MULTILEVEL APPROXIMATION OF GAUSSIAN RANDOM FIELDS:

COVARIANCE COMPRESSION, ESTIMATION AND SPATIAL PREDICTION

Christoph Schwab
Seminar for Applied Mathematics (SAM)
ETH Zürich, Switzerland

Helmut Harbrecht, University of Basle, Switzerland,
Lukas Herrmann, RICAM, Linz, Austria,
Kristin Kirchner, TU Delft, The Netherlands

BIRS Workshop on Statistical Aspects of Non-Linear Inverse Problems, November 2021

Layout

- Synopsis
- Gaussian Random Fields (GRFs) on Manifolds $\mathcal{M}$
- Multi Resolution Analysis on $\mathcal{M}$
- Covariance and Precision Operator Preconditioning
- Covariance Tapering via Matrix Compression
- Sampling through Multilevel Path Simulation
- Multilevel Monte-Carlo Covariance Estimation
- Sparse Spatial Prediction ('Kriging')
- Numerical examples
- References

Synopsis

- **Wavelet Compression** of [Matrix Representations of] **Pseudodifferential Operators** leveraged for **Optimal Numerical Covariance Matrix Tapering, Estimation, Kriging**
- Multilevel MC Covariance Estimation from Samples of GRF $\mathcal{Z}$
- Multilevel MC Path Simulation of GRF $\mathcal{Z}$
- Multilevel Kriging Algorithm
- Optimal (linear) complexities of these algorithms
- References:

1. **L. Herrmann, K. Kirchner, and ChS**
   Multilevel approximation of Gaussian Random Fields: fast simulation.
   Math. Models Methods Appl. Sci. **30**(1):181-223, 2020.
2. **H. Harbrecht, L. Herrmann, K. Kirchner, and ChS**
   Multilevel Approximation of Gaussian Random Fields:
   Covariance Compression, Estimation and Spatial Prediction
   arXiv:2103.04424

Gaussian Random Fields (GRFs) on Manifolds $\mathcal{M}$

- $\mathcal{M}$ closed, bounded orientable Riemannian Manifold, $n = \dim(\mathcal{M})$, $\partial\mathcal{M} = \varnothing$.
- $(\Omega, \mathcal{F}, \mathbb{P})$ probability space, $(\mathcal{M}, \mathscr{B}(\mathcal{M}))$ Borel sets
- $(\mathcal{Z}(x))_{x\in\mathcal{M}}$ family of $\mathcal{F}$-measurable $\mathbb{R}$-valued RVs:

$$\forall \ \{x_1, , \ldots, x_m\} \subset \mathcal{M} : \ (\mathcal{Z}(x_1), \ldots, \mathcal{Z}(x_m))^\top \in \mathbb{R}^m \text{ centered Gaussian}$$

$$\mathcal{Z} : \mathcal{M} \times \Omega \to \mathbb{R} \quad \mathscr{B}(\mathcal{M}) \otimes \mathcal{F} - \text{measurable.}$$

- sPDE:

$$\mathcal{A}\mathcal{Z} = \mathcal{W} \quad \text{on} \quad \mathcal{M}.$$

  $\mathcal{W}$ white noise on $L^2(\mathcal{M})$: $L^2(\mathcal{M})$-valued, weak random var. with

$$L^2(\mathcal{M}) \ni \varphi \mapsto \mathbb{E}[\exp(i(\varphi, \mathcal{W})_{L^2(\mathcal{M})}] = \exp(-\frac{1}{2}\|\varphi\|^2_{L^2(\mathcal{M})}) \ ,$$

  $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ elliptic, s.a., order $\hat{r} > n/2$ "coloring" operator

- $\mathcal{Z}$ centered, Covariance Operator $\mathcal{C} : L^2(\mathcal{M}) \to L^2(\mathcal{M})$ given by

$$(\mathcal{C}v, w)_{L^2(\mathcal{M})} = \mathbb{E}\big[(\mathcal{Z}, v)_{L^2(\mathcal{M})}(\mathcal{Z}, w)_{L^2(\mathcal{M})}\big] \quad \forall v, w \in L^2(\mathcal{M}).$$

- $\mathcal{W} \in H^{-n/2-\varepsilon}(\mathcal{M})$ ($\mathbb{P}$-a.s.) for any $\varepsilon > 0$ implies

$$\mathcal{Z} \in H^s(\mathcal{M}), \quad \text{for every} \quad s < \hat{r} - n/2 \quad (\mathbb{P}\text{-a.s.}),$$

Sample paths of GRF $\mathcal{Z}$ on sphere $\mathcal{M} = \mathbb{S}^2$

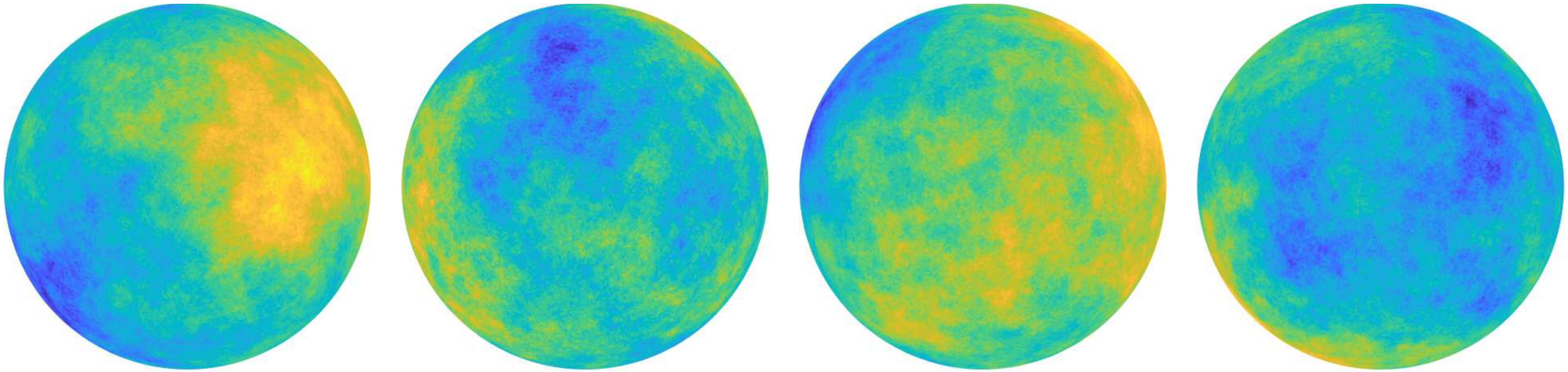

Figure 1: Four realizations of a Gaussian random field on $\mathbb{S}^2$ for the Matèrn covariance $k_{1/2}$ with respect to the geodesic distance.

- for any $q \in (0, \infty)$, $0 \leq s < \hat{r} - n/2$,

$$\mathbb{E}\left[\|\mathcal{Z}\|^q_{H^s(\mathcal{M})}\right] < \infty.$$

Whittle–Matérn models:
$\mathcal{A} = (\mathcal{L} + \kappa^2)^\beta$, with $\mathcal{L} \in OPS^{\bar{r}}_{1,0}(\mathcal{M})$ for some $\beta, \bar{r} > 0$.
$\kappa \in C^\infty(\mathcal{M})$: local correlation scale of GRF $\mathcal{Z}$. $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ with $\hat{r} = \beta\bar{r} > 0$.
No stationarity, isotropy, ... (no circulant embedding, etc for fast simulation).
Example: $\mathcal{L} = -\nabla_\mathcal{M} \cdot a(x)\nabla_\mathcal{M} \in OPS^2_{1,0}(\mathcal{M})$ implies $\mathcal{A} \in OPS^{2\beta}_{1,0}(\mathcal{M})$

Covariance and Precision Operator

Assumption:

1. $\mathcal{M}$ smooth, closed, bounded and connected orientable Riemannian manifold of dimension n.
2. $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ for some $\hat{r} > n/2$, self-adjoint and positive: ex. $a_- > 0$ such that

$$\forall w \in H^{\hat{r}/2}(\mathcal{M}) : \quad \langle \mathcal{A}w, w \rangle \geq a_- \|w\|^2_{H^{\hat{r}/2}(\mathcal{M})}.$$

Proposition:
Let $\hat{r} > n/2$ and $\mathcal{M}$ and $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ satisfy **Assumption** . Then:

1. Covariance Operator $\mathcal{C}$ of GRF $\mathcal{Z}$ is
$$\mathcal{C} = \mathcal{A}^{-2} \in OPS^{-2\hat{r}}_{1,0}(\mathcal{M}) \; .$$
For every $s \in \mathbb{R}$, $\mathcal{C}\colon H^s(\mathcal{M}) \to H^{s+2\hat{r}}(\mathcal{M})$ is an isomorphism.
2. $\mathcal{C}$ self-adjoint, (strictly) positive definite and compact on $L^2(\mathcal{M})$, trace-class.
3. Precision operator $\mathcal{P}$ of GRF $\mathcal{Z}$ is
$$\mathcal{P} = \mathcal{A}^2 \in OPS^{2\hat{r}}_{1,0}(\mathcal{M})$$
For every $s \in \mathbb{R}$, $\mathcal{P}\colon H^s(\mathcal{M}) \to H^{s-2\hat{r}}(\mathcal{M})$ isomorphism.
4. $\mathcal{P} = \mathcal{A}^2$ self-adjoint, positive definite, unbounded on $L^2(\mathcal{M})$, spectrum discrete, accumulates only at ∞.

