Sparse Matrix Techniques for Interpolation of Large Environmental and Biological Datasets

February 8, UCDHCS

Motivation

Precipitation anomaly in April 1948

Best Linear Unbiased Predictor

Suppose a spatial process of the form
\[ Z(x) = m(x)^T \beta + \varepsilon(x), \quad E(\varepsilon) = 0, \quad \text{Cov}(\varepsilon) = C \]
with observations \( Z = (Z(x_1), \ldots, Z(x_n))^T, x_i \in D \subseteq \mathbb{R}^d \).

The kriging estimator (BLUP) is
\[ \hat{Z}(x_0) = \lambda^T Z = c^T C^{-1} Z \]

where
\[ \lambda = C^{-1}(I - M(M^T C^{-1} M)^{-1}(M^T C^{-1} C - m(x_0))) = C^{-1} c \]
with \( c_i = \text{Cov}(Z(x_0), Z(x_i)) \), and \( M = (m(x_1), \ldots, m(x_n))^T \).

Motivation

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.

Tapering and Kriging

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

In collaboration with Marc Genton (T&M), Doug Nychka (NCAR) and Steve Sain (CUD)

Motivation

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

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

Ordinary kriging

Nearest neighbor kriging with 8 observations

Objective

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

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

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

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: \( C_\theta(h) = (1 - \frac{|h|}{\theta}_+) \) where \( (x)_+ = \max(0, x) \)
- spherical: \( C_\theta(h) = (1 - \frac{|h|}{\theta}_+)^2 (1 + \frac{|h|}{\theta}_+) \)
- Wendland-type: \( C_\theta(h) = (1 - \frac{|h|}{\theta}_+^{\nu+k}) \) - a polynomial in \( \frac{|h|}{\theta}_+ \) of degree \( k \)

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

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

we have asymptotic optimality

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

Taper Theorem

Infill Condition: Let \( x^* \in \mathcal{D} \) and \( x_1, x_2, \ldots \) be a dense sequence in \( \mathcal{D} \).

Taper Condition: Let \( f_0 \) be the spectral density of the taper covariance, \( C_\theta \), and for some \( \epsilon > 0 \) and \( M < \infty \)

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

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

\[
\lim_{n \to \infty} \frac{\text{MSE}(x^*, C_{\theta, \nu} C_\theta)}{\text{MSE}(x^*, C_\theta)} = 1 \quad \lim_{n \to \infty} \frac{g(x^*, C_{\theta, \nu} C_\theta)}{\text{MSE}(x^*, C_\theta)} = 1(= \gamma)
\]

Examples

- Exponential
- Spheric
- Exp * Spheric
- Matern (1.5)
- Wendland
- Matern * Wendland

Tapered Covariances

Tapering is a form of misspecification if

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

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

Simulation Study

When does infill asymptotics kick in?

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

When does infill asymptotics kick in?

![Graph showing MSE taper vs n (log scale) for different values of ν and θ.]

- Spherical
- Wendland type

<table>
<thead>
<tr>
<th>ν</th>
<th>MSE taper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.00</td>
</tr>
<tr>
<td>1.0</td>
<td>1.02</td>
</tr>
<tr>
<td>1.5</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Simulation Study

What is an efficient taper?

![Graph showing MSE taper vs θ for different values of ν.]

- Spherical
- Wendland type

<table>
<thead>
<tr>
<th>ν</th>
<th>MSE taper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.00</td>
</tr>
<tr>
<td>1.0</td>
<td>1.02</td>
</tr>
<tr>
<td>1.5</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Recap

To summarize we have discussed:

- the BLUP, requiring solutions of linear systems
- optimal covariances, decay exponentially
- that tapering preserves optimality,

Is there a computational gain?

Computational Efficiency

To solve the system \( Cx = Z \), we:

- perform a Cholesky factorisation \( C = LL^T \)
- solve two triangular systems \( Lu = Z \) and \( L^T x = u \)

Use a special format to store sparse matrices:

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

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</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</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>Creating the matrix C</td>
<td>41.34</td>
<td>21.59</td>
<td>6.35</td>
<td>6.35</td>
</tr>
<tr>
<td>Solving $Cx = Z$</td>
<td>169.09</td>
<td>169.09</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Backsolve</td>
<td>6.13</td>
<td>6.13</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Multiplying $C^T$ with $C^{-1}Z$</td>
<td>4638.01</td>
<td>1830.86</td>
<td>733.82</td>
<td>26.99</td>
</tr>
<tr>
<td>Creating the figure</td>
<td>6.19</td>
<td>6.19</td>
<td>6.19</td>
<td>6.19</td>
</tr>
<tr>
<td>Total</td>
<td>1.4h 34min 12min 41sec</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Application: Microarrays

Additive mixed-effects ‘spatial’ process:

$$Y = \text{fixed effects} + \text{remaining random structure} = X\beta + h + \varepsilon$$

where $h$ represents a random, zero-mean spatial process.

$$\varepsilon$$ represents a random error process orthonormal to $h$.

$X$ includes chip specific and gene specific (across chip) terms.

$$X = \begin{bmatrix} 1 & R & C & 0 & \ldots & 0 & G \\ 0 & 1 & R & C & \ldots & 0 & G \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ldots & 0 & 1 & R & C & G \end{bmatrix}$$

where $R$ and $C$ represent row and column effects.

$G$ indicate gene effects.

Application: Microarrays

Y = mean = row effects + column effects + spatial process + gene effects + error

Discussion

Completed:
- Efficient methodological approach
- Spatial models with drift and/or nugget effect
- Rule of thumb: 16-24 observation in taper range

In progress:
- Relate the tail behavior of the spectrum and the behavior at the origin of the covariance
- Apply to other statistical problems, like microarrays

In prospect:
- Nonstationarity, MLE, …