Covariance Tapering
for Interpolation
of Large Spatial Datasets

October 7, OSU

Tapering and Kriging

Introduce a sparseness structure in the covariance via tapering to gain computational advantages in large kriging problems constraint to maintaining asymptotic optimality.

In collaboration with Marc Genton and Doug Nychka.

Motivation

Monthly aggregated precipitation in April 1948

Precipitation anomaly in April 1948

Best Linear Unbiased Predictor

Suppose a spatial process of the form

\[ Z(x) = m(x)^T \beta + \epsilon(x), \quad E(\epsilon) = 0, \quad \text{Cov}(\epsilon) = C \]

with observations \( Z = (Z(x_1), \ldots, Z(x_n))^T, x_i \in D \subseteq \mathbb{R}^d \).

The kriging estimator (BLUP) is

\[ \hat{Z}(x_0) = \lambda^T Z \]

where

\[ \lambda = C^{-1}(I - M(M^T C^{-1} M)^{-1}(M^T C^{-1} \epsilon - m(x_0))) \]

with \( c_i = \text{Cov}(Z(x_0), Z(x_i)), M = (m(x_1), \ldots, m(x_n))^T. \)
Prediction at a large number of points involves either

- solving one large linear system,
- solving many tiny linear systems

or, being smarter and

- “approximating” the covariance and solving “efficiently” one large linear system.

Easy to think in one dimension: precipitation anomaly along 40° lat.

For an isotropic and stationary process with covariance $C_0(h)$, find a taper $C_0(h)$, such that kriging with the product $C_0(h)C_0(h)$ is asymptotically optimal.

$$\frac{\text{MSE}(\mathbf{x}^*, C_0 C_0)}{\text{MSE}(\mathbf{x}^*, C_0)} \to 1 \quad \frac{\varphi(\mathbf{x}^*, C_0 C_0)}{\text{MSE}(\mathbf{x}^*, C_0)} \to \gamma$$

$$\varphi(\mathbf{x}^*, C) = C(0) - c^T c^{-1} c^*$$
Matérn Covariance

We need a broad, flexible class of covariances to describe spatial processes.

Matérn class covariance

\[ C_{\alpha,\nu}(h) \propto (\alpha h)^{\nu} K_\nu(\alpha h) \quad h = |h| \]

and spectral density

\[ f_{\alpha,\nu}(\rho) \propto \frac{1}{(\alpha^2 + \rho^2)^{\nu + d/2}} \quad \rho = |\omega| \]

Differentiability at the origin of the covariance is related to the tail behavior of the spectrum, i.e., the smoothness of the process.

The process is \( m \) times mean squared differentiable iff \( m < \nu \).

Taper Functions

We impose on the taper \( C_\theta \) the conditions

- \( C_\theta \) is a positive definite function in \( \mathbb{R}^d \)
- \( C_\theta(h) = 0 \) for \( h > \theta \)

For example:

- triangular: \( C_\theta(h) = \max(0, \frac{|h|}{\theta}) \)
- spherical: \( C_\theta(h) = \max(0, \frac{|h|}{\theta})^2 \left( 1 + \frac{|h|}{\theta} \right) \)
- Wendland-type: \( C_\theta(h) = \max(0, \frac{|h|}{\theta})^{\ell+k} \)-polynom in \( \frac{|h|}{\theta} \) of deg \( k \)

Misspecified Covariances

In a series of (Annals) papers, Stein gives asymptotic results for misspecified covariances.

Suppose the true covariance \( C_0 \) and spectrum \( f_0 \). If we krig with the misspecified covariance \( C_1 \) characterized by \( f_1 \) then under the Tail Condition

\[ \frac{f_1(\omega)}{f_0(\omega)} = \gamma \quad 0 < \gamma < \infty \quad \text{as} \quad |\omega| \to \infty \quad \text{and} \quad \ldots \]

we have asymptotic optimality

\[ \lim_{n \to \infty} \frac{\text{MSE}(\mathbf{x}^*, C_1)}{\text{MSE}(\mathbf{x}^*, C_0)} = 1 \quad \lim_{n \to \infty} \frac{g(\mathbf{x}^*, C_1)}{g(\mathbf{x}^*, C_0)} \to \gamma \]

Tapered Covariances

Tapering is a form of misspecification if

\[ \frac{\mathcal{F}(C_{\alpha,\nu}(h)C_\theta(h))}{\mathcal{F}(C_{\alpha,\nu}(h))} \to \gamma \quad \text{as} \quad |\omega| \to \infty \]

Which taper satisfies this condition?

