GRF Approximations using GMRFs

Calder, Pratola, Gory, Liaw, Olsen, Vaidyanathan, Zheng

January 30, 2014

Outline

Outline

Introduction to Gaussian Markov Random Fields

- A GMRF is essentially a Normal random vector with a certain type of covariance structure.
- Specifically, $\mathbf{x} = (x_1, \ldots, x_n)$ is a GMRF with mean $\boldsymbol{\mu}$ and symmetric positive definite precision matrix $\mathbf{Q}$ if it has density

$$\pi(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{Q}|^{1/2}\exp(-1/2(\mathbf{x}-\boldsymbol{\mu})'\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu}))$$

  and the property that

$$Q_{ij} = 0 \iff x_i \perp x_j|\mathbf{x}_{-\{i,j\}} \text{ for all } i \neq j.$$

- GMRFs are often used as a prior where $\mathbf{x}$ is a collection of latent variables.

Advantages of GMRFs

- Sparse precision matrix makes problems more computationally tractable.
- Exact algorithms (up to numerical error) are available and provide an alternative to MCMC.
- Fast because of linear algebra tricks (such as the reordering of band matrices to compute the Cholesky decomposition of the precision matrix).
- Can approximate a Gaussian Random Field with a GMRF to speed up computation.

Outline

Gaussian Field

- A continuous random field $x(\boldsymbol{u})$ defined on a domain $\mathcal{D} \in R^d$ is a Gaussian Field if all finite dimensional distributions of the field are Gaussian.
- It is completely defined by the mean $\boldsymbol{\mu}(.)$ and covariance function $C(.,.)$
- Observe data $X = (X_1, \ldots, X_n)^T$ at spatial locations $\{\boldsymbol{u}_i, i = 1, \ldots, n\}$
- Assume an exponential covariance function

$$l(\boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{X}'\Sigma(\boldsymbol{\theta})^{-1}\mathbf{X})\right\}$$

$$\Sigma_{ij} \equiv Cov(x(\mathbf{u_i}), x(\mathbf{u_j})) \equiv C(\mathbf{u_i}, \mathbf{u_j}) \propto \exp\left\{\frac{-3||\mathbf{u_i} - \mathbf{u_j}||}{\theta}\right\}$$

- Computational cost of factorizing $\boldsymbol{\Sigma}$ for a GF is generally $O(n^3)$

GMRF to approximate GF?

- Computational cost of factorizing $\boldsymbol{Q}$ for a GMRF is typically $O(n^3/2)$
- Why do we model GF and not a GMRF?
  - Deal with continuous processes (not defined on a regular grid)
  - It is easier to specify the covariance structure for two distinct sites than to specify conditional properties

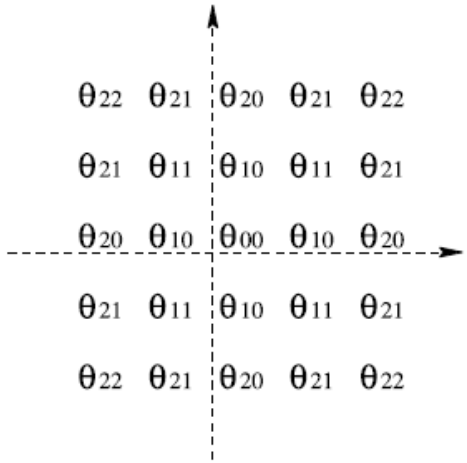
What we would like to do

- Model the process using a GF on a set of locations $\{\boldsymbol{u}_i\}$ to construct a discretized GF with a covariance matrix
- Find a GMRF with local neighborhood and precision matrix Q that optimally represents the GF
- Do the computations using the GMRF representation using numerical methods for sparse matrices

Literature Review

- Besag & Kooperberg (1995)
  - Estimate parameters for a GMRF using the GF covariance matrix
  - Used KL divergence

Parameters to be estimated $\theta_{00}, \theta_{10}, \theta_{20}, \theta_{11}, \theta_{21}, \theta_{22}$