Multiresolution Analysis on $\mathcal{M}$

Multiresolution ("wavelet") Analysis (MRAs):

80/90ies: Signal processing (R. Coifman, I. Daubecies, Y. Meyer), Operator Equations (Y. Meyer)
90/00ies: Finite Elements, Integral Operators (W. Dahmen, R. Schneider, R. Stevenson)
Here: Use **MRAs on** $\mathcal{M}$ to **optimally precondition and compress** $\mathcal{C}$ and $\mathcal{P}$

- $\{V_j\}_{j>j_0}$ nested, linear subspaces $V_j \subset V_{j+1} \subset \ldots \subset L^2(\mathcal{M})$.
- $\{V_j\}_{j>j_0}$ has *regularity* $\gamma > 0$ and *(approximation) order* $d \in \mathbb{N}$ if

$$\gamma = \sup\{s \in \mathbb{R} : V_j \subset H^s(\mathcal{M})\ \forall j > j_0\},$$
$$d = \sup\Big\{s \in \mathbb{R} : \inf_{v_j \in V_j} \|v - v_j\|_{L^2(\mathcal{M})} \lesssim 2^{-js}\|v\|_{H^s(\mathcal{M})}\ \forall v \in H^s(\mathcal{M})\ \forall j > j_0\Big\}.$$

- $\{V_j\}_{j>j_0}$ $H^{r/2}(\mathcal{M})$-conforming, i.e.,

$$\gamma > \max\{0, r/2\} \text{ for some fixed order } r \in \mathbb{R}.$$

- $\dim(V_j) = \mathcal{O}(2^{nj})$,

$$\forall j > j_0 : \quad V_j = \operatorname{span} \mathbf{\Phi}_j, \quad \text{where} \quad \mathbf{\Phi}_j := \{\varphi_{j,k} : k \in \Delta_j\} \text{ (Single Scale Basis).}$$

- *Dual single-scale bases* defined by the *biorthogonality relation*

$$\forall j > j_0 : \quad \widetilde{\mathbf{\Phi}}_j := \{\widetilde{\varphi}_{j,k} : k \in \Delta_j\}, \quad \text{with} \quad \langle \varphi_{j,k}, \widetilde{\varphi}_{j,k'} \rangle = \delta_{k,k'} \quad \forall k, k' \in \Delta_j.$$

Multiresolution Analysis on $\mathcal{M}$

- Projector $Q_j : L^2(\mathcal{M}) \to V_j$:

$$\forall v \in L^2(\mathcal{M}) : \quad Q_j v := \sum_{k \in \Delta_j} \langle v, \widetilde{\varphi}_{j,k} \rangle \varphi_{j,k}.$$

- Biorthogonal Complement Basis: set $\nabla_j := \Delta_{j+1} \backslash \Delta_j$.

$$\mathbf{\Psi}_j = \{\psi_{j,k} : k \in \nabla_j\} \quad \text{and} \quad \widetilde{\mathbf{\Psi}}_j = \{\widetilde{\psi}_{j,k} : k \in \nabla_j\}, \qquad j > j_0,$$

- *biorthogonality relation*:

$$\langle \psi_{j,k}, \widetilde{\psi}_{j',k'} \rangle = \delta_{(j,k),(j',k')} = \begin{cases} 1, & \text{if } j = j' \text{ and } k = k', \\ 0, & \text{otherwise,} \end{cases}$$

- Local, isotropic supports: $\operatorname{diam}(\operatorname{supp} \psi_{j,k}) \simeq 2^{-j}, \quad j > j_0,$
- Biorthogonality:

$$\forall j > j_0 : \quad V_{j+1} = W_j \oplus V_j, \quad \widetilde{V}_{j+1} = \widetilde{W}_j \oplus \widetilde{V}_j, \quad \widetilde{V}_j \perp W_j, \quad V_j \perp \widetilde{W}_j.$$

  Convention $W_{j_0} := V_{j_0+1}, \quad \widetilde{W}_{j_0} := \widetilde{V}_{j_0+1}, \quad$ and $\quad \mathbf{\Psi}_{j_0} := \mathbf{\Phi}_{j_0+1}, \quad \widetilde{\mathbf{\Psi}}_{j_0} := \widetilde{\mathbf{\Phi}}_{j_0+1}.$

- Biorthogonal $\mathbf{\Psi}, \widetilde{\mathbf{\Psi}}$ wavelet bases (primal, resp. dual, *multiresolution analysis* (MRAs)).

$$\mathbf{\Psi} = \bigcup_{j \geq j_0} \mathbf{\Psi}_j, \qquad \widetilde{\mathbf{\Psi}} = \bigcup_{j \geq j_0} \widetilde{\mathbf{\Psi}}_j.$$

Multiresolution Analysis on $\mathcal{M}$

- **Equivalent, bi-infinite Matrix Representations of $\mathcal{C}$ and $\mathcal{P}$:**

$$\mathbf{C} = \mathcal{C}(\mathbf{\Psi})(\mathbf{\Psi}) \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}} \quad \text{and} \quad \mathbf{P} = \mathcal{P}(\mathbf{\Psi})(\mathbf{\Psi}) \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$$

- **Vanishing Moment Property**: ($\rightarrow$ Covariance operator compression)

$$|\langle v, \psi_{j,k}\rangle| \lesssim 2^{-j(\widetilde{d}+n/2)} \sup_{|\alpha|=\widetilde{d},\, x\in \operatorname{supp}(\psi_{j,k})} |\partial^\alpha v(x)| \quad \forall (j,k) \in \mathcal{J} := \{(j,k) : j \geq j_0,\ k \in \nabla_j\},$$

- **Norm Equivalences**: ($\rightarrow$ optimal diagonal preconditioning) $\mathbf{\Psi}, \widetilde{\mathbf{\Psi}}$ Riesz bases in scale $H^t(\mathcal{M})$

$$\|v\|^2_{H^t(\mathcal{M})} \simeq \sum_{j\geq j_0} \sum_{k\in\nabla_j} 2^{2jt} |\langle v, \widetilde{\psi}_{j,k}\rangle|^2, \qquad t \in (-\widetilde{\gamma}, \gamma),$$

$$\|v\|^2_{H^t(\mathcal{M})} \simeq \sum_{j\geq j_0} \sum_{k\in\nabla_j} 2^{2jt} |\langle v, \psi_{j,k}\rangle|^2, \qquad t \in (-\gamma, \widetilde{\gamma}).$$

- **Index Set**: $\mathcal{J} = \{(j,k) : j \geq j_0, k \in \nabla_j\}$, j-scale parameter, k localization parameter

Covariance and Precision Operator Preconditioning

Diagonal scaling matrix:

$$\mathbf{D}^s := \mathrm{diag}\big(2^{s|\lambda|} : \lambda \in \mathcal{J}\big), \quad s \in \mathbb{R}.$$

Theorem [Optimal Covariance and Precision Matrix Preconditioning]

If $\mathcal{M}$ and $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ satisfy **Assumption** and $\tilde{\gamma} > r$, $\gamma > 0$, then

(C1) $\mathbf{C} = \mathcal{C}(\mathbf{\Psi})(\mathbf{\Psi})$ self-adjoint, positive, compact operator on $\ell^2(\mathcal{J})$:

Ex. $0 < c_- \leq c_+ < \infty$ such that $\sigma(\mathbf{D}^{\hat{r}}\mathbf{C}\mathbf{D}^{\hat{r}}) \subset [c_-, c_+]$ and $\mathrm{cond}_2(\mathbf{D}^{\hat{r}}\mathbf{C}\mathbf{D}^{\hat{r}}) \simeq 1$,

(C2) for every $\Lambda \subset \mathcal{J}$ with $p = \#(\Lambda) < \infty$, $\mathbf{C}_\Lambda = \{\mathbf{C}_{\lambda,\lambda'} : \lambda, \lambda' \in \Lambda\} \in \mathbb{R}^{p\times p}$ is SPD and

$$\sigma(\mathbf{D}^{\hat{r}}_\Lambda \mathbf{C}_\Lambda \mathbf{D}^{\hat{r}}_\Lambda) \subset [c_-, c_+], \qquad \mathbf{D}^{\hat{r}}_\Lambda := \{\mathbf{D}^{\hat{r}}_{\lambda,\lambda'} : \lambda, \lambda' \in \Lambda\} \in \mathbb{R}^{p\times p}.$$

(P1) $\mathbf{P} = \mathcal{P}(\mathbf{\Psi})(\mathbf{\Psi})$ self-adjoint, positive, unbounded operator on $\ell^2(\mathcal{J})$.

