Geostatistical Modeling for Large Data Sets

Whitney Huang

Department of Statistics
Purdue University

October 28, 2014
Outline

Motivation

Methods

- Covariance tapering
- Low–rank approximation
- Likelihood approximation
- Gaussian Markov random field approximation
Gaussian process (GP) geostatistics

Model:

\[ Y(s) = \mu(s) + \eta(s) + \epsilon(s), \quad s \in S \subset \mathbb{R}^d \]

where

\[ \mu(s) = X^T(s)\beta, \quad \{\eta(s)\}_{s \in S} \sim GP(0, C(\cdot, \cdot)) \]

\[ C(s, s') = \sigma^2 \rho_\theta(\|s - s'\|), \text{ and } \epsilon(s) \sim N(0, \tau^2) \quad \forall s \in S \]

Log-likelihood:
Given data \( Y = (Y(s_1), \cdots, Y(s_n))^T \)

\[ l_n(\beta, \theta, \sigma^2, \tau^2) \propto -\frac{1}{2} \log |\Sigma(\theta, \sigma^2) + \tau^2 I_n| \]

\[ -\frac{1}{2}(Y - X^T\beta)^T [\Sigma(\theta, \sigma^2) + \tau^2 I_n]^{-1}(Y - X\beta) \]

where \( \Sigma(\theta, \sigma^2)_{i,j} = \sigma^2 \rho_\theta(\|s_i - s_j\|), i,j = 1, \cdots, n \)
“Big $n$ Problem” in geostatistics

- Modern environmental instrument has produced a wealth of space–time data ⇒ $n$ is big
- Evaluation of the likelihood function involves factorizing large covariance matrices that generally requires
  - $O(n^3)$ operations
  - $O(n^2)$ memory
- Modeling strategies are needed to deal with large spatial data set.
  - parameter estimation ⇒ MLE, Bayesian
  - spatial interpolation ⇒ Kriging
  - multivariate spatial data, spatio-temporal data
“Big $n$ Problem” in geostatistics

- Modern environmental instrument has produced a wealth of space–time data ⇒ $n$ is big
- Evaluation of the likelihood function involves factorizing large covariance matrices that generally requires
  - $\mathcal{O}(n^3)$ operations
  - $\mathcal{O}(n^2)$ memory
- Modeling strategies are needed to deal with large spatial data set.
  - parameter estimation ⇒ MLE, Bayesian
  - spatial interpolation ⇒ Kriging
  - multivariate spatial data, spatio-temporal data
"Big n Problem" in geostatistics

- Modern environmental instrument has produced a wealth of space–time data ⇒ \( n \) is big
- Evaluation of the likelihood function involves factorizing large covariance matrices that generally requires
  - \( O(n^3) \) operations
  - \( O(n^2) \) memory
- Modeling strategies are needed to deal with large spatial data set.
  - parameter estimation ⇒ MLE, Bayesian
  - spatial interpolation ⇒ Kriging
  - multivariate spatial data, spatio-temporal data
Modeling strategies in the literature

- Covariance tapering (Furrer et al. 06, Kaufman et al. 08, Du et al. 09)
- Low–rank approximation (Cressie & Johannesson 08, Banerjee et al. 08)
- Likelihood approximation (Vecchia 88, Stein 04)
- Gaussian Markov random field approximation (Rue & Tjelmeland 02, Rue & Held 05, Lindgren et al. 11)
Modeling strategies in the literature

- Covariance tapering (Furrer et al. 06, Kaufman et al. 08, Du et al. 09)
- Low–rank approximation (Cressie & Johannesson 08, Banerjee et al. 08)
- Likelihood approximation (Vecchia 88, Stein 04)
- Gaussian Markov random field approximation (Rue & Tjelmeland 02, Rue & Held 05, Lindgren et al. 11)
Modeling strategies in the literature