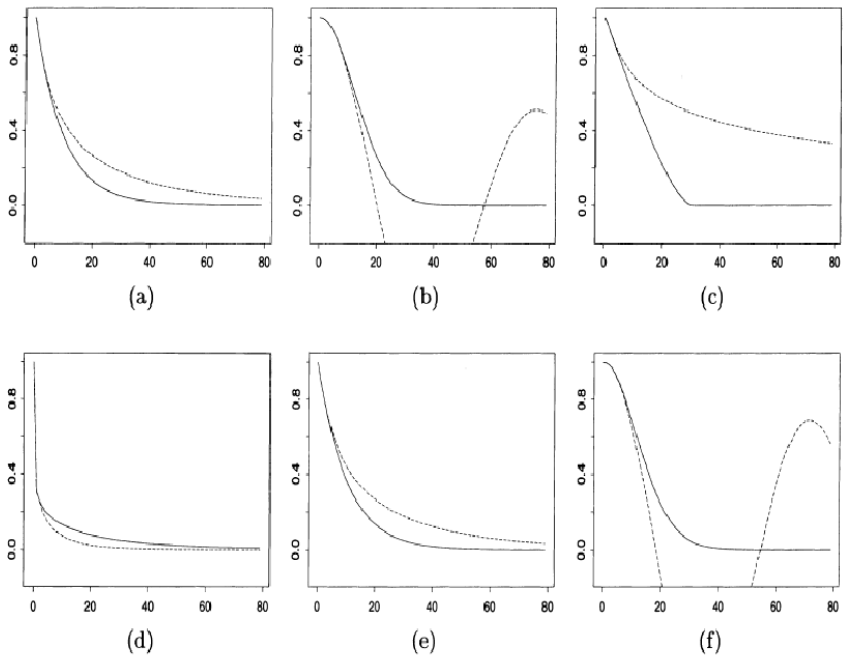


Literature Review

- Rue & Tjelmeland (2002)
  - Show that KL divergence is not optimal to fit GMRF to a GF
  - Define a modified discrepancy metric based on weighted differences between correlations

Results from Rue & Tjelmeland (2002)

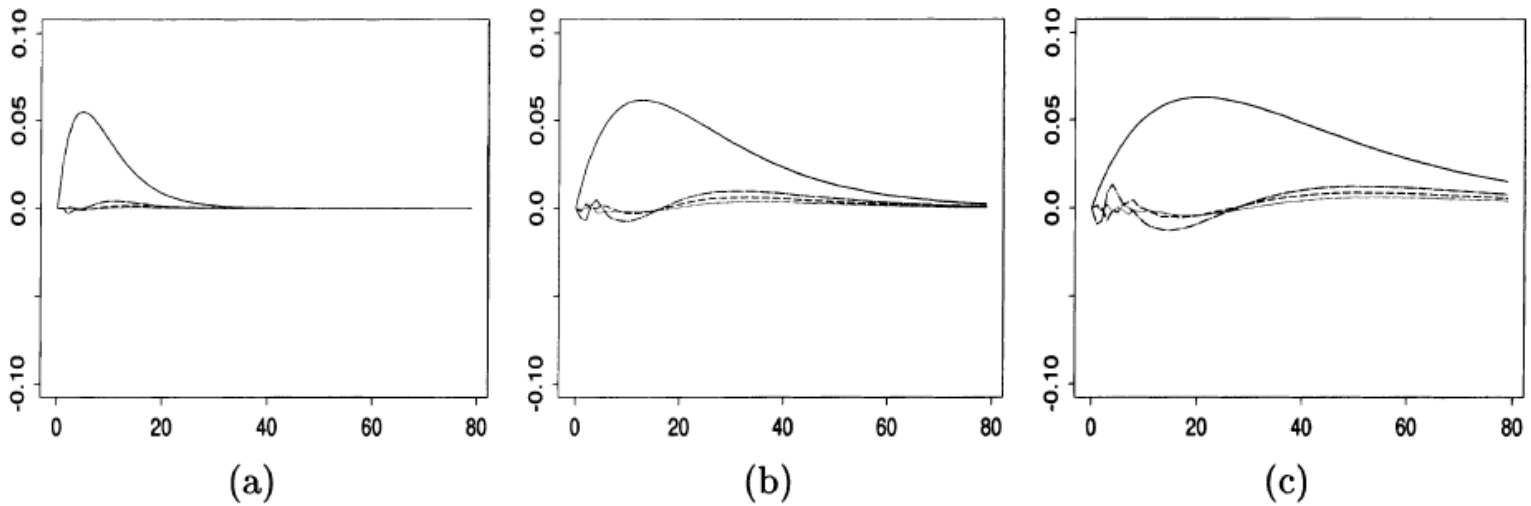
Solid line - Target, Dashed line - Fitted by KLD



