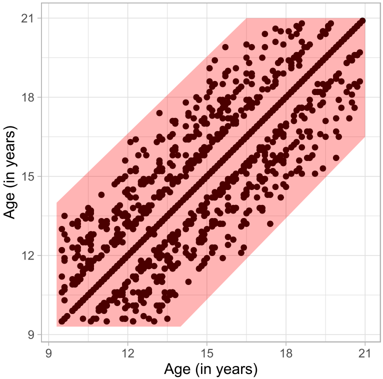
Covariance Estimation



- **Figure.** The pairs (s, t) s.t. BMD measurements of an individual for the both ages s and t are available.

Motivation

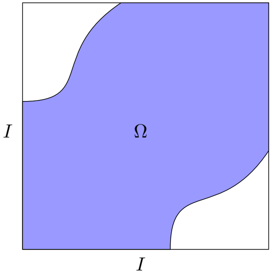
Covariance Estimation



- **Figure.** The pairs (s, t) s.t. BMD measurements of an individual for the both ages s and t are available.
- The covariance can be estimated only over a relatively small region around the diagonal.

Problem Statement

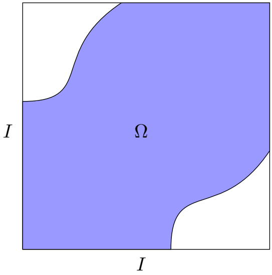
Covariance Recovery



- Let $X = \{X_t : t \in I\}$ be a 2nd-order stochastic process on an interval $I \subset \mathbb{R}$ with mean zero and covariance K.

Problem Statement

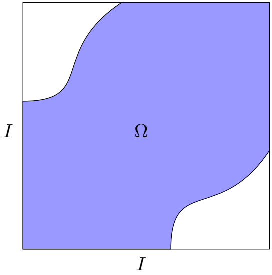
Covariance Recovery



- Let $X = \{X_t : t \in I\}$ be a 2nd-order stochastic process on an interval $I \subset \mathbb{R}$ with mean zero and covariance K.
- Given a consistent estimator $\hat{K}_\Omega$ of the partial covariance $K_\Omega = K|_\Omega$ is it possible to construct a consistent estimator of K?

Problem Statement

Covariance Recovery

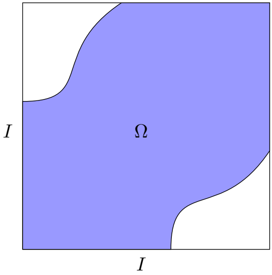


- Let $X = \{X_t : t \in I\}$ be a 2nd-order stochastic process on an interval $I \subset \mathbb{R}$ with mean zero and covariance K.
- Given a consistent estimator $\hat{K}_\Omega$ of the partial covariance $K_\Omega = K|_\Omega$ is it possible to construct a consistent estimator of K?

  1. Under what conditions?

Problem Statement

Covariance Recovery

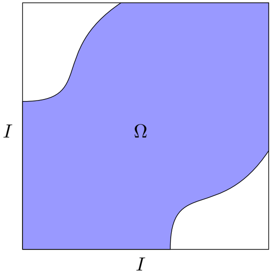


- Let $X = \{X_t : t \in I\}$ be a 2nd-order stochastic process on an interval $I \subset \mathbb{R}$ with mean zero and covariance K.
- Given a consistent estimator $\hat{K}_\Omega$ of the partial covariance $K_\Omega = K|_\Omega$ is it possible to construct a consistent estimator of K?

  1. Under what conditions?
  2. How? And how accurate?

Problem Statement

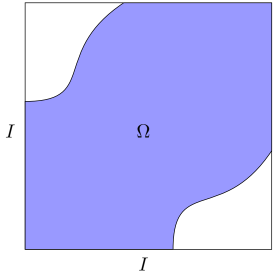
Covariance Completion



- Under what conditions is it possible to extend a kernel $K_\Omega : \Omega \to \mathbb{R}$ to a covariance kernel K on I?

Problem Statement

Covariance Completion

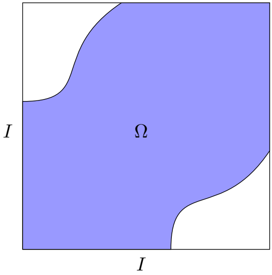


- Under what conditions is it possible to extend a kernel $K_\Omega : \Omega \to \mathbb{R}$ to a reproducing kernel K on I?

    1. Is there a unique completion? If not, how many completions can there be?

Problem Statement

Covariance Completion



- Under what conditions is it possible to extend a kernel $K_\Omega : \Omega \to \mathbb{R}$ to a reproducing kernel K on I?

  1. Is there a unique completion? If not, how many completions can there be?
  2. Is there a special completion which has a nice interpretation in terms of the process X?

Related Work

Fragmented Functional Data

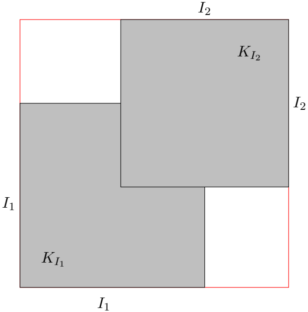
1. **Markov Chains:** Delaigle and Hall (2016).
2. **Unique Completion:** Delaigle, Hall, Huang, and Kneip (2021); Lin, Wang, and Zhong (2021), Descary and Panaretos (2019).

Positive-Definite Completion

1. **Stationary on $\mathbb{Z}$:** Carathéodory (1907); Calderón and Pepinsky (1952).
2. **Stationary on $\mathbb{R}$:** Krein (1940); Rudin (1963).
3. **Nonstationary on finite set:** Dym and Gohberg (1981); Grone, Johnson, Sá, and Wolkowicz (1984).
4. **Nonstationary on $\mathbb{R}$: ?**

Solution

Completion Formula



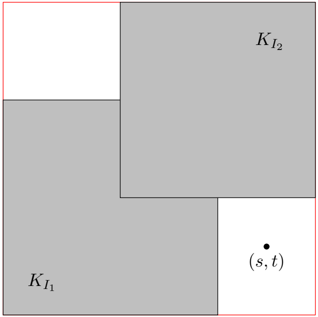
- Let K_Ω be a kernel such that

$$K_{I_1} = K_\Omega|_{I_1 \times I_1}$$
$$K_{I_2} = K_\Omega|_{I_2 \times I_2}$$

are reproducing kernels.