- **Covariance tapering** (Furrer et al. 06, Kaufman et al. 08, Du et al. 09)
- **Low–rank approximation** (Cressie & Johannesson 08, Banerjee et al. 08)
- **Likelihood approximation** (Vecchia 88, Stein 04)
- **Gaussian Markov random field approximation** (Rue & Tjelmeland 02, Rue & Held 05, Lindgren et al. 11)
Modeling strategies in the literature

- **Covariance tapering** (Furrer et al. 06, Kaufman et al. 08, Du et al. 09)
- **Low–rank approximation** (Cressie & Johannesson 08, Banerjee et al. 08)
- **Likelihood approximation** (Vecchia 88, Stein 04)
- **Gaussian Markov random field approximation** (Rue & Tjelmeland 02, Rue & Held 05, Lindgren et al. 11)
Outline

Motivation

Methods
  Covariance tapering
  Low–rank approximation
  Likelihood approximation
  Gaussian Markov random field approximation
Covariance tapering (Furrer et al. 06)

We replace the $C(\|h\|)$ by

$$C_{\text{tap}}(h; \gamma) = \rho_{\text{tap}}(h; \gamma) \circ C(h)$$

where $\rho_{\text{tap}}(h; \gamma)$ is an isotropic correlation function with compact support ($\rho_{\text{tap}}(h) = 0$ if $h \geq \gamma$) and $\circ$ denotes the Schur product.
Covariance tapering cont’d

- $C_{\text{tap}}(h)$ is a valid covariance function
- Sparse matrix algorithm can be used
Low–rank approximation

Hierarchical Representation (assume zero mean spatial process)

\[ Y = \eta + \varepsilon, \quad \varepsilon \sim MVN(0, \Sigma_\varepsilon) \]
\[ \eta = H\alpha + \xi, \quad \xi \sim MVN(0, \Sigma_\xi) \]
\[ \alpha \sim MVN(0, \Sigma_\alpha) \]

where \( \alpha = (\alpha_1, \cdots, \alpha_p)^T \) such that \( p \ll n \) and \( H \) is mapping from the latent process, \( \alpha \), to the true spatial process of interest, \( \eta \). \( \Sigma_\varepsilon \) and \( \Sigma_\xi \) and diagonal.
Low–rank approximation cont’d

To carry out the spatial interpolation (i.e. kriging) of \( \eta(s_0)|\{Y(s_i)\}_{i=1}^n \) one need to compute

\[
\left( H\Sigma_\alpha H^T + V \right)^{-1}
\]

where \( V = \Sigma_\varepsilon + \Sigma_\xi \).

Sherman–Morrison–Woodbury formula

\[
(A + BCD)^{-1} = A^{-1} - A^{-1} B \left( C^{-1} + DA^{-1}B \right)^{-1} DA^{-1}
\]

In the case of low–rank model, we have

\[
\left( H\Sigma_\alpha H^T + V \right)^{-1} = V^{-1} - V^{-1} H \left( \Sigma_\alpha^{-1} + H^T V^{-1} H \right)^{-1} H^T V^{-1}
\]
Fixed Rank Kriging (Cressie & Johannesson 08)

\[ Y = X\beta + ZW^* + \varepsilon \]

Let \( W^* = \{w(s_i^*)\}_{i=1}^p \) be latent variables at \( p \ll n \) known knots \( \{s_i^*\}_{i=1}^p \) and \( Z(\cdot) \) be a known basis function.

The fixed rank kriging is equivalent to the following low rank model

\[ Y(s) = X(s)\beta + \sum_{j=1}^{p} Z(s - s_j^*)W_j + \varepsilon(s) \]
Gaussian Predictive Process (Banerjee et al. 08)

Use a model

\[ Y(s) = X(s)^T \beta + H\alpha(s) + \epsilon(s) \]

to approximate the original spatial process

\[ Y(s) = X(s)^T \beta + \eta(s) + \epsilon(s) \]