Ex. $0 < c_- \leq c_+ < \infty$ such that $\sigma(\mathbf{D}^{-\hat{r}}\mathbf{P}\mathbf{D}^{-\hat{r}}) \subset [c_-, c_+]$ and $\mathrm{cond}_2(\mathbf{D}^{-\hat{r}}\mathbf{P}\mathbf{D}^{-\hat{r}}) \simeq 1$,

(P2) for every $\Lambda \subseteq \mathcal{J}$ with $p = \#(\Lambda) < \infty$, $\mathbf{P}_\Lambda$ is SPD and $\sigma(\mathbf{D}^{-\hat{r}}_\Lambda \mathbf{P}_\Lambda \mathbf{D}^{-\hat{r}}_\Lambda) \subset [c_-, c_+]$.

Remark: Λ: "tapering pattern", $p = \#(\Lambda) < \infty$ number of "graphical" parameters.

Compression Estimates

Proposition [CZ Estimates]

- if $\mathcal{B} \in OPS^r_{1,0}(\mathcal{M})$, mutually biorthogonal MRAs $\mathbf{\Psi}$, $\widetilde{\mathbf{\Psi}}$, with $n+r+2\widetilde{d} > 0$ in local coordinates on $\mathcal{M}$,
- and $\mathcal{M}$ fulfills **Assumption**,

then, with the **supports**

$$S_{j,k} := \operatorname{conv\,hull}(\operatorname{supp}(\psi_{j,k})) \subset \mathcal{M}, \qquad S'_{j,k} := \operatorname{sing\,supp}(\psi_{j,k}) \subset \mathcal{M},$$

bi-infinite matrix $\mathbf{B} = \mathcal{B}(\mathbf{\Psi})(\mathbf{\Psi}) \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ has entries which admit, uniformly in $j \in \mathbb{N}$:

(i) $\forall (j,k), (j',k') \in \mathcal{J}$ s.t. $S_{j,k} \cap S_{j',k'} = \varnothing$, we have

$$|\langle \mathcal{B}\psi_{j',k'}, \psi_{j,k}\rangle| \lesssim 2^{-(j+j')(\widetilde{d}+n/2)} \operatorname{dist}(S_{j,k}, S_{j',k'})^{-(n+r+2\widetilde{d})},$$

(ii) $\forall (j,k), (j',k') \in \mathcal{J}$ s.t. $\operatorname{dist}(S'_{j,k}, S_{j',k'}) \gtrsim 2^{-j'}$, we have

$$|\langle \mathcal{B}\psi_{j',k'}, \psi_{j,k}\rangle| + |\langle \mathcal{B}\psi_{j,k}, \psi_{j',k'}\rangle| \lesssim 2^{jn/2} 2^{-j'(\widetilde{d}+n/2)} \operatorname{dist}(S'_{j,k}, S_{j',k'})^{-(r+\widetilde{d})}.$$

Apply for $\mathcal{B} \in \{\mathcal{A}, \mathcal{A}^{-1}, \mathcal{C}, \mathcal{P}\}$.

Compression Estimates

Definition [Tapering Strategy]

A-priori matrix compression: *block truncation (or "tapering") parameters* $\{\tau'_{jj'}, \tau_{jj'} : j_0 \le j, j' \le J\}$:

$$[\mathbf{B}_p^\varepsilon]_{\lambda,\lambda'} := \begin{cases} 0 & \mathrm{dist}(S_\lambda, S_{\lambda'}) > \tau_{jj'} \text{ and } j, j' > j_0, \\ 0 & \mathrm{dist}(S_\lambda, S_{\lambda'}) \le 2^{-\min\{j,j'\}} \text{ and } \mathrm{dist}(S'_\lambda, S_{\lambda'}) > \tau'_{jj'} \text{ if } j' > j \ge j_0, \\ & \text{ and } \mathrm{dist}(S_\lambda, S'_{\lambda'}) > \tau'_{jj'} \text{ if } j > j' \ge j_0, \\ \langle \mathcal{B}\psi_{\lambda'}, \psi_\lambda \rangle & \text{otherwise.} \end{cases}$$

Here, with **fixed, real-valued constants**

$$a, a' > 1 \text{ sufficiently large and } \quad d < d' < \widetilde{d} + r,$$

Truncation ("Tapering") Parameters $\tau_{jj'}$ and $\tau'_{jj'}$

$$\tau_{jj'} := a \max \left\{ 2^{-\min\{j,j'\}}, 2^{[2J(d'-r/2)-(j+j')(d'+\widetilde{d})]/(2\widetilde{d}+r)} \right\},$$
$$\tau'_{jj'} := a' \max \left\{ 2^{-\max\{j,j'\}}, 2^{[2J(d'-r/2)-(j+j')d'-\max\{j,j'\}\widetilde{d}]/(\widetilde{d}+r)} \right\}.$$

$\mathcal{B}_p^\varepsilon$: Operator corresponding to tapered matrix $\mathbf{B}_p^\varepsilon = \mathcal{B}_p^\varepsilon(\mathbf{\Psi})(\mathbf{\Psi})$.

Numerical Illustration (Matérn Covariance, $n = 1$)

$\mathcal{M}$ boundary of domain $D \subset \mathbb{R}^2$ given by the 2π-periodic, analytic parametrization

$$\gamma : [0, 2\pi] \to \mathcal{M} = \partial D, \quad \gamma(\varphi) = g(\varphi) \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \end{bmatrix},$$

where

$$g(\varphi) = \alpha_0 + \frac{1}{100} \sum_{k=1}^{5} \big(\alpha_{-k} \sin(k\varphi) + \alpha_k \cos(k\varphi)\big)$$

finite Fourier series with the following coefficients:

$$\begin{array}{llllll} \alpha_{-5} = 2.2, & \alpha_{-4} = 0.56, & \alpha_{-3} = 0.14, & \alpha_{-2} = 1.1, & \alpha_{-1} = 1.4, & \alpha_0 = 50, \\ \alpha_5 = 0.89, & \alpha_4 = -1.5, & \alpha_3 = -1.2, & \alpha_2 = -1.5, & \alpha_1 = -0.57. & \end{array}$$

Covariance kernels: Matérn family

$$k_\nu(z) = \frac{2^{1-\nu}\sigma^2}{\Gamma(\nu)} \big(\sqrt{2\nu}\frac{z}{\ell}\big)^\nu K_\nu\big(\sqrt{2\nu}\frac{z}{\ell}\big),$$

with $\sigma^2 = 1$, as a product of an exponential and a polynomial for $\nu = q - 1/2$ with $q \in \mathbb{N}$.

$$k_{1/2}(z) = \exp\left(-\frac{z}{\ell}\right), \quad k_{3/2}(z) = \left(1 + \frac{\sqrt{3}z}{\ell}\right) \exp\left(-\frac{\sqrt{3}z}{\ell}\right), \quad k_{5/2}(z) = \left(1 + \frac{\sqrt{5}}{\ell}z + \frac{5}{3\ell}z^2\right) \exp\left(-\frac{\sqrt{5}z}{\ell}\right),$$

$z = \|x - y\|_2$ for $x, y \in \Gamma$, $\ell > 0$ spatial correlation legth.