Solution

Completion Formula



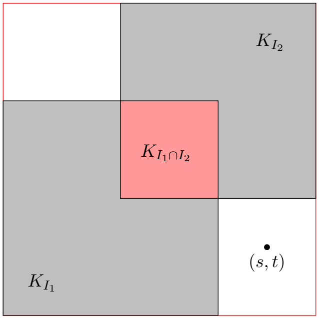
- Let K_Ω be a kernel such that

$$K_{I_1} = K_\Omega|_{I_1 \times I_1}$$
$$K_{I_2} = K_\Omega|_{I_2 \times I_2}$$

  are reproducing kernels.
- How to define $K(s, t)$ such that K is a reproducing kernel?

Solution

Completion Formula

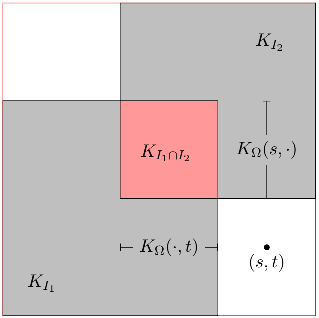


- Consider the *subkernel* $K_{I_1 \cap I_2}$. It has an RKHS

$$\mathcal{H} = \mathcal{H}(K_{I_1 \cap I_2})$$

Solution

Completion Formula



- Consider the *subkernel* $K_{I_1 \cap I_2}$. It has an RKHS

$$\mathcal{H} = \mathcal{H}(K_{I_1 \cap I_2})$$

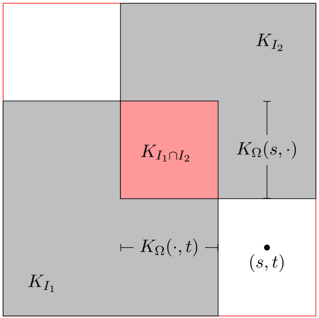
- And consider the *cross-covariances*

$$K_\Omega(t, \cdot) : u \mapsto K_\Omega(t, u)$$
$$K_\Omega(\cdot, s) : u \mapsto K_\Omega(u, s)$$

for $u \in I_1 \cap I_2$.

Solution

Completion Formula



- Consider the *subkernel* $K_{I_1\cap I_2}$. It has an RKHS

$$\mathcal{H} = \mathcal{H}(K_{I_1\cap I_2})$$

- And consider the *cross-covariances*

$$K_\Omega(t,\cdot) : u \mapsto K_\Omega(t,u)$$
$$K_\Omega(\cdot,s) : u \mapsto K_\Omega(u,s)$$

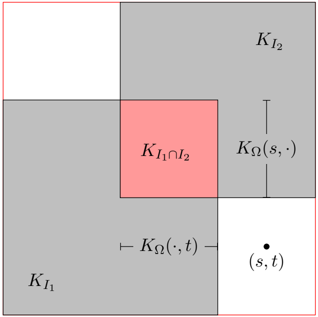
  for $u \in I_1 \cap I_2$.

- Define

$$K(s,t) = \langle K_\Omega(s,\cdot), K_\Omega(\cdot,t)\rangle_{\mathcal{H}_{I_1\cap I_2}}$$

Solution

Completion Formula



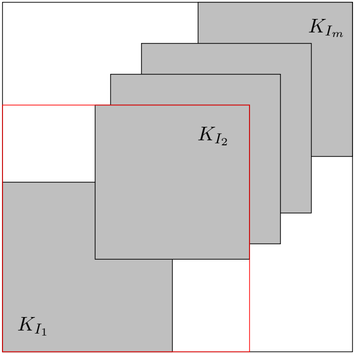
If we define $K_\star : I \times I \to \mathbb{R}$ as

$$\boxed{K_\star(s,t) = \begin{cases} K_\Omega(s,t) & (s,t) \in \Omega \\ \langle K_\Omega(s,\cdot), K_\Omega(\cdot,t)\rangle_{\mathcal{H}} & \text{otherwise.} \end{cases}}$$

the result is a valid covariance on I.

Solution

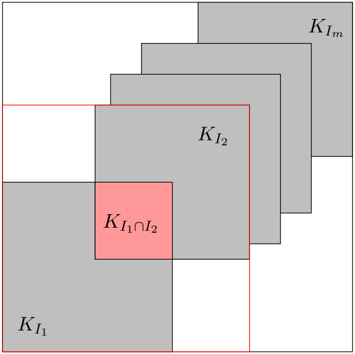
Completion Algorithm



The procedure can be iterated in many ways, but regardless of the manner of completion one recovers the same completion.

Solution

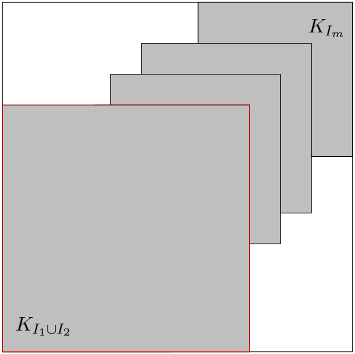
Completion Algorithm



The procedure can be iterated in many ways, but regardless of the manner of completion one recovers the same completion.

Solution

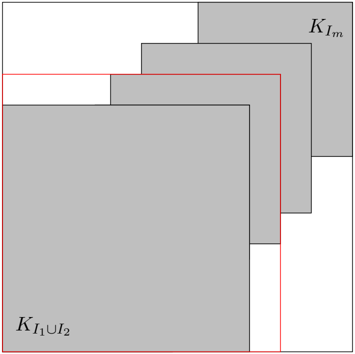
Completion Algorithm



The procedure can be iterated in many ways, but regardless of the manner of completion one recovers the same completion.