Knots: \( \{s_1^*, \cdots, s_p^*\} \) where \( p \ll n \)

\[ \Rightarrow \alpha = \{\alpha(s_i^*)\}_{i=1}^p, \quad H(\theta) = [\text{Cov}(s_i, s_j^*; \theta)]^T [\Sigma_\alpha]^{-1} \]
Likelihood approximation (Vecchia 88)

Partition the observation vector \( \mathbf{Y} \) into sub–vector \( \mathbf{Y}_1, \cdots, \mathbf{Y}_b \) and let \( \mathbf{Y}(j) = (\mathbf{Y}_1^T, \cdots, \mathbf{Y}_j^T)^T \)

The exact likelihood

\[
p(\mathbf{Y}; \beta, \theta) = p(\mathbf{Y}_1; \beta, \theta) \prod_{j=2}^{b} p(\mathbf{Y}_j|\mathbf{Y}_{(j-1)}; \beta, \theta)
\]

Approximate the exact likelihood by replacing \( \mathbf{Y}_{(j-1)} \) by a sub–vector \( \mathbf{S}_{(j-1)} \) of \( \mathbf{Y}_{(j-1)} \)
Markov Random Fields

- Covariance tapering
- Low-rank approximation
- Likelihood approximation
- Gaussian Markov random field approximation
Gaussian Markov Random Fields (GMRF)

Definition

Let the neighbors to a point $i$ be the points $\mathcal{N}_i$ that are "close" to $i$. A Gaussian random field $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma = Q^{-1})$ that satisfies

$$p(X_i|X_j, j \neq i) = p(X_i|X_j : j \in \mathcal{N}_j)$$

is a Gaussian Markov random field (GMRF) with $Q_{ij} = 0$ iff $X_i \perp X_j|X_{-ij}$
Remarks: GP vs. GMRF in geostatistical modeling

- **+:** GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- **−:** Inference involves factorizing covariance matrices
- **+:** GMRF model is computationally efficient due to the sparse precision matrix
- **−:** Only for discretely indexed spatial data

Main idea of GMRF approach:

\[ \text{GP inference} \iff \text{SPDE} \iff \text{GMRF computation} \]
Remarks: GP vs. GMRF in geostatistical modeling

- +: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- -: Inference involves factorizing covariance matrices
- +: GMRF model is computationally efficient due to the sparse precision matrix
- -: Only for discretely indexed spatial data

Main idea of GMRF approach:

\[
\begin{align*}
\text{GP inference} & \quad \leftrightarrow \quad \text{SPDE} \\
\text{GMRF computation} &
\end{align*}
\]
Remarks: GP vs. GMRF in geostatistical modeling

- +: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- -: Inference involves factorizing covariance matrices
- +: GMRF model is computationally efficient due to the sparse precision matrix
- -: Only for discretely indexed spatial data

Main idea of GMRF approach:

\[
\begin{align*}
&\text{GP inference} \quad \overset{\text{SPDE}}{\leftrightarrow} \quad \text{GMRF computation}
\end{align*}
\]
Remarks: GP vs. GMRF in geostatistical modeling

- **+**: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- **-**: Inference involves factorizing covariance matrices
- **+**: GMRF model is computationally efficient due to the sparse precision matrix
- **-**: Only for discretely indexed spatial data

Main idea of GMRF approach:

\[
\text{GP inference} \quad \leftrightarrow \quad \text{SPDE} \quad \rightarrow \quad \text{GMRF computation}
\]
Remarks: GP vs. GMRF in geostatistical modeling

- **+:** GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- **−:** Inference involves factorizing covariance matrices
- **+:** GMRF model is computationally efficient due to the sparse precision matrix
- **−:** Only for discretely indexed spatial data

Main idea of GMRF approach:

\[ \text{GP inference} \quad \leftrightarrow \quad \text{SPDE} \quad \leftrightarrow \quad \text{GMRF computation} \]
Remarks: GP vs. GMRF in geostatistical modeling