The taper has to be

- as differentiable at the origin as the original covariance
- more differentiable throughout the domain than at the origin

Taper Theorem

Infill Condition: Let \( \mathbf{x}^* \in D \) and \( \mathbf{x}_1, \mathbf{x}_2, \ldots \) be a dense sequence in \( D \).

Taper Condition: Let \( f_\theta \) be the spectral density of the taper covariance, \( C_\theta \), and for some \( \epsilon > 0 \) and \( M < \infty \)

\[ f_\theta(\rho) < \frac{M}{(1 + \rho^2)^{\nu + d/2 + \epsilon}} \]

Taper Theorem: Assume that \( C_{\alpha,\nu} \) is a Matérn covariance with smoothness parameter \( \nu \) and the Infill and Taper Conditions hold. Then

\[ \lim_{n \to \infty} \frac{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu}C_\theta)}{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu})} = 1 \quad \lim_{n \to \infty} \frac{g(\mathbf{x}^*, C_{\alpha,\nu}C_\theta)}{g(\mathbf{x}^*, C_{\alpha,\nu})} = 1(= \gamma) \]
Conditions in Terms of Covariances

The principal irregular term (PIT) relates the tail behavior of the spectrum and the behavior at the origin of the covariance.

Formally, the PIT of $C$ is the first term as a function of $h$ in this series expansion about zero that is not raised to an even power.

**Conjecture:** Assume a polynomial isotropic covariance function $C_0$ in $\mathbb{R}^d$ that is integrable with PIT $B h^p$ and ... Then the PIT and the tail behaviour are related by

$$\lim_{\rho \to \infty} \rho^{p+d} f_0(\rho) = |B \frac{\mu^1}{2 \pi} (d+1/2)|$$

Simulation Study

When does infill asymptotics kick in?

Simulation setup:
- $n$ equispaced or random (100 samples) observations in $[0,1]^2$
- $C_{\alpha,\nu}$: Matérn covariance, eff. range 0.4
- $C_0$: spherical and Wendland-type, range $\theta = 0.4$
- prediction $\theta = 0.5$.

Simulation Study

What is an efficient taper?

Simulation setup:
- $n = 400$ equispaced or random observations in $[0,1]^2$
- $C_{\alpha,\nu}$: Matérn covariance, eff. range 0.4
- $C_0$: spherical, Wendland-type, range $\theta$
- NN-kriging with neighborhood $\theta$
- prediction on $(0.5,0.5)$.

Recap

To summarize we have discussed

- the BLUP, requiring solutions of linear systems
  - one point $c^Tc^{-1}Z$, field $C^Tc^{-1}Z$

- optimal covariances, decay exponentially
  - no zeros in $C$, $c$ or $C$

- that tapering preserves optimality,
  - covariance matrices are sparse

Is there a computational gain?
To solve the system $Cx = Z$, we

- perform a Cholesky factorisation $C = LL^T$
- solve two triangular systems $Lu = Z$ and $L^Tx = u$

Use a special format to store sparse matrices:

$$8z + 4z + 4n + 1 \text{ bytes} \text{ vs } 8 \times n^2 \text{ bytes}$$

and $z \approx n \times \#\text{tapered neighbors}$.

The Cholesky factor of a sparse matrix “is” sparse.

Matlab and R contain a toolbox and libraries (SparseM and KriSp).

### Computational Efficiency

**Linux powered 2.6 GHz Xeon processor with 2 Gbytes RAM.**
R with KriSp, SparseM, Fields and Base libraries.

<table>
<thead>
<tr>
<th>Action</th>
<th>Classic Time (sec)</th>
<th>Classic optim.</th>
<th>Sparse</th>
<th>Sparse + FFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading data, variable setup</td>
<td>0.54 0.54 0.54 0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating the matrix $C$</td>
<td>41.34 21.59 6.35 6.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solving $Cx = Z$ { Cholesky Backsolve }</td>
<td>169.09 169.09 0.28 0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplying $C^T$ with $C^{-1}Z$</td>
<td>4638.01 1830.86 733.82 26.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating the figure</td>
<td>6.19 6.19 6.19 6.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.4h 34min 12min 41sec</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

**Completed:**
- Efficient methodological approach
- Spatial models with drift and/or nugget effect
- Rule of thumb: 16–20 observation in taper range

**In progress:**
- Conjecture
- Apply to other statistical problems, like microarrays

**In prospection:**
- Nonstationarity, MLE, ...