Numerical Illustration (Matérn Covariance, $n = 1$)

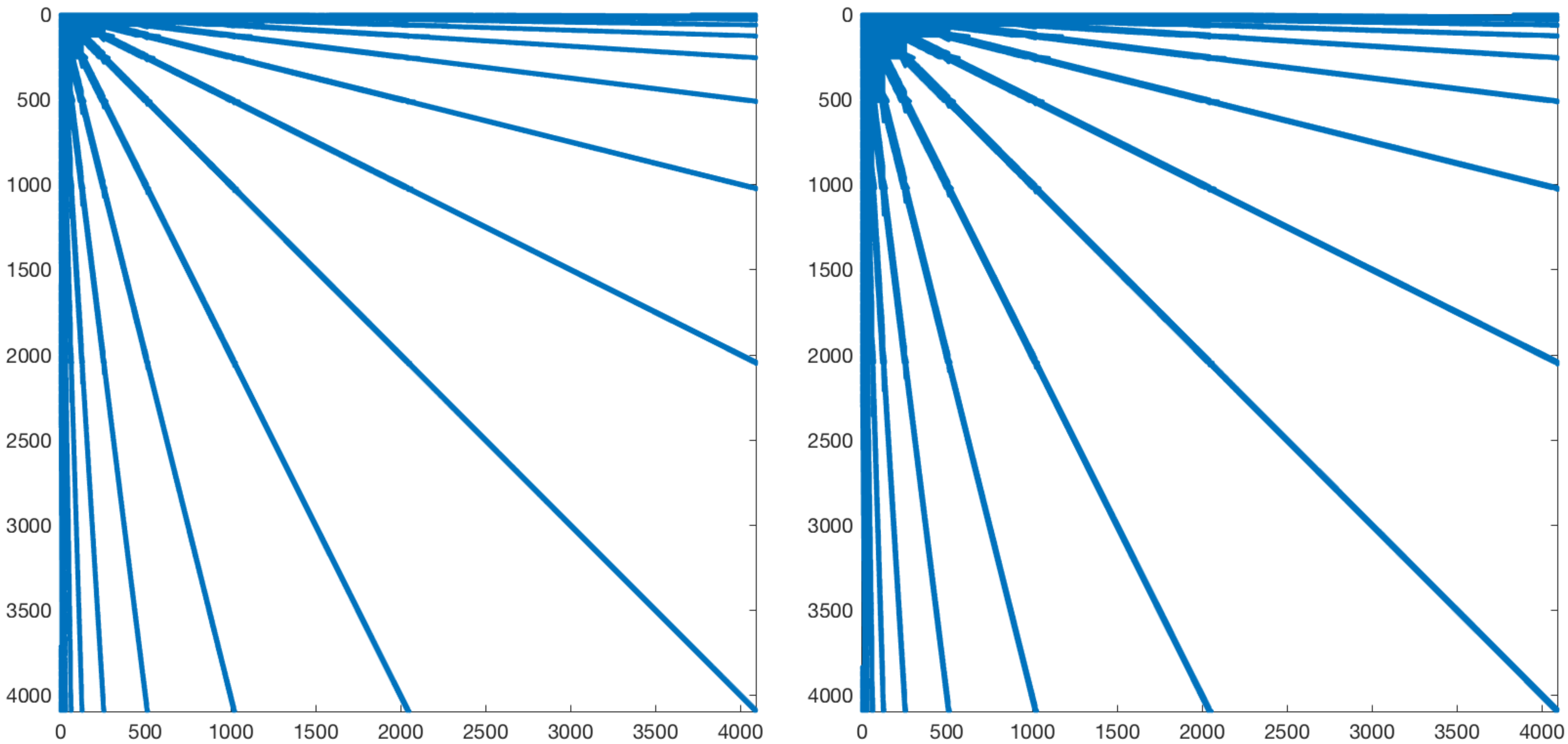

Figure 2: A-priori compression pattern for $p = 4096$ wavelets in case of the Matérn covariance kernel $k_{1/2}$ and $\boldsymbol{\Psi}^{(2,6)}$ (*left*) and in case of the Matérn covariance kernel $k_{3/2}$ and $\boldsymbol{\Psi}^{(2,8)}$ (*right*). In the left and right matrix, only 5.0 % and 6.8 % of the matrix coefficients are relevant, respectively.

Covariance and Precision Operator Compression

Theorem [Covariance Matrix Compression]

- $\mathcal{M}$ and $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ satisfy **Assumption** for some $\hat{r} > n/2$, $n = \dim(\mathcal{M})$.
- $\mathbf{\Psi}$ MRA, Norm equivalences with $\widetilde{\gamma} > \hat{r}$ and $\gamma > 0$.
- $\mathcal{C} = \mathcal{A}^{-2}$ covariance operator of GRF $\mathcal{Z}$, tapered covariance matrix by $\mathbf{C}^{\varepsilon}_{p(J)}$, and covariance tapering parameters $\{\tau_{jj'}(\mathcal{C}), \tau'_{jj'}(\mathcal{C}) : j_0 \le j, j' \le J\}$, with $-2\hat{r}$ in place of r.

Then, ex. $\varepsilon_0 > 0$ s.t. for every $\varepsilon \in (0, \varepsilon_0)$, ex. $a, a' > 0$ independent of $p(J)$ with:

(i) For every $J \ge j_0$, tapered matrix $\mathbf{C}^{\varepsilon}_{p(J)}$ is SPD.

(ii) Diag. Precond. renders $\mathbf{C}^{\varepsilon}_{p(J)}$ **uniformly well-conditioned**:

$$\exists 0 < \widetilde{c}_- \le \widetilde{c}_+ < \infty \text{ s.t. } \forall J \ge j_0 : \quad \sigma\big(\mathbf{D}^{\hat{r}}_{p(J)} \mathbf{C}^{\varepsilon}_{p(J)} \mathbf{D}^{\hat{r}}_{p(J)}\big) \subset [\widetilde{c}_-, \widetilde{c}_+].$$

(iii) $\{\mathbf{C}^{\varepsilon}_{p(J)}\}_{J \ge j_0}$ **optimally sparse**: as $J \to \infty$, $\#\mathrm{nnz}(\mathbf{C}^{\varepsilon}_{p(J)})$ is $\mathcal{O}(p(J))$.

(iv) $\{\mathbf{C}^{\varepsilon}_{p(J)}\}_{J \ge j_0}$ **optimally consistent**:

$$\forall J \ge j_0, v \in H^{t'}(\mathcal{M}), w \in H^t(\mathcal{M}) : \quad \big|\big\langle \big(\mathcal{C} - \mathcal{C}^{\varepsilon}_{p(J)}\big) Q_J w, Q_J v \big\rangle\big| \lesssim \varepsilon 2^{J(-2\hat{r}-t-t')} \|w\|_{H^t(\mathcal{M})} \|v\|_{H^{t'}(\mathcal{M})}.$$

Covariance and Precision Operator Compression

Theorem [Precision Matrix Compression]

- $\mathcal{M}$, $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$ satisfy **Assumption**, for some $\hat{r} > n/2$.
- $\boldsymbol{\Psi}$ MRA with $\gamma > \hat{r}$ and $\widetilde{\gamma} > 0$.
  $\mathcal{P} = \mathcal{A}^2$ **precision operator** of GRF $\mathcal{Z}$, **tapered precision matrix** $\mathbf{P}^{\varepsilon}_{p(J)}$,
  tapering parameters $\{\tau_{jj'}(\mathcal{P}), \tau'_{jj'}(\mathcal{P}) : j_0 \leq j, j' \leq J\}$, **with** $2\hat{r}$ **in place of** r.

Then, exists $\varepsilon_0 > 0$ such that, for every $\varepsilon \in (0, \varepsilon_0)$, ex $a, a' > 0$ independent of $p(J)$, such that

(i) For every $J \geq j_0$, tapered matrix $\mathbf{P}^{\varepsilon}_{p(J)}$ is SPD.

(ii) Diagonal preconditioning renders $\mathbf{P}^{\varepsilon}_{p(J)}$ uniformly well-conditioned: ex. $0 < \widetilde{c}_- \leq \widetilde{c}_+ < \infty$ such that

$$\forall J \geq j_0 : \quad \sigma\big(\mathbf{D}^{-\hat{r}}_{p(J)} \mathbf{P}^{\varepsilon}_{p(J)} \mathbf{D}^{-\hat{r}}_{p(J)}\big) \subset [\widetilde{c}_-, \widetilde{c}_+].$$

(iii) $\{\mathbf{P}^{\varepsilon}_{p(J)}\}_{J \geq j_0}$ **optimally sparse**: as $J \to \infty$, $\#\mathrm{nnz}(\mathbf{P}^{\varepsilon}_{p(J)}) = \mathcal{O}(p(J))$.