- +: GP model is widely used in modeling continuously indexed spatial data in which the covariance function characterizes the process properties
- -: Inference involves factorizing covariance matrices
- +: GMRF model is computationally efficient due to the sparse precision matrix
- -: Only for discretely indexed spatial data

Main idea of GMRF approach:

\[
\begin{align*}
\underbrace{\text{GP}}_{\text{inference}} & \iff \underbrace{\text{SPDE}}_{\text{GMRF}} \\
\text{computation} &
\end{align*}
\]

Gaussian process \( Y(s) \) with Matern covariance function is a stationary solution to the linear fractional stochastic partial differential equation:

\[
(\alpha^2 - \Delta)^{\frac{\kappa}{2}} Y(s) = \mathcal{W}(s), \quad \kappa = \nu + \frac{d}{2}, \nu > 0
\]

where

- \( \mathcal{W}(s) \) is a spatial Gaussian white noise
- \( \Delta = \sum_i \frac{\partial^2}{\partial s_i^2} \) is the Laplacian operator
- \( d \) is the dimension of the spatial domain
An explicit link between GP and GMRF via SPDE 
(Lindgren et al. 11)

- Establish the link between GP with Matérn covariance function (with $\nu + \frac{d}{2}$ are integers) and GMRF
- (Bayesian) inference can be done by using Integrated nested Laplace approximation (INLA) approach
- The extensions to nonstationary models, models on manifolds, multivariate models, spatio-temporal models are relatively easy
An explicit link between GP and GMRF via SPDE (Lindgren et al. 11)

- Establish the link between GP with Matérn covariance function (with $\nu + \frac{d}{2}$ are integers) and GMRF
- (Bayesian) inference can be done by using Integrated nested Laplace approximation (INLA) approach
- The extensions to nonstationary models, models on manifolds, multivariate models, spatio-temporal models are relatively easy
An explicit link between GP and GMRF via SPDE (Lindgren et al. 11)

- Establish the link between GP with Matérn covariance function (with $\nu + \frac{d}{2}$ are integers) and GMRF
- (Bayesian) inference can be done by using Integrated nested Laplace approximation (INLA) approach
- The extensions to nonstationary models, models on manifolds, multivariate models, spatio-temporal models are relatively easy
Extensions

- non-stationary model on a sphere

\[(\alpha^2(s) + \Delta)^{\frac{\kappa}{2}} \tau(s) Y(s) = \mathcal{W}(s), \quad s \in S^2\]

- non-separable anisotropic space-time model

\[\left(\frac{\partial}{\partial t} + (\alpha^2 + \mathbf{m} \cdot \nabla - \nabla \cdot \mathbf{H} \nabla)\right)^{\frac{\kappa}{2}} Y(s, t) = \mathcal{W}(s, t)\]

where \((s, t) \in S^2 \times \mathbb{R}\)
For Further Reading I

H. Rue, and L. Held

*Gaussian Markov Random Fields: Theory and Applications.*

S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang

Gaussian Predictive Process Models for Large Spatial Data Sets


N. A. C. Cressie, and G. Johannesson

Fixed Rank Kriging for Very Large Spatial Data Sets

For Further Reading II

J. Du, H. Zhang, and V. S. Mandrekar
Fixed–Domain Asymptotic Properties of Tapered Maximum Likelihood Estimators

R. Furrer, M. G. Genton, and D. W. Nychka
Covariance Tapering for Interpolation of Large Spatial Datasets

C. G. Kaufman, M. J. Schervish, and D. W. Nychka
Covariance Tapering for Likelihood–Based Estimation in Large Spatial Data Sets
For Further Reading III

Lindgren, F., Rue, H., & Lindström, J.
An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach.
*JRSSB*, 73:423–498

H. Rue, and H. Tjelmeland
Fitting Gaussian Markov Random Fields to Gaussian Field.

M. L. Stein, Z. Chi, and L. J. Welty
Approximating Likelihoods for Large Spatial Data Sets
A. V. Vecchia

Estimation and Model Identification for Continuous Spatial Processes