(a) Exponential (b)Gaussian (c) Spherical (d) Matern ($\nu = 0.05$) (e) ($\nu = 0.5$) (f) ($\nu = 10$)

Results from Rue & Tjelmeland (2002)contd

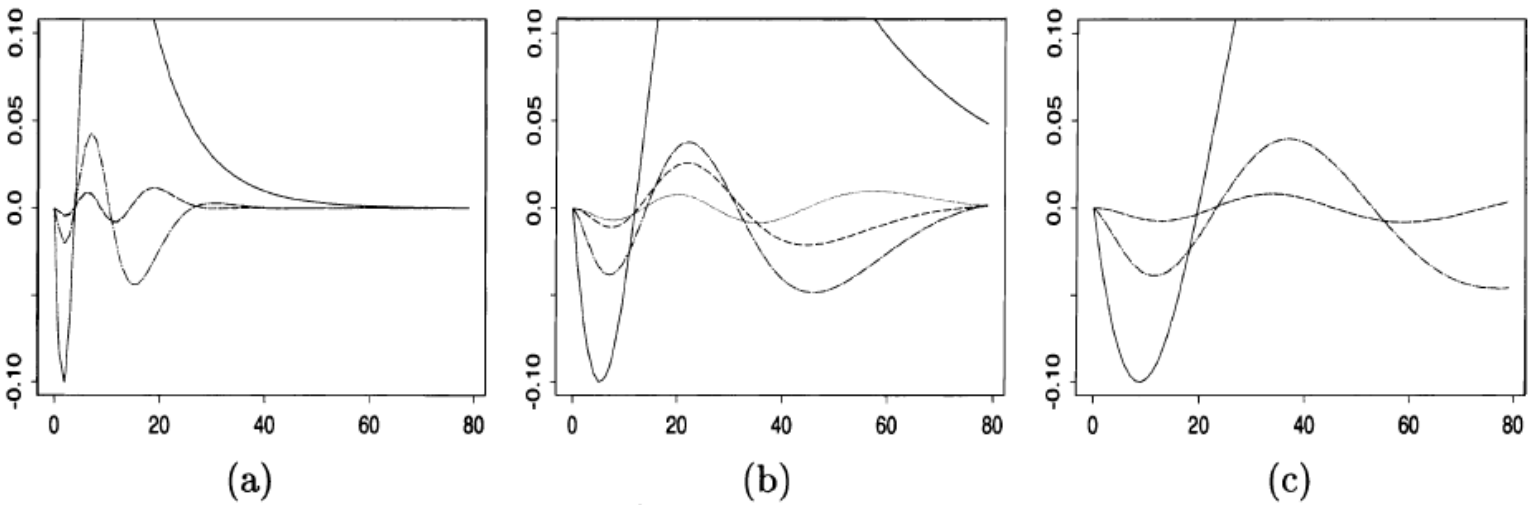
Exponential Correlation: solid line - 3x3 dashed dot line - 5x5 dashed line - 7x7 dotted line - 9x9



(a) $r = 10$ (b)$r = 30$ (c) $r = 50$

Results from Rue & Tjelmeland (2002)contd

Gaussian Correlation: solid line - 3x3 dashed dot line - 5x5 dashed line - 7x7 dotted line - 9x9



(a) $r = 10$ (b)$r = 30$ (c) $r = 50$

Problems with Rue & Tjelmeland (2002)

- Only works for GF's defined on a Lattice or Torus
- Need to precompute fit for different neighborhood sizes and different parameter values of $C(.,.)$

SPDE Approach - Lindgren et al (2011)

- Provide an explicit link between GF with Matern covariance structure and GMRF's on any triangulation of R^d
- Matern Covariance function

$$r(u,v) = \frac{\sigma^2}{\Gamma(\nu) \times 2^{\nu-1}} (\kappa||v-u||)^\nu K_\nu(\kappa||v-u||) \quad (1)$$

- Controlled by the smoothness parameter ν and range parameter κ

SPDE Approach

- SPDE

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}) \tag{2}$$

$$\alpha = \nu + d/2$$

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial {x_i}^2}$$

- Solution to the SPDE, is a GF with a matern covariance structure with parameters ν and κ

SPDE Approach

- Main Idea : Use a finite element representation of a weak solution to (2) to then link the GF to a particular GMRF precision structure

$$x(\boldsymbol{u}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{u}) w_k \tag{3}$$

- ψ_k basis functions defined on the triangulated input domain
- w_k Gaussian distributed weights

What makes this exciting!