(iv) $\mathcal{P}^{\varepsilon}_{p(J)}(\boldsymbol{\Psi})(\boldsymbol{\Psi}) = \mathbf{P}^{\varepsilon}_{p(J)}$. Assume $\hat{r} \leq t, t' \leq d < \widetilde{d} + 2\hat{r}$. Then

$$\forall J \geq j_0, v \in H^{t'}(\mathcal{M}), w \in H^t(\mathcal{M}) : \quad \big|\big\langle \big(\mathcal{P} - \mathcal{P}^{\varepsilon}_{p(J)}\big) Q_J w, Q_J v \big\rangle\big| \lesssim \varepsilon 2^{J(2\hat{r} - t - t')} \|w\|_{H^t(\mathcal{M})} \|v\|_{H^{t'}(\mathcal{M})}.$$

$k_{1/2}$								
p	J	single-scale	nnz	$\mathbf{\Psi}^{(2,4)}$	nnz	$\mathbf{\Psi}^{(2,6)}$	nnz	$\mathbf{\Psi}^{(2,8)}$
32	5	$2.6 \cdot 10^3$	100	$2.4 \cdot 10^2$	100	$1.8 \cdot 10^2$	100	$6.6 \cdot 10^2$
64	6	$1.1 \cdot 10^4$	80	$2.7 \cdot 10^2$	88	$1.9 \cdot 10^2$	98	$6.7 \cdot 10^2$
128	7	$4.5 \cdot 10^4$	60	$3.1 \cdot 10^2$	65	$1.9 \cdot 10^2$	71	$6.8 \cdot 10^2$
256	8	$1.9 \cdot 10^5$	40	$3.4 \cdot 10^2$	42	$1.9 \cdot 10^2$	48	$6.8 \cdot 10^2$
512	9	$7.6 \cdot 10^5$	25	$3.7 \cdot 10^2$	26	$1.9 \cdot 10^2$	30	$6.8 \cdot 10^2$
1024	10	$3.1 \cdot 10^6$	16	$3.9 \cdot 10^2$	16	$1.9 \cdot 10^2$	18	$6.8 \cdot 10^2$
2048	11	$1.2 \cdot 10^7$	9.4	$4.0 \cdot 10^2$	9.0	$1.9 \cdot 10^2$	10	$6.8 \cdot 10^2$
4096	12	$5.0 \cdot 10^7$	5.0	$4.2 \cdot 10^2$	5.0	$1.9 \cdot 10^2$	5.7	$6.8 \cdot 10^2$

Table 1: Condition numbers and compression rates in case of the Matérn covariance kernel $k_{1/2}$. The compression rates validate the asymptotically linear behaviour. The condition numbers stay bounded for $\mathbf{\Psi}^{(2,6)}$ and $\mathbf{\Psi}^{(2,8)}$, whereas for $\mathbf{\Psi}^{(2,4)}$ a slight increase is observed.

$k_{3/2}$								
p	J	single-scale	nnz	$\mathbf{\Psi}^{(2,6)}$	nnz	$\mathbf{\Psi}^{(2,8)}$	nnz	$\mathbf{\Psi}^{(2,10)}$
32	5	$3.2 \cdot 10^5$	100	$2.3 \cdot 10^3$	100	$1.9 \cdot 10^4$	100	$1.9 \cdot 10^4$
64	6	$5.8 \cdot 10^6$	91	$3.3 \cdot 10^3$	98	$2.3 \cdot 10^4$	100	$2.0 \cdot 10^4$
128	7	$1.1 \cdot 10^8$	69	$4.9 \cdot 10^3$	75	$2.5 \cdot 10^4$	79	$2.0 \cdot 10^4$
256	8	$1.9 \cdot 10^9$	48	$6.9 \cdot 10^3$	51	$2.6 \cdot 10^4$	55	$2.0 \cdot 10^4$
512	9	$3.3 \cdot 10^{10}$	31	$1.0 \cdot 10^4$	33	$2.6 \cdot 10^4$	36	$2.0 \cdot 10^4$
1024	10	$5.4 \cdot 10^{11}$	19	$1.3 \cdot 10^4$	20	$2.7 \cdot 10^4$	21	$2.0 \cdot 10^4$
2048	11	$8.8 \cdot 10^{12}$	11	$1.8 \cdot 10^4$	12	$2.7 \cdot 10^4$	12	$2.1 \cdot 10^4$
4096	12	$1.4 \cdot 10^{14}$	6.7	$2.5 \cdot 10^4$	6.8	$2.8 \cdot 10^4$	7.0	$2.8 \cdot 10^4$

Table 2: Condition numbers and compression rates in case of the Matérn covariance kernel $k_{3/2}$. The numerical compression rates validate the asymptotically linear behaviour. The numerical condition numbers stay bounded for $\mathbf{\Psi}^{(2,8)}$ and $\mathbf{\Psi}^{(2,10)}$, whereas for $\mathbf{\Psi}^{(2,6)}$ a slight increase is observed.

Wrap-Up A

- GRFs $\mathcal{Z}$ indexed by smooth, Riemannian manifold $\mathcal{M}$ represented in pair of biorthogonal MRAs $(\mathbf{\Psi}, \tilde{\mathbf{\Psi}})$ as

$$\mathcal{Z} = \sum_{\lambda \in \mathcal{J}} \langle \mathcal{Z}, \psi_\lambda \rangle \widetilde{\psi}_\lambda \quad \Longleftrightarrow \quad \widetilde{\mathbf{A}} \widetilde{\mathbf{z}} = \mathbf{w}.$$

- optimal (diagonal) preconditioning of (finite $p \times p$) sections of $\mathbf{C}$ and $\mathbf{P}$,
- optimally $(O(p))$ sparse tapering of $\mathbf{C}$ and $\mathbf{P}$ for any $\mathcal{A} \in OPS^{\hat{r}}_{1,0}(\mathcal{M})$:

$$\#\mathrm{nnz}(\mathbf{C}^{\varepsilon}_p) = O(p), \quad \#\mathrm{nnz}(\mathbf{P}^{\varepsilon}_p) = O(p).$$

- Naturally allows *multilevel path-simulation* of GRF $\mathcal{Z}$,
- No group invariances (stationarity / isotropy etc.) required.

Multilevel Path Simulation of GRF $\mathcal{Z}$

$$\mathcal{Z} = \sum_{\lambda \in \mathcal{J}} \langle \mathcal{Z}, \psi_\lambda \rangle \widetilde{\psi}_\lambda = \sum_{j \geq j_0} \sum_{|\lambda|=j} \langle \mathcal{Z}, \psi_\lambda \rangle \widetilde{\psi}_\lambda \quad \Longleftrightarrow \quad \widetilde{\mathbf{A}}\widetilde{\mathbf{z}} = \mathbf{w}$$

where

$$\widetilde{\mathbf{A}} = \mathcal{A}(\widetilde{\mathbf{\Psi}})(\widetilde{\mathbf{\Psi}}), \quad \widetilde{z}_\lambda = \langle \mathcal{Z}, \psi_\lambda \rangle, \quad w_\lambda = \langle \mathcal{W}, \widetilde{\psi}_\lambda \rangle, \quad \mathbf{w} \sim \mathsf{N}(\mathbf{0}, \widetilde{\mathbf{M}}), \;\; \widetilde{\mathbf{M}} = \mathrm{Id}(\widetilde{\mathbf{\Psi}})(\widetilde{\mathbf{\Psi}}).$$

$\boldsymbol{\xi}$ seqn. of i.i.d $\mathsf{N}(0,1)$ RVs. Then

$$\mathbf{w} \overset{d}{=} \sqrt{\widetilde{\mathbf{M}}}\,\boldsymbol{\xi} \quad \text{and} \quad \widetilde{\mathbf{z}} \overset{d}{=} \widetilde{\mathbf{A}}^{-1}\sqrt{\widetilde{\mathbf{M}}}\,\boldsymbol{\xi}, \quad \widetilde{\mathbf{z}} \sim \mathsf{N}(\mathbf{0}, \mathbf{C}), \quad \mathbf{C} = \widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{M}}\widetilde{\mathbf{A}}^{-1}.$$

Problem: often $\mathbf{C} = \mathcal{C}(\mathbf{\Psi})(\mathbf{\Psi})$ available (e.g. estimation from samples of GP, or from explicit covariance kernel $k\colon \mathcal{M} \times \mathcal{M} \to \mathbb{R}$, $k(x,x') := \mathbb{E}[\mathcal{Z}(x)\mathcal{Z}(x')]$), but not $\mathbf{A}_p$, $\tilde{\mathbf{A}}_p$!