Solution

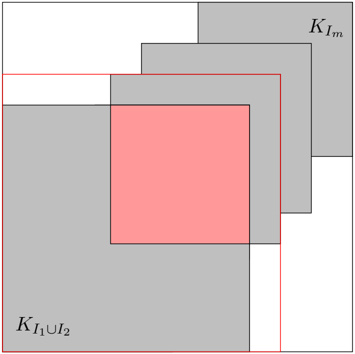
Completion Algorithm



The procedure can be iterated in many ways, but regardless of the manner of completion one recovers the same completion.

Solution

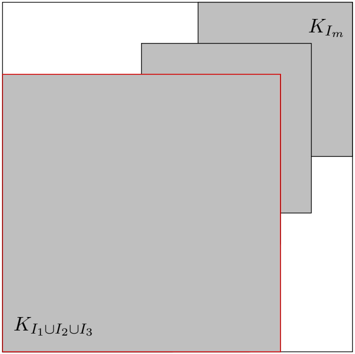
Completion Algorithm



The procedure can be iterated in many ways, but regardless of the manner of completion one recovers the same completion.

Solution

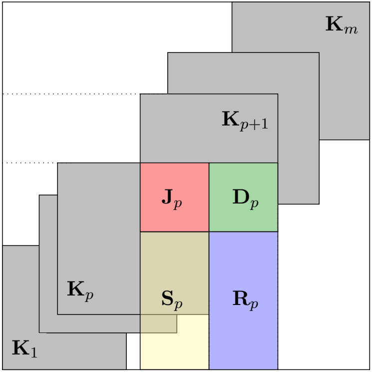
Completion Algorithm



The procedure can be iterated in many ways, but regardless of the manner of completion one recovers the same completion.

Estimation

Construction of $K_\star$



The formula $(*)$ reduces to

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

Using the eigendecomposition

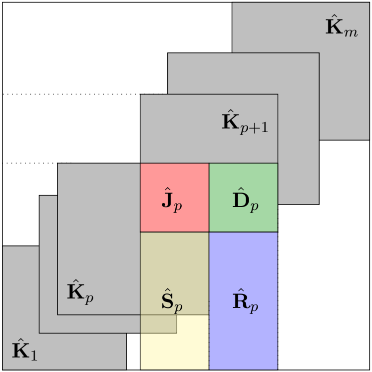
$$\mathbf{J}_p = \sum_{k=1}^{\infty} \lambda_{p,k} e_{p,k} \otimes e_{p,k}$$

we have

$$\boxed{\mathbf{R}_p = \sum_{k=1}^{\infty} \frac{1}{\lambda_{p,k}} \mathbf{S}_p e_{p,k} \otimes \mathbf{D}_p^* e_{p,k}}$$

Estimation

Construction of $\hat{K}_\star$



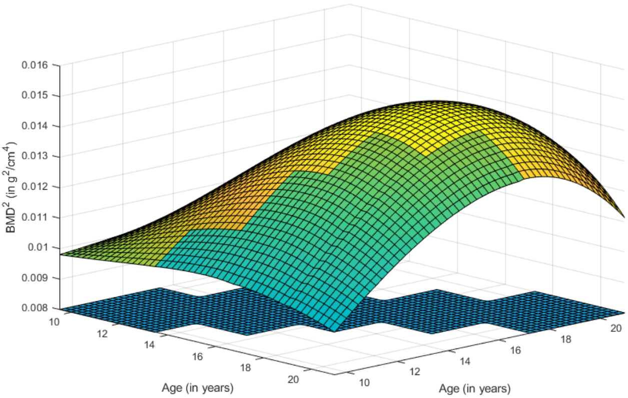
Using the eigendecomposition

$$\hat{\mathbf{J}}_p = \sum_{k=1}^{\infty} \hat{\lambda}_{p,k} \hat{e}_{p,k} \otimes \hat{e}_{p,k}$$

we have

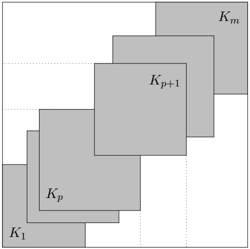
$$\boxed{\hat{\mathbf{R}}_p = \sum_{k=1}^{N(p)} \frac{1}{\hat{\lambda}_{p,k}} \hat{\mathbf{S}}_p \hat{e}_{p,k} \otimes \hat{\mathbf{D}}_p^* \hat{e}_{p,k}}$$

Illustration



Canonical Completion

Naturalness of Completion



If we have

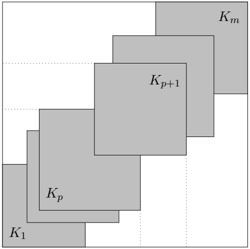
$$K_\Omega(s,t) = \min(s,t) \text{ for } (s,t) \in \Omega,$$

then the completion algorithm gives us

$$\boxed{K_*(s,t) = \min(s,t) \text{ for } s,t \in I.}$$

Canonical Completion

Naturalness of Completion



In fact, the same applies for

- Analytic covariances:

$$K(s,t) = e^{-(s-t)^2}, \cos(s-t), \ldots$$

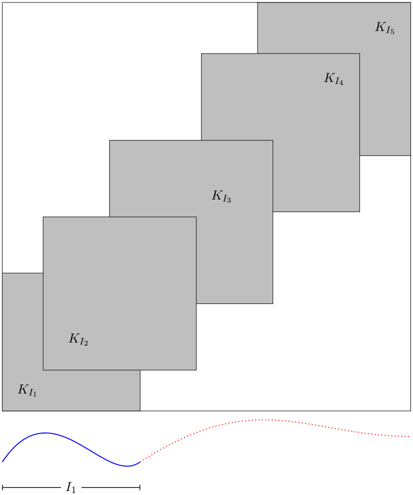
- Covariances of Markov Gaussian process:

$$K(s,t) = \min(s,t), e^{-|s-t|}, \ldots$$

- *Nice* finite rank covariances:

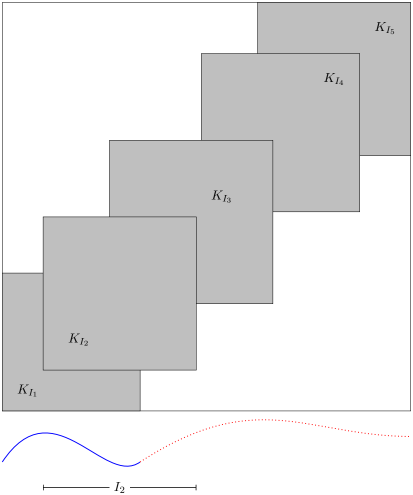
$$K(s,t) = st + s^2t^2, \ldots$$

Connection to Linear Prediction[6]



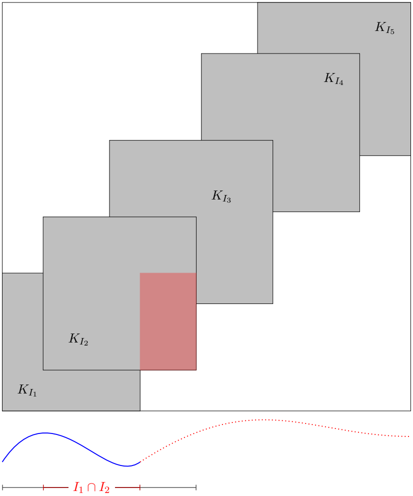
[6] Kneip, A. and Liebl, D., 2020. On the optimal reconstruction of partially observed functional data. *The Annals of Statistics, 48(3)*, pp.1692-1717.

Connection to Linear Prediction[6]



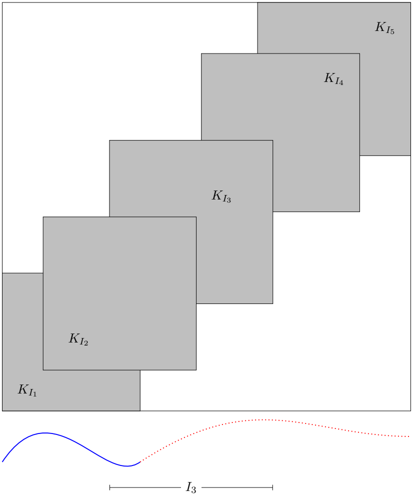
[6] Kneip, A. and Liebl, D., 2020. On the optimal reconstruction of partially observed functional data. *The Annals of Statistics, 48(3)*, pp.1692-1717.

Connection to Linear Prediction[6]



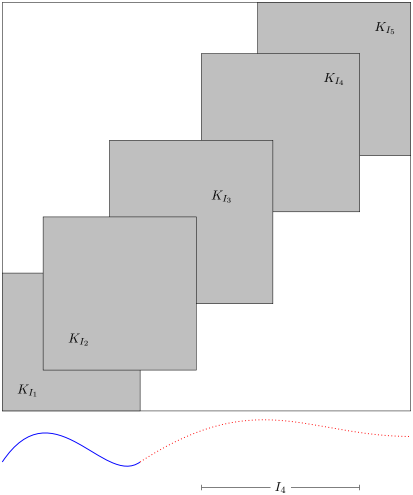
[6] Kneip, A. and Liebl, D., 2020. On the optimal reconstruction of partially observed functional data. *The Annals of Statistics, 48(3)*, pp.1692-1717.

Connection to Linear Prediction[6]



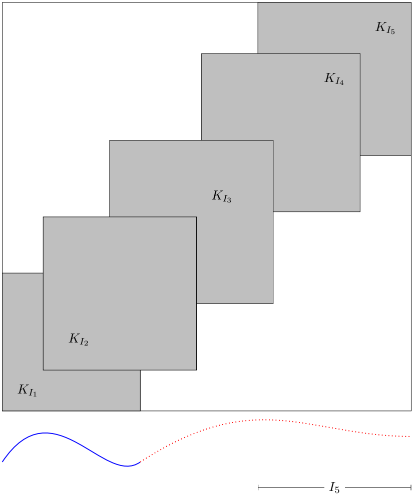
[6] Kneip, A. and Liebl, D., 2020. On the optimal reconstruction of partially observed functional data. *The Annals of Statistics, 48(3)*, pp.1692-1717.

Connection to Linear Prediction[6]



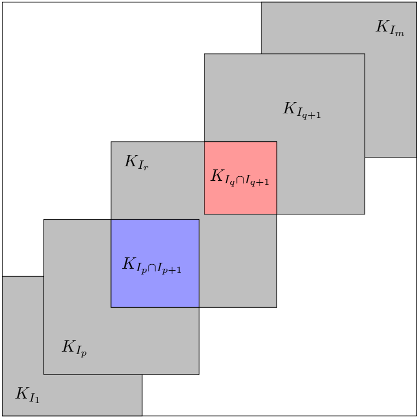
[6] Kneip, A. and Liebl, D., 2020. On the optimal reconstruction of partially observed functional data. *The Annals of Statistics, 48(3)*, pp.1692-1717.

Connection to Linear Prediction[6]



[6] Kneip, A. and Liebl, D., 2020. On the optimal reconstruction of partially observed functional data. *The Annals of Statistics, 48(3)*, pp.1692-1717.