- Any GF model defined with a Matern covariance structure such that $\nu + d/2$ is integer valued, can be approximated by a GMRF
- Result is not limited to regular lattices
- Can be extended to GF's on manifolds, nonstationary and anisotropic covariance structures etc

A discrete model on a regular grid

Let $\mathbf{x}$ be a GMRF on a regular two-dimensional lattice indexed by ij, with the Gaussian full conditionals:

$$E(x_{ij}|\mathbf{x}_{-ij}) = \frac{1}{a}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}). \tag{4}$$

$$\text{var}(x_{ij}|\mathbf{x}_{-ij}) = 1/a$$

and $|a| > 4$. This model can be written as the elements of the precision matrix related to a single location

$$\begin{array}{|cc} \text{-1} & \\ \text{a} & \text{-1} \end{array} \tag{5}$$

A discrete model on a regular grid

- Besag (1981) obtained the following approximation result from the general Matérn covariance function letting $\nu \to 0$, $\kappa^2 = a - 4$, and $\sigma^2 = 1/4\pi$:

$$\text{cov}(x_{ij}, x_{i'j'}) \approx \frac{1}{2\pi} K_0\{l\sqrt{(a-4)}\}, \quad l \neq 0,$$

where l is the Euclidean distance between ij and $i'j'$.
"Informally this means that the discrete model defined by expression (4) generates approximate solutions to the SPDE equation (2) on a unit distance regular grid, with $\nu = 0$." (Lindgren et al., 2011).

- Set $\alpha = 1$ and solve the SPDE equation (2), a generalized random field has spectrum:

$$R_1(\mathbf{k}) \propto (a - 4 + ||\mathbf{k}||^2)^{-1},$$

- Some discretized version of the SPDE can be considered as a linear filter with the squared transfer function R_1.
- Replace the noise term $\mathcal{W}(\mathbf{u})$ in SPDE equation(2) by Gaussian noise with spectrum R_1, the resulting solution has spectrum $R_2 = R_1^2$.

- Therefore the GMRF representations for the Matérn fields are convolutions of the coefficients in representation (5).
  - $\nu = 1$

$$\begin{array}{|ccc} 1 & & \\ -2a & 2 & \\ 4 + a^2 & -2a & 1 \end{array}$$

  - $\nu = 2$

$$\begin{array}{|cccc} -1 & & & \\ 3a & -3 & & \\ -3(a^2+3) & 6a & -3 & \\ a(a^2+12) & -3(a^2+3) & 3a & -1 \end{array}$$

Fig. 1 in Lindgren et al. (2011) shows the performance of the approximations when $\nu = 1$.