Multilevel Path Simulation of GRF $\mathcal{Z}$

How to numerically sample $\mathcal{Z}$ at cost $O(p)$ / realization?

$$\mathbf{R}_p := \mathbf{D}_p^{\hat{r}}\mathbf{C}_p\mathbf{D}_p^{\hat{r}} \in \mathbb{R}^{p\times p}, \qquad \mathbf{R}_p^{\varepsilon} := \mathbf{D}_p^{\hat{r}}\mathbf{C}_p^{\varepsilon}\mathbf{D}_p^{\hat{r}} \in \mathbb{R}^{p\times p},$$

$$\widetilde{\mathbf{z}}_p^{\varepsilon} := \mathbf{D}_p^{-\hat{r}}\sqrt{\mathbf{D}_p^{\hat{r}}\mathbf{C}_p^{\varepsilon}\mathbf{D}_p^{\hat{r}}}\,\boldsymbol{\xi}_p = \mathbf{D}_p^{-\hat{r}}\sqrt{\mathbf{R}_p^{\varepsilon}}\,\boldsymbol{\xi}_p, \qquad \widetilde{\mathbf{z}}_p^{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{C}_p^{\varepsilon}).$$

$\sigma(\mathbf{R}_p), \sigma(\mathbf{R}_p^{\varepsilon}) \subset [\widetilde{c}_-, \widetilde{c}_+]$ for $\varepsilon \in (0, \varepsilon_0)$ sufficiently small.

[Hale, Higham, and Trefethen]: Computing $\mathbf{A}^{\alpha}$, $\log(\mathbf{A})$, and related matrix functions by contour integrals, SINUM **46** (2008) $\Longrightarrow$

$$\sqrt{\mathbf{R}_p^{\varepsilon}} \approx \mathbf{S}_K := \frac{2E\sqrt{\widetilde{c}_-}}{\pi K}\mathbf{R}_p^{\varepsilon}\sum_{k=1}^{K}\frac{\operatorname{dn}\left(t_k|1-\widehat{\varkappa}_{\mathrm{R}}^{-1}\right)}{\operatorname{cn}^2\left(t_k|1-\widehat{\varkappa}_{\mathrm{R}}^{-1}\right)}\left(\mathbf{R}_p^{\varepsilon}+w_k^2\mathbf{I}_p\right)^{-1}.$$

$\operatorname{sn}, \operatorname{cn}$ and dn: Jacobian elliptic functions, E complete elliptic integral of 2nd kind, parameter $\widehat{\varkappa}_{\mathrm{R}}^{-1}$, $\widehat{\varkappa}_{\mathrm{R}} := \widetilde{c}_+/\widetilde{c}_-$,

$$w_k := \sqrt{\widetilde{c}_-}\,\frac{\operatorname{sn}\left(t_k|1-\widehat{\varkappa}_{\mathrm{R}}^{-1}\right)}{\operatorname{cn}\left(t_k|1-\widehat{\varkappa}_{\mathrm{R}}^{-1}\right)} \quad \text{and} \quad t_k := \frac{\left(k-\frac{1}{2}\right)E}{K}, \quad k \in \{1,\dots,K\}.$$

Computable Approximation in work $O(p)$/ realization (at any level of spatial resolution!):

$$\widetilde{\mathbf{z}}_{p,K}^{\varepsilon} := \mathbf{D}_p^{-\hat{r}}\mathbf{S}_K\boldsymbol{\xi}_p, \qquad \widetilde{\mathbf{z}}_{p,K}^{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{D}_p^{-\hat{r}}\mathbf{S}_K^2\mathbf{D}_p^{-\hat{r}}).$$

Example [Matérn-like GRF on $\mathbb{S}^2$]

Sphere						
p	J	nnz$(\mathbf{C}_J)$	cpu$(\mathbf{C}_J)$	nnz$(\mathbf{L}_J)$	cpu$(\mathbf{L}_J)$	cpu(sample)
6144	5	4.70	18	10.3	0.65	0.0017
24576	6	1.22	113	4.43	5.1	0.015
98304	7	0.43	692	1.68	26	0.096
393216	8	0.12	4108	0.59	151	0.46
1572864	9	0.03	23374	0.20	865	2.7

Table 3: Compression rates and computing times in case of the Matérn covariance kernel $k_{1/2}$ on the sphere. Once Cholesky decomposition has been computed, each sample generated in $O(N)$.

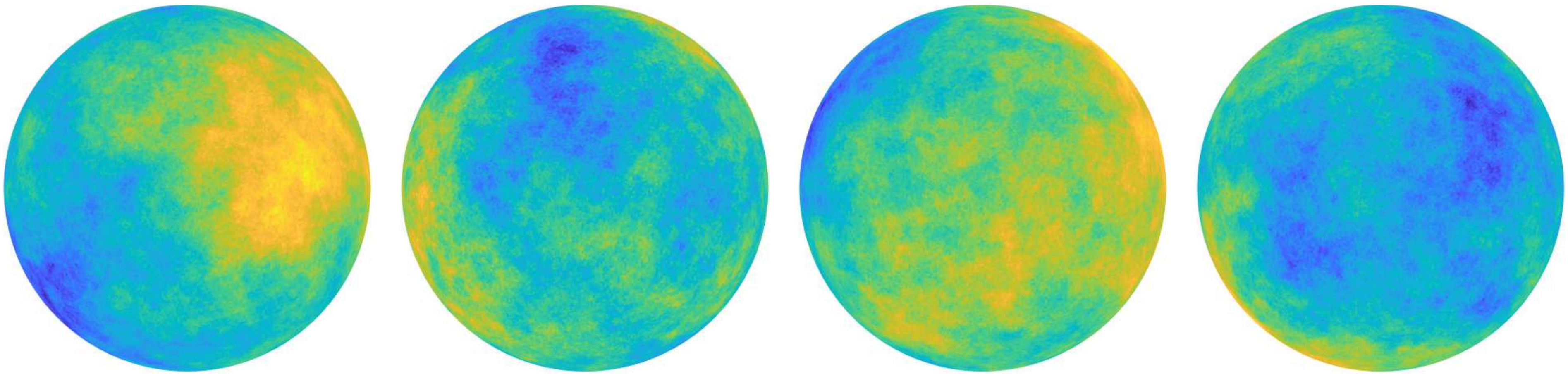

Figure 3: Four realizations of a Gaussian random field on $\mathbb{S}^2$ for the Matèrn covariance $k_{1/2}$ with respect to the geodesic distance.

Multi-Level Monte-Carlo Covariance Estimation

For $J \geq j_0$, define the MLMC estimator by

$$\mathbf{C}_p^\varepsilon \approx E_J^*(\mathbf{C}_p^\varepsilon) := \sum_{j,j'=j_0}^{J} E_{M_{j,j'}}(\mathbf{C}_{\text{global}}^\varepsilon(j,j')).$$

Monte Carlo estimator $E_{M_{j,j'}}(\mathbf{C}_{\text{global}}^\varepsilon(j,j'))$ realized by $M_{j,j'}$ i.i.d. samples of coefficient vector $\widetilde{\mathbf{z}}$ at levels j, j',

$$\mathbf{C}_p^\varepsilon(j,j') \approx E_{M_{j,j'}}(\mathbf{C}^\varepsilon(j,j')) := \frac{1}{M_{j,j'}} \sum_{i=1}^{M_{j,j'}} \left(\widetilde{\mathbf{z}}_i(j)\widetilde{\mathbf{z}}_i(j')^\top\right)^\varepsilon$$

Sample Numbers:

$$M_{j,j'} := \widetilde{M}_{\max\{j,j'\}}, \quad \text{where} \quad \widetilde{M}_j := \sum_{j'=j}^{J} M_{j'}.$$

Proposition:

$$\left\| \sup_{u \in H^t(\mathcal{M})\setminus\{0\}} \sup_{v \in H^{t'}(\mathcal{M})\setminus\{0\}} \frac{|\langle(\mathcal{C}_p^\varepsilon - E_J^*(\mathcal{C}_p^\varepsilon))Q_J u, Q_J v\rangle|}{\|u\|_{H^t(\mathcal{M})}\|v\|_{H^{t'}(\mathcal{M})}} \right\|_{L^2(\Omega)}$$
$$\leq \frac{2C}{1 - 2^{-(\min\{t,t'\}+\beta)}} \sum_{j=j_0}^{J} \frac{1}{\sqrt{\widetilde{M}_j}} 2^{-j(\min\{t,t'\}+\beta)} \|\mathcal{Z}\|^2_{L^4(\Omega;H^\beta(\mathcal{M}))}.$$