Uniqueness of Completion

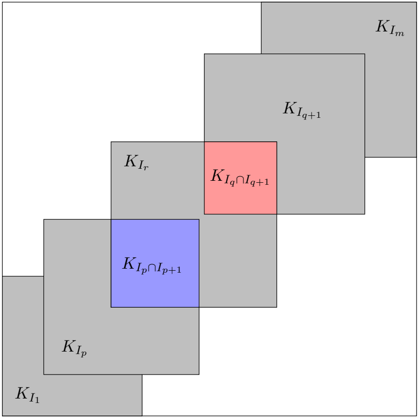


We define the *Schur complement* of a kernel K_I on I with respect to $J \subset I$ as the kernel K_I / K_J

$$= K_I(s,t) - \langle K_I(s,\cdot), K_I(\cdot,t) \rangle_{\mathcal{H}_J}$$

for $s, t \in I \setminus J$.

Uniqueness of Completion



The partial covariance K_Ω admits a unique completion iff

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

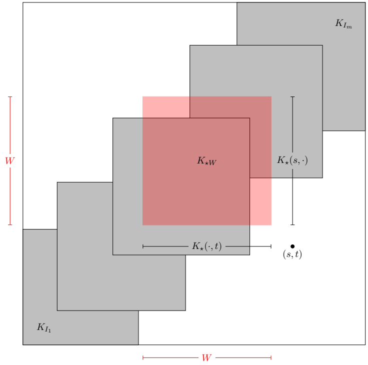
Consequently, for some I_r

$$X_t \in \mathrm{Span}\{X_u : u \in I_r\}$$

for every $t \in I$.

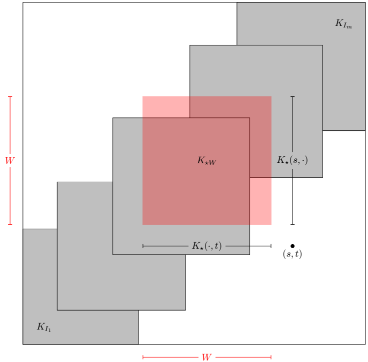
The Separation Property

The Graph Ω



The Separation Property

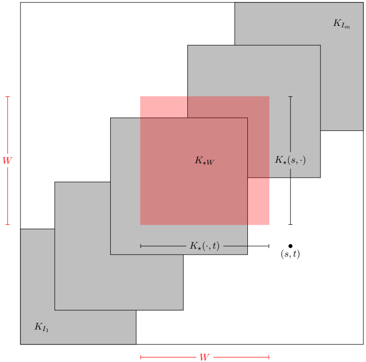
The Graph Ω



- The region Ω can be thought of as a graph with the vertices $s \in I$ and edges $(s,t) \in \Omega$.

The Separation Property

The Graph Ω

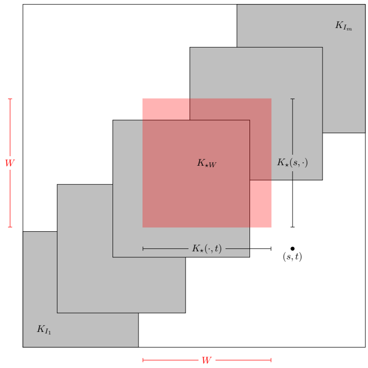


- The region Ω can be thought of as a graph with the vertices $s \in I$ and edges $(s,t) \in \Omega$.
- It turns out that so long as W *separates* $s,t \in I$ with respect to Ω, we have

$$\boxed{K_\star(s,t) = \langle K_\star(s,\cdot), K_\star(\cdot,t)\rangle_{\mathcal{H}_W}}$$

The Separation Property

Conditional Independence

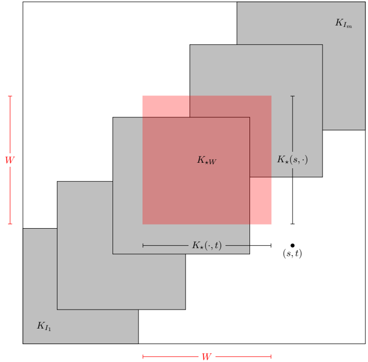


- If X was a Gaussian process, then by *Loève isometry* this implies

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

The Separation Property

Conditional Independence



- If X was a Gaussian process, then by *Loève isometry* this implies

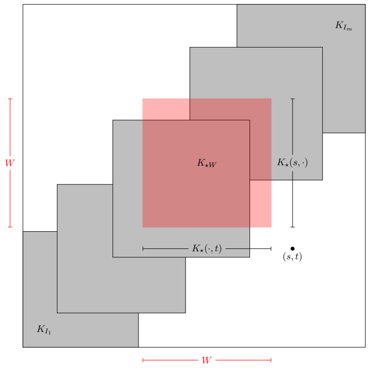
$$\mathbb{E}\left[X_s X_t\right] = \mathbb{E}\left[\mathbb{E}\left[X_s | X_W\right]\mathbb{E}\left[X_t | X_W\right]\right]$$

- And moreover,

$$\operatorname{Cov}(X_s, X_t | X_W) = 0$$

The Separation Property

Conditional Independence



- If X was a Gaussian process, then by *Loève isometry* this implies

$$\mathbb{E}[X_s X_t] = \mathbb{E}[\mathbb{E}[X_s|X_W]\,\mathbb{E}[X_t|X_W]]$$

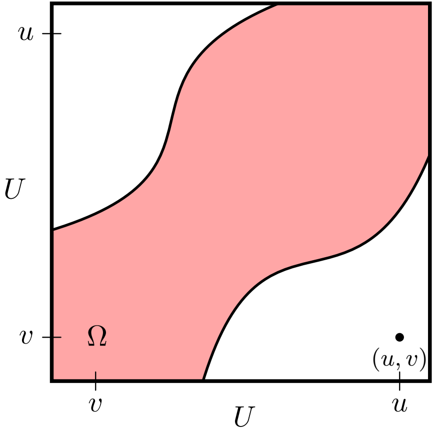
- And moreover,

$$\mathrm{Cov}(X_s, X_t|X_W) = 0$$

- In other words,

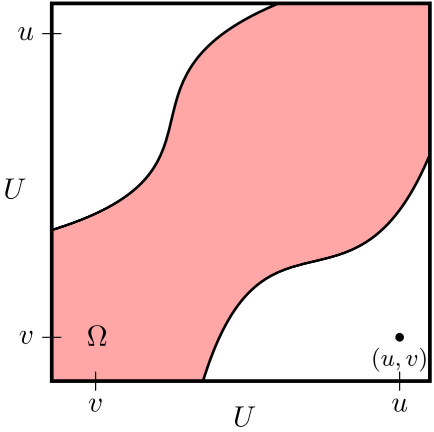
$$X_s \perp\!\!\!\perp X_t \mid \{X_u : u \in W\}.$$

Visualizing Uncountably Infinite Graphs



Assume that U is an interval.