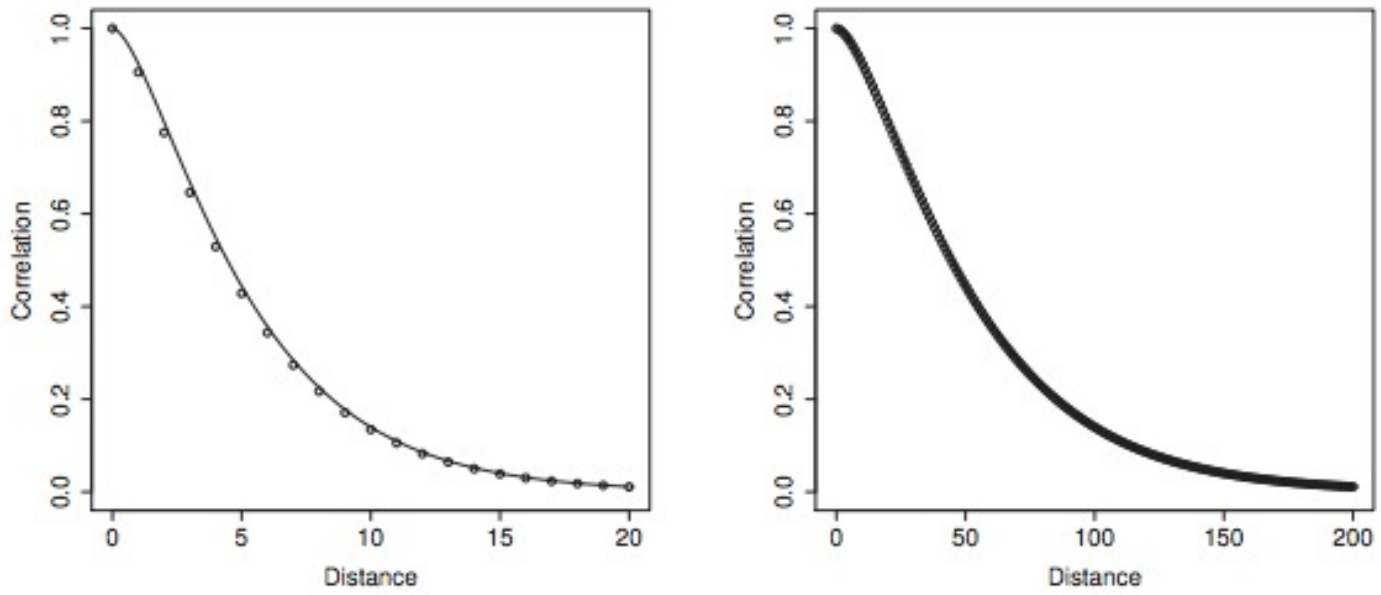


Fig. 1. The Matérn correlations (solid line) for range 10 (left) and 100 (right), and the correlations for the GMRF representation (circles).

Main Result 1

Result 1. The coefficients in the GMRF representation of SPDE equation(2) *on a regular unit distance two-dimensional infinite lattice for* $\nu = 1, 2, \ldots,$ *is found by convolving model* (5) *by itself* ν *times.*

Main Result 2: Triangulation

Goal: extend the SPDE approach to irregular grids.

Build an irregular grid of non-intersecting triangles (share edges or corners) over the space. Corners are denoted **vertices**.

Triangulation process is based on collected data.

- Initialize observed locations as vertices.
- Add vertices using heuristics such as
  - maximize minimum interior angles of triangles.
  - minimize the total number of triangles.

Main Result 2: Triangulation Example

Leukemia survival data in England (Henderson et al, 2002)

- 1043 observations
- Triangulation resulted in 1749 vertices, 3446 triangles.
- fine resolution near observed locations; coarse resolution elsewhere.

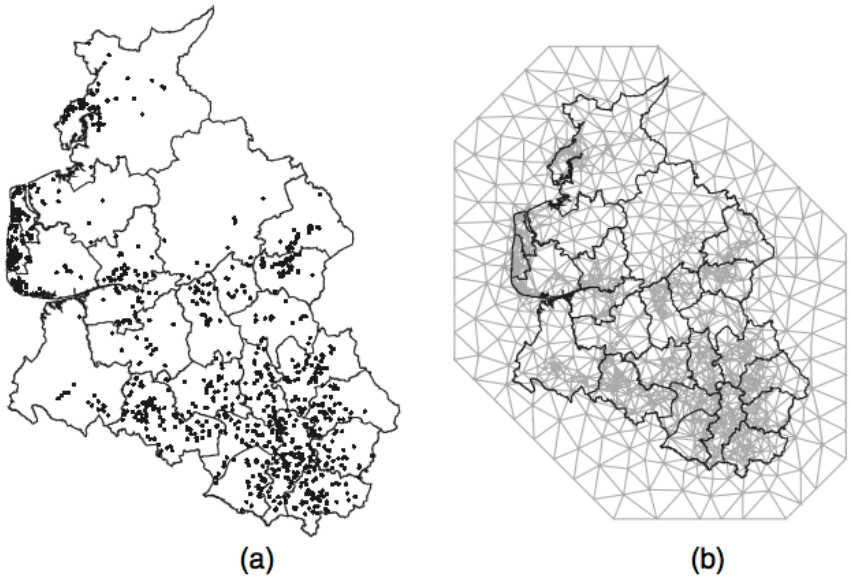


Fig 2 from Lindgren et al: (a) observed locations, (b) triangulation.

Main Result 2: SPDE

Use a stochastic weak formulation of the SPDE:

$$\{\langle \phi_j, (\kappa^2 - \Delta)^{\alpha/2} x \rangle, j = 1, \ldots, m\} \stackrel{\mathrm{d}}{=} \{\langle \phi_j, \mathcal{W} \rangle, j = 1, \ldots, m\}$$

The solution must fulfill the above for every appropriate finite set of test functions $\{\phi_k, j = 1, ..., m\}$.