Multi-Level Monte-Carlo Covariance Estimation

Theorem [MLMC Covariance Estimation]
Let **Assumption** hold, and assume $\alpha_0 \in [\alpha, 2\hat{r} + t + t']$ for $\alpha < \hat{r} - n/2 + \min\{t, t'\}$.
Choose sample numbers

$$\widetilde{M}_j := \left\lceil \widetilde{M}_{j_0} 2^{-j(n+\alpha)2/3} \right\rceil, \quad j = j_0 + 1, \ldots, J, \quad \widetilde{M}_{j_0} := \begin{cases} 2^{J2\alpha_0}, & \text{if } 2\alpha > n, \\ 2^{J2\alpha_0} J^2, & \text{if } 2\alpha = n, \\ 2^{J(2\alpha_0 + 2n/3 - 4\alpha/3)}, & \text{if } 2\alpha < n. \end{cases}$$

Then

$$\left\| \sup_{u \in H^t(\mathcal{M}) \setminus \{0\}} \sup_{v \in H^{t'}(\mathcal{M}) \setminus \{0\}} \frac{|\langle (\mathcal{C}_p^\varepsilon - E_J^*(\mathcal{C}_p^\varepsilon)) Q_J u, Q_J v \rangle|}{\|u\|_{H^t(\mathcal{M})} \|v\|_{H^{t'}(\mathcal{M})}} \right\|_{L^2(\Omega)} = \mathcal{O}(\varepsilon)$$

can be achieved with computational cost

$$\text{work} = \begin{cases} \mathcal{O}(\varepsilon^{-2}) & \text{if } 2\alpha > n, \\ \mathcal{O}(\varepsilon^{-2} |\log(\varepsilon^{-1})|) & \text{if } 2\alpha = n, \\ \mathcal{O}(\varepsilon^{-(n/\alpha_0 - 2(1 - \alpha/\alpha_0))}) & \text{if } 2\alpha < n. \end{cases}$$

p	J	$\widetilde{M}_j$	ℓ^2-error	
8	3	51200	$1.1 \cdot 10^{-1}$	—
16	4	25600	$5.4 \cdot 10^{-2}$	(2.1)
32	5	12800	$4.9 \cdot 10^{-2}$	(1.1)
64	6	6400	$2.9 \cdot 10^{-2}$	(1.7)
128	7	3200	$1.9 \cdot 10^{-2}$	(1.5)
256	8	1600	$1.3 \cdot 10^{-2}$	(1.4)
512	9	800	$1.1 \cdot 10^{-2}$	(1.3)
1024	10	400	$8.5 \cdot 10^{-3}$	(1.2)
2048	11	200	$5.3 \cdot 10^{-3}$	(1.6)
4096	12	100	$2.9 \cdot 10^{-3}$	(1.3)

Table 4: MLMC Covariance estimation.
Sample sizes $\widetilde{M}_j$, accuracy of MLMC covariance estimation,
$\widetilde{M}_J = 100$, $\widetilde{M}_j = \widetilde{M}_J 2^{J-j}$ shown here for $J = 12$.
Error in operator norm w.r.to (densely populated) truth covariance matrix $\mathbf{C}_p$ in wavelet coordinates.

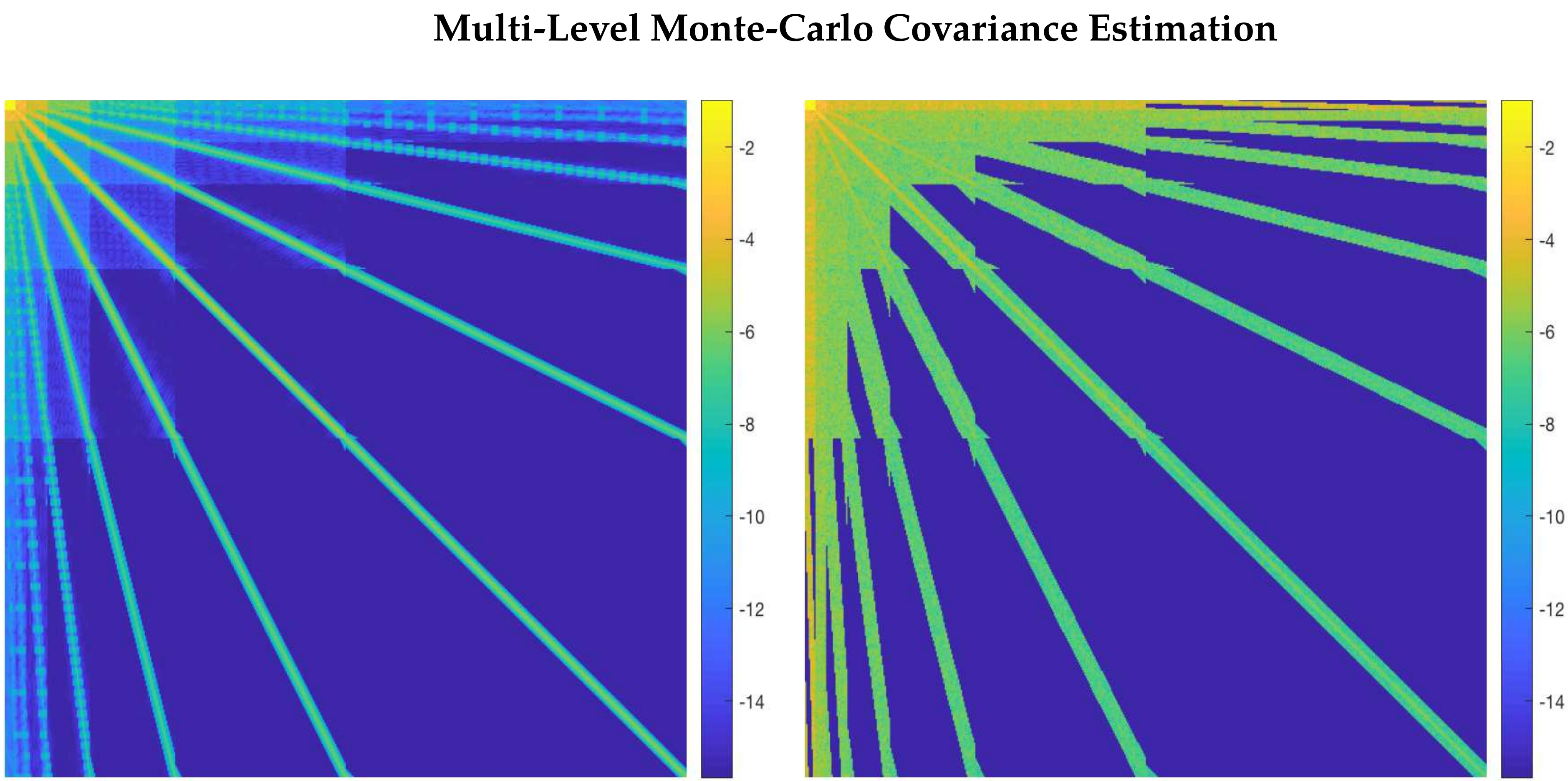

Figure 4: Truth covariance matrix (*left*) in wavelet representation and its multilevel Monte Carlo estimation (*right*) for $p = 512$ parameters. Spatial dimension $n = 1$, Matèrn-covariance kernel $k_{1/2}$, spatial correlation length $\ell = 1$, wavelet $\mathbf{\Psi}^{(2,6)}$.

Spatial Prediction: Sparse Approximate Kriging

Assume:

- $\mathcal{Z}$ observed at K distinct, known spatial locations $\{x_i\}_{i=1}^K \subset \mathcal{M}$
- observations subject to i.i.d. centered Gaussian measurement noise:

$$y_i = \mathcal{Z}(x_i) + \eta_i, \quad i = 1, \dots, K, \qquad \eta_i \sim \mathsf{N}(0, \sigma^2) \quad \text{i.i.d.}$$

Goal:

- predict $\mathcal{Z}$ at an unobserved location $x_* \in \mathcal{M}$ (or several locations), conditioned on observations $\{y_i\}_{i=1}^K$.
- i.e. calculate posterior mean $\mathbb{E}[\mathcal{Z}(x_*)|y_1, \dots, y_K]$.