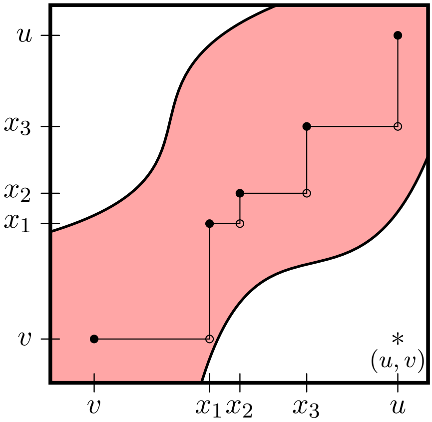
Visualizing Uncountably Infinite Graphs



Assume that U is an interval.

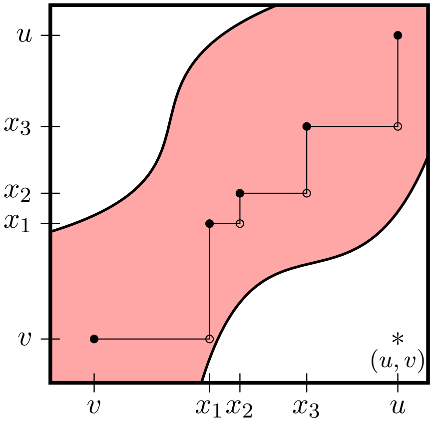
The region Ω can be interpreted as an "adjacency matrix". Two vertices $u, v \in U$ are adjacent if and only if $(u, v) \in \Omega$.

Visualizing Uncountably Infinite Graphs



Visualizing paths is tricky.

Visualizing Uncountably Infinite Graphs



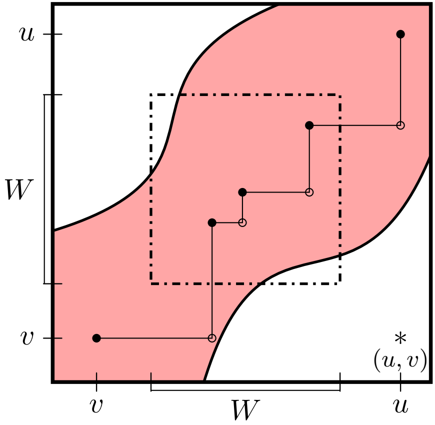
Visualizing paths is tricky. Consider for example the path

$$v, x_1, x_2, x_3, u.$$

The unfilled circles $\circ$ represent the edges (v, x_1), (x_1, x_2), (x_2, x_3) and (x_3, u). For a valid path they must lie within Ω.

On the other hand $*$ represents the non-edge (u, v).

Visualizing Uncountably Infinite Graphs

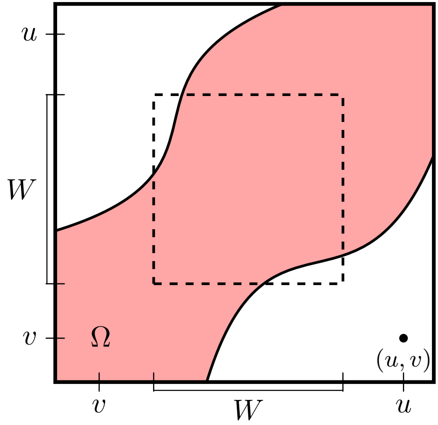


Separators are sets which intercept paths.

Here our path (v, x_1, x_2, x_3, u) passes through W because some of the points in the path lie in W. In fact, notice that this is true for all the paths from v to u in Ω.

Thus, W separates u and v in Ω.

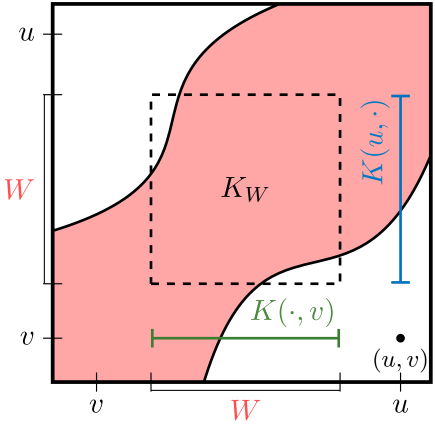
Continuously Indexed Graphical Model



We say that X has the graph $\Omega \subset U \times U$ if for every $u, v \in U$ separated by $W \subset U$ in Ω, it satisfies the *global Markov property*:

$$X_u \perp\!\!\!\perp X_v \mid X_W.$$

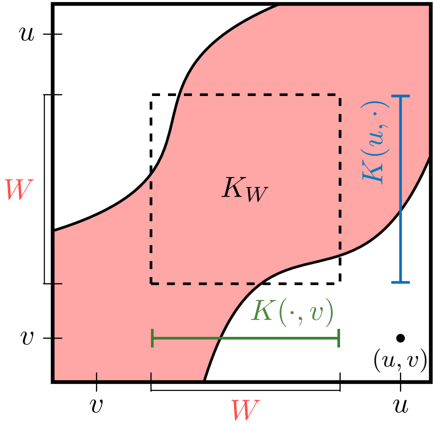
Continuously Indexed Graphical Model



We say that X has the graph $\Omega \subset U \times U$ if for every $u, v \in U$ separated by $W \subset U$ in Ω, it satisfies the *separation equation*:

$$K(u,v) = \langle K(u,\cdot), K(\cdot,v) \rangle_{\mathcal{H}(K_W)}$$

Continuously Indexed Graphical Model

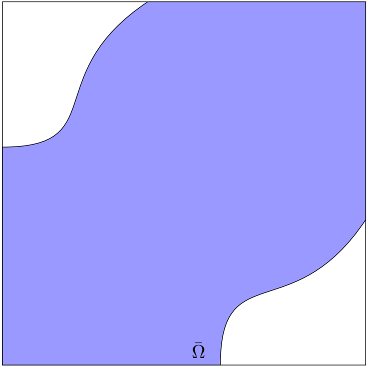


We say that X has the graph $\Omega \subset U \times U$ if for every $u, v \in U$ separated by $W \subset U$ in Ω, it satisfies the *separation equation*:

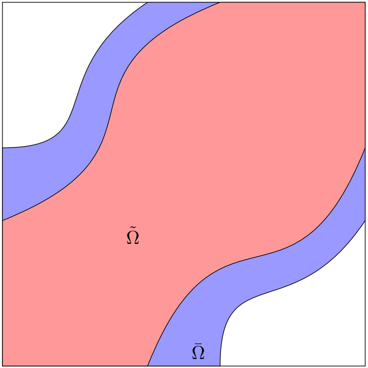
$$K(u,v) = \langle K(u,\cdot), K(\cdot,v) \rangle_{\mathcal{H}(K_W)}$$

A reproducing kernel characterization of conditional independence in a Gaussian process!

Covariance Recovery

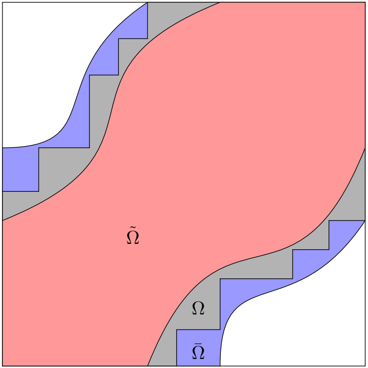


Covariance Recovery



- Assume that X admits a graphical structure with the graph $\tilde{\Omega}$, and we can estimate the covariance over $\bar{\Omega}$.

Covariance Recovery



- Assume that X admits a graphical structure with the graph $\tilde{\Omega}$, and we can estimate the covariance over $\bar{\Omega}$.
- If $\tilde{\Omega} \subset \Omega \subset \bar{\Omega}$ then it is possible to recover the covariance K of X from $K_{\bar{\Omega}}$.

More on Positive-Definite Completion

1. Using the canonical completion $K_\star$, one can characterize all completions of K_Ω.
2. One can derive more elegant expressions for $K_\star$:

$$K_\star(x, y) = -\frac{1}{2}[K_\Omega(x, x) + K_\Omega(y, y] + \sup_f \left[f(x) + f(y) - \frac{1}{2}\|f\|_*^2\right]$$

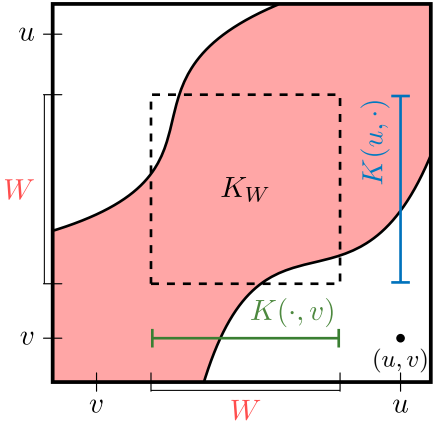
   where

$$\|f\|_*^2 = \|f_{I_1}\|^2 - \|f_{I_1 \cap I_2}\|^2 + \|f_{I_2}\|^2 - \cdots + \|f_{I_p}\|^2$$

3. The concept of conditional independence supplies the correct definition of a special solution of positive-definite completion!

Recovering Graph from Covariance

Recovering Uncountably Large Graphs



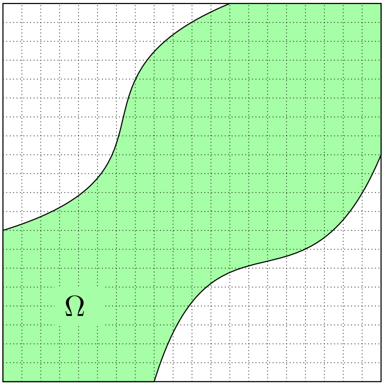
We say that X has the graph $\Omega \subset U \times U$ if for every $u, v \in U$ separated by $W \subset U$ in Ω, it satisfies the *separation equation*:

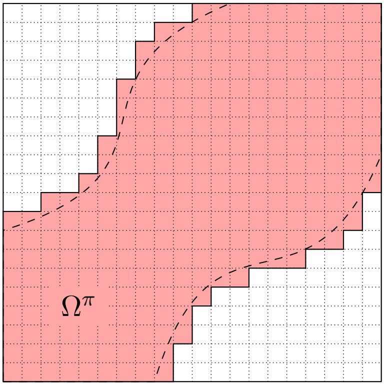
$$K(u,v) = \langle K(u,\cdot), K(\cdot,v) \rangle_{\mathcal{H}(K_W)}$$

This is not a very handy characterization unlike he multivariate case.

Resolving Uncountably Large Graphs

Let's partition the domain U into $\pi = \{U_1, \ldots, U_p\}$.





Instead of asking whether there is an edge between $u, v \in U$, we pose the related question of whether there is an edge between (some point in) U_i and (some point in) U_j.

Resolving Uncountably Large Graphs

Turns out Ω^π admits a nice inverse zero characterization!