Issues:

- computationally challenging: assuming finite spatial resolution of $O(2^{-J})$ with parameter dimension $p \sim 2^{Jn}$ for approximating $\mathcal{Z}$, direct solve: cubic either in K or in p or in both.

Our Contribution:

- compressed *approximate* numerical kriging estimator consistent with approximation error of $\mathcal{Z}$, with work $\mathcal{O}((K \log(p) + p)\sigma^{-1} \log(\delta^{-1}\sigma^{-2}))$, where

$$N = \#(\text{pccg-steps}) \gtrsim \sigma^{-1} \log(\delta^{-1}\sigma^{-2}).$$

Spatial Prediction: Sparse Approximate Kriging

Multilevel Compression for Kriging:

- abstract setting: linear functionals $g_1, \ldots, g_K$, and model

$$\mathbf{y} = \mathbf{G}\widetilde{\mathbf{z}} + \boldsymbol{\eta},$$

  where $\mathbf{y} = (y_1, \ldots, y_K)^\top$ random vector corresponding to noisy observations,

- $\mathbf{G} \in \mathbb{R}^{K \times p}$ **observation matrix**, entries $G_{i(j,k)} := \langle g_i, \widetilde{\psi}_{j,k} \rangle$,
- $\widetilde{\mathbf{z}}$, $\boldsymbol{\eta}$ centered multivariate Gaussian random vectors with covariance matrices

$$\mathbf{C}_p \in \mathbb{R}^{p \times p}, \quad \sigma^2 \mathbf{I} \in \mathbb{R}^{K \times K}, \quad \text{respectively.}$$

- Assume: g_i **local averages** at $x_i \in \mathcal{M}$.
- Joint distribution of $\widetilde{\mathbf{z}}$ and $\mathbf{y}$ given by

$$\begin{pmatrix} \widetilde{\mathbf{z}} \\ \mathbf{y} \end{pmatrix} \sim \mathsf{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_p & \mathbf{C}_p \mathbf{G}^\top \\ \mathbf{G}\mathbf{C}_p & \mathbf{G}\mathbf{C}_p\mathbf{G}^\top + \sigma^2 \mathbf{I} \end{pmatrix} \right).$$

- **Law of posterior** $\widetilde{\mathbf{z}}|\mathbf{y}$ again Gaussian,
- **Kriging Predictor** given by posterior mean:

$$\boldsymbol{\mu}_{\widetilde{\mathbf{z}}|\mathbf{y}} = \mathbf{C}_p \mathbf{G}^\top \left(\mathbf{G}\mathbf{C}_p\mathbf{G}^\top + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{y} \;, \quad \boldsymbol{\mu}^{\varepsilon}_{\widetilde{\mathbf{z}}|\mathbf{y}} := \mathbf{C}^{\varepsilon}_p \mathbf{G}^\top \left(\mathbf{G}\mathbf{C}^{\varepsilon}_p\mathbf{G}^\top + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{y}$$

Spatial Prediction: Sparse Approximate Kriging

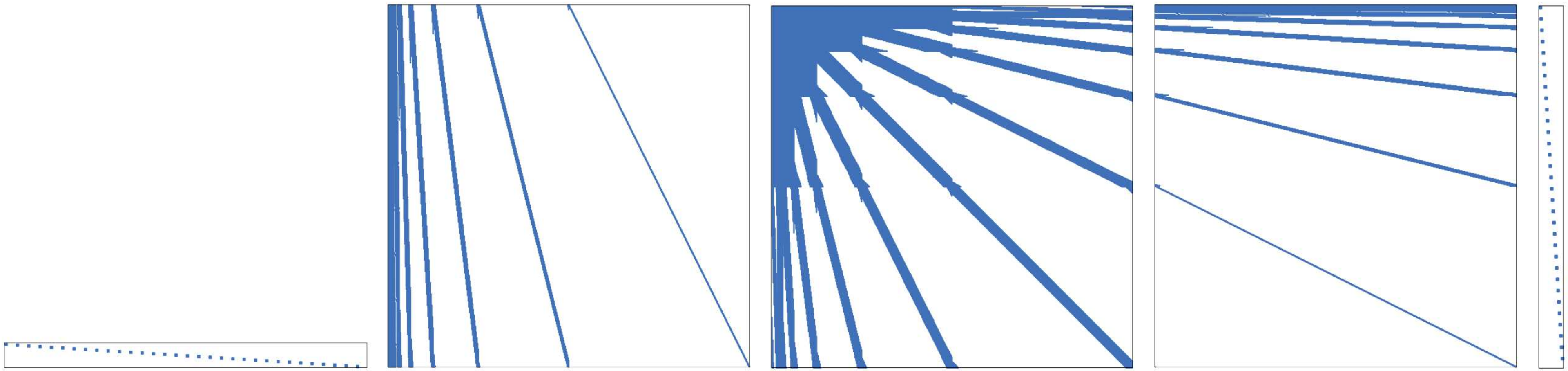

Figure 5: Sparse factorization of the approximate kriging matrix $\mathbf{G}\mathbf{C}_{\mathbf{p}}^{\varepsilon}\mathbf{G}^{\top} = \mathbf{G}_{\tilde{\varphi}}\mathbf{T}_{\tilde{\varphi}\to\tilde{\psi}}^{\top}\mathbf{C}_{p}^{\varepsilon}\mathbf{T}_{\tilde{\varphi}\to\tilde{\psi}}\mathbf{G}_{\tilde{\varphi}}^{\top}$

Theorem [Kriging Predictor with N steps of pccg]
The computational cost of N steps of pccg with ε wavelet compression and with diagonal preconditioning for $\boldsymbol{\mu}_{\tilde{\mathbf{z}}|\mathbf{y}}^{\varepsilon,N}$ to achieve a consistency error $\delta \in (0,1)$

$$\left\|\boldsymbol{\mu}_{\tilde{\mathbf{z}}|\mathbf{y}}^{\varepsilon} - \boldsymbol{\mu}_{\tilde{\mathbf{z}}|\mathbf{y}}^{\varepsilon,N}\right\|_2 = \mathcal{O}(\delta)$$

is $\mathcal{O}((K\log(p)+p)\sigma^{-1}\log(\delta^{-1}\sigma^{-2}))$, where

$$N \gtrsim \sigma^{-1}\log(\delta^{-1}\sigma^{-2}).$$

Wrap-Up B: Conclusions

- Fast Numerical Methods for GRFs $\mathcal{Z}$ indexed by compact $\mathcal{M}$,
- Multi-level numerical Sampling, Covariance Estimation, Kriging at linear in p cost,
- at consistency afforded by path-regularity and suitable MRA.
- No stationarity (or other group invariances) of $\mathcal{Z}$ on $\mathcal{M}$ required.
- Multi-Level Methods for UQ in PDEs with GRF input.

Open Problems

- $\partial\mathcal{M} \neq \emptyset$: Boundary Singularities of GRF and of covariance (w. Melenk, Faustmann)

References

- **SPDE Approach to Coloring of GRFs**

1. **F. Lindgren, H. v. Rue, and J. Lindstrom**.
   An explicit link between Gaussian Fields and Gaussian Markov Random Fields: the SPDE approach.
   J. R. Stat. Soc. Ser. B Stat. Methodol., **73** (4):423-498, 2011.
2. **D. Bolin and F. Lindgren**.
   Spatial models generated by nested stochastic partial differential equations,
   with an application to global ozone mapping.
   Ann. Appl. Stat. **5**(1):523-550, 2011.

References

- **Wavelet Compression of Nonlocal Operators**

  1. **Yves Meyer**
     Ondelettes et opérateurs. II. Opérateurs de Calderón-Zygmund. Hermann, Paris, 1990.
  2. **Reinhold Schneider**
     Multiscale and wavelet matrix compression:
     Analysis-based methods for the efficient solution of large nonsparse systems of equations.
     Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 1998.

- **Wavelet Constructions in Domains and on Surfaces**

  1. **W. Dahmen and R. Schneider**.
     Wavelets on manifolds. I. Construction and domain decomposition.
     SIAM J. Math. Anal. **31**(1):184-230, 1999.
  2. **N. Rekatsinas and R. Stevenson**.
     A quadratic Finite Element Wavelet Riesz basis.
     Int. J. Wavelets Multiresolut. Inf. Process., **16**(4):1850033, 17, 2018.

References

- **Covariance Matrix Tapering**

1. **P. J. Bickel and E. Levina**.
   Regularized estimation of large covariance matrices.
   Ann. Statist. **36**(1):199-227, 2008.

Thank You.