For $1 \leq i, j \leq p$, let $\mathbf{K}_{ij} : L^2(U_j, \mu) \to L^2(U_i, \mu)$ be the integral operator induced by the integral kernel $K_{ij} = K|_{U_i \times U_j}$ 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

$$\mathbf{R}_{ij} = \mathbf{K}_{ii}^{-1/2} \mathbf{K}_{ij} \mathbf{K}_{jj}^{-1/2}.$$

Resolving Uncountably Large Graphs

The correlation operator matrix $\mathbf{R}_\pi$ behaves much better than $\mathbf{K}_\pi$. In fact, it is often invertible!

Theorem

Under some technical conditions, if $\mathbf{R}_\pi$ is invertible, then the graph Ω^π is related to the inverse $\mathbf{P}_\pi = \mathbf{R}_\pi^{-1}$ as follows:

$$\Omega^\pi \equiv \lim_{\epsilon \to 0} (\Omega + \mathbb{B}_\epsilon)^\pi = \cup \{U_i \times U_j : \|\mathbf{P}_{ij}\| \neq 0\}.$$

The $\subset$ statement holds even without the said technical condition.

Choosing a finer partition π yields a higher resolution version Ω^π of Ω.

Graphs of some processes

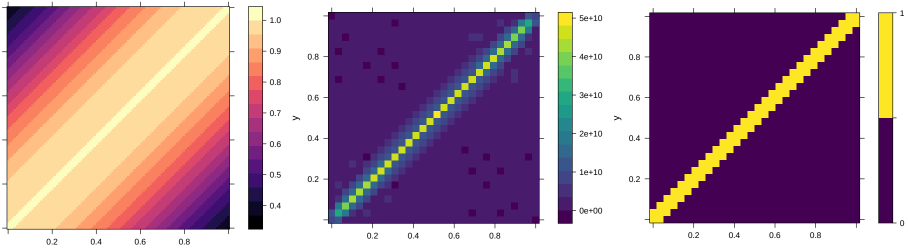


Figure: The covariance K, precision matrix $\mathbf{P}_\pi$ and the graph Ω^π of Gaussian kernel: $K(u, v) = \exp[-(u - v)^2]$.

Graphs of some processes

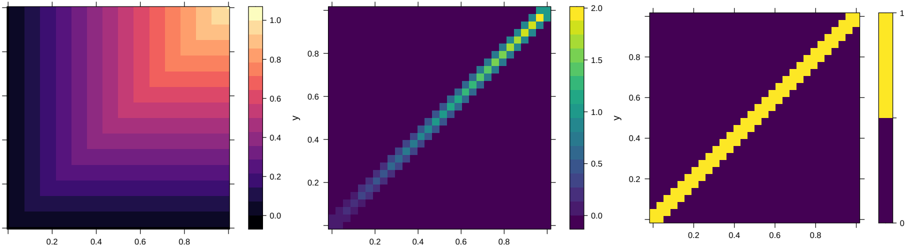


Figure: The covariance K, precision matrix $\mathbf{P}_\pi$ and the graph Ω^π of Brownian motion: $K(u,v) = \min(u,v)$.

Graphs of some processes

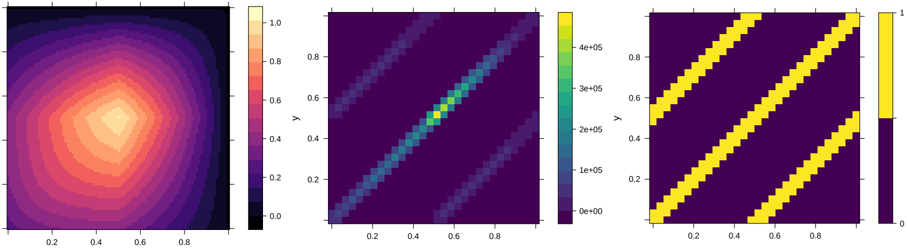


Figure: The covariance K, precision matrix $\mathbf{P}_\pi$ and the graph Ω^π of integrated Brownian motion: $K(u,v) = \mathrm{Cov}(X_u, X_v)$ where X_t is given by $X_t = \int_{\max(0,t-1/2)}^{t} B_u \, du$.

Publications

1. **The Completion of Covariance Kernels.**
   Kartik G. Waghmare and Victor M. Panaretos.
   The Annals of Statistics, 50.6 (2022), pp. 3281–33.
2. **Continuously Indexed Graphical Models.**
   Kartik G. Waghmare and Victor M. Panaretos.
   Journal of the Royal Statistical Society: Series B (2024).
3. **The Positive-Definite Completion Problem.**
   Kartik G. Waghmare and Victor M. Panaretos.
   Transactions of the American Mathematical Society (2024).

References I

- Calderón, A. and R Pepinsky (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, pp. 339–348.
- Carathéodory, C. (1907). "Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen". In: *Math. Ann.* 64, pp. 95–115.
- Delaigle, A. and P. Hall (2016). "Approximating fragmented functional data by segments of Markov chains". In: *Biometrika* 103.4, pp. 779–799.
- Delaigle, A. et al. (2021). "Estimating the Covariance of Fragmented and Other Related Types of Functional Data". In: *Journal of the American Statistical Association* 116.535, pp. 1383–1401.

References II

- Descary, M.H. and V.M. Panaretos (2019). "Functional data analysis by matrix completion". In: *The Annals of Statistics* 47.1, pp. 1–38.
- Dym, H. and I. Gohberg (1981). "Extensions of band matrices with band inverses". In: *Linear algebra and its applications* 36, pp. 1–24.
- Grone, R. et al. (1984). "Positive definite completions of partial Hermitian matrices". In: *Linear algebra and its applications* 58, pp. 109–124.
- Krein, M. G. (1940). "Sur le probleme du prolongement des fonctions hermitiennes positives et continues". In: *CR (Doklady) Acad. Sci. URSS (NS)* 26.1, pp. 17–22.
- Lin, Z., J.L. Wang, and Q. Zhong (2021). "Basis expansions for functional snippets". In: *Biometrika* 108.3, pp. 709–726.
- Rudin, W. (1963). "The extension problem for positive-definite functions". In: *Illinois J. Math.* 7, pp. 532–539.