The inner product is defined as (integral over area of interest):

$$\langle f, g \rangle = \int f(\mathbf{u})\, g(\mathbf{u})\, \mathrm{d}\mathbf{u}$$

Main Result 2: Finite Element Representation

The finite element representation of the solution to the SPDE is:

$$x(\mathbf{u}) = \sum_{k=1}^{n} \psi_k(\mathbf{u}) w_k$$

- continuously indexed.
- n = number of vertices in the triangulation.
- $\{\psi_k\}$ - piecewise linear basis fcns = 1 at vertex k, 0 at other vertices.
- $\{w_k\}$ - Gaussian-distributed weights = value of the field at the vertices.

Then the values of the field...

- at the vertices are determined by the weights.
- at triangle interiors are determined by linear interpolation.

Main Result 2: Finite Element Representation

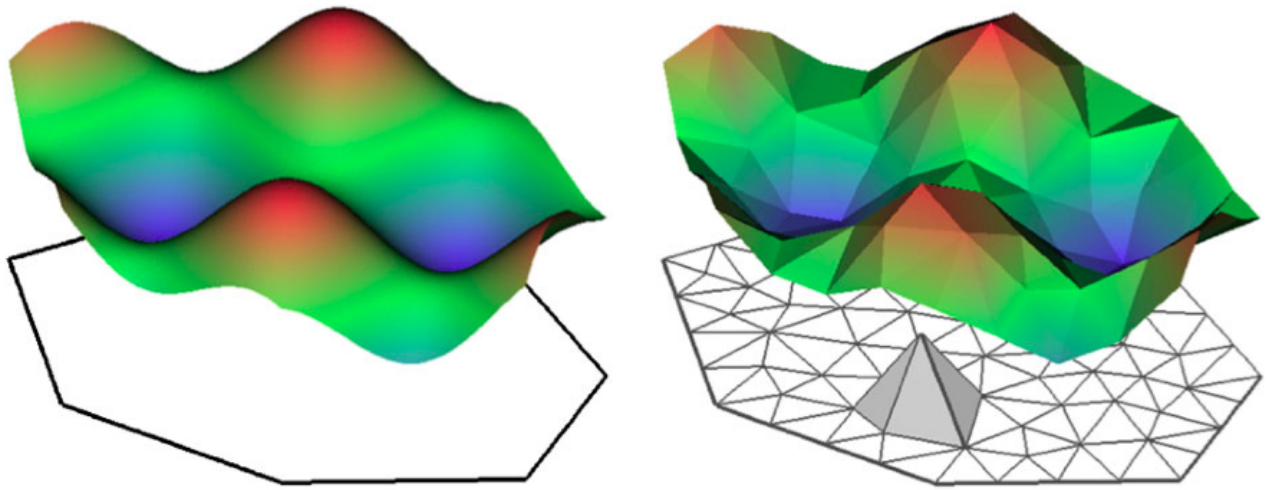


Fig 3 from Calmeletti et al, 2012: Left panel is the continuously indexed spatial random field. Right panel is the corresponding finite element representation.

Main Result 2: Finite Dimension Solution

The full distribution of the solution is determined by the joint distribution of the Gaussian-distributed weights. We would like:

- the precision matrix for the Gaussian weights of the finite element field.
- to approximate that precision matrix with a GMRF precision matrix.

Main Result 2: Finite Dimension Solution

For the finite dimension solution, we find the distribution of weights $\{w_k\}$ that fulfills the weak SPDE for a **specific set of m=n (number of vertices) test functions**. Lindgren et al chose:

- for $\alpha = 1$, $\phi_k = (\kappa^2 - \Delta)^{1/2}\psi_k$ (denoted least squares solution)
- for $\alpha = 2$, $\phi_k = \psi_k$ (denoted Galerkin solution)
- for $\alpha \geq 3$, recursive Galerkin

Define the following (n x n) matrices with entries:

$$C_{ij} = \langle \psi_i \psi_j \rangle,$$
$$G_{ij} = \langle \nabla\psi_i \nabla\psi_j \rangle,$$
$$(\mathbf{K}_{\kappa^2})_{ij} = \kappa^2 C_{ij} + G_{ij}.$$

Main Result 2

With some boundary conditions, we then have the result in Lindgren et al:

Result 2. Let $\mathbf{Q}_{\alpha,\kappa^2}$ be the precision matrix for the Gaussian weights $\mathbf{w}$ as defined in equation (9) for $\alpha = 1, 2, \ldots$, as a function of κ^2. Then the finite dimensional representations of the solutions to equation (2) have precisions

$$\left.\begin{aligned} \mathbf{Q}_{1,\kappa^2} &= \mathbf{K}_{\kappa^2}, \\ \mathbf{Q}_{2,\kappa^2} &= \mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\mathbf{K}_{\kappa^2}, \\ \mathbf{Q}_{\alpha,\kappa^2} &= \mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\mathbf{Q}_{\alpha-2,\kappa^2}\mathbf{C}^{-1}\mathbf{K}_{\kappa^2}, \qquad \text{for } \alpha = 3, 4, \ldots. \end{aligned}\right\} \tag{10}$$

- C and G are non-zero only where basis fcn pairs share common triangles.
- C^{-1} is dense, resulting in a dense Q (finite element solution, not Markov).
- Approximate C by a diagonal matrix with elements $\tilde{c}_{ii} = < \psi_i, 1 >$.
  - The precision matrix is then sparse (GMRF model).
  - And we have an explicit mapping from GF to GMRF.

Main Result 2: Leukemia Example

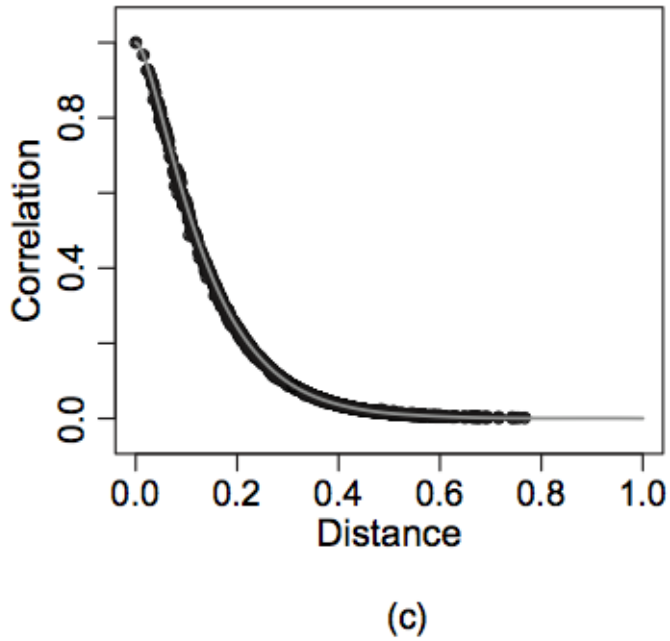


Fig 2c from Lindgren et al: stationary correlation function and the corresponding GMRF approximation for ν=1, approx range=0.26

Extensions

- Matérn fields on manifolds
- Non-stationary fields
- Oscillating covariance functions
- Non-isotropic models and spatial deformations
- Non-separable space-time models

All these extensions give explicit GMRF representations (the finite element representations and equation (10) in Main Result 2), individually or combined. Therefore one can construct the GMRF representations of non-stationary oscillating GFs on the sphere, without requiring any computation beyond the geometric properties of the triangulation.

Extensions

Matérn fields on manifolds

To define Matérn fields on a sphere, we can define Gaussian white noise on $\mathbb{S}^2$ as a zero-mean random GF $W(\cdot)$ with the property that the covariance between $W(A)$ and $W(B)$, for any subsets A and B of $\mathbb{S}^2$, is proportional to the surface integral over $A \cap B$. The GMRF representation of the weak solution only needs to change the definition of the inner product to a surface integral on $\mathbb{S}^2$.

Extensions

Non-stationary fields

Let the the parameters κ^2 and the innovation variance in SPDE equation(2) depend on the coordinate $\mathbf{u}$,

$$\{\kappa^2(\mathbf{u}) - \Delta\}^{\alpha/2}\{\tau(\mathbf{u})x(\mathbf{u})\} = \mathcal{W}(\mathbf{u}) \tag{6}$$

It is non-stationary when one or both parameters are non-constant. If we let the parameters $\kappa^2(\mathbf{u})$ and $\tau(\mathbf{u})$ vary slowly, the local interpretation of equation (6) is still a Matérn field. With a similar approach for the stationary case, the GMRF representation of equation (6) can be obtained.

Extensions

Oscillating covariance functions

Consider a complex version of SPDE equation(2), let $\alpha = 2$,

$$\{\kappa^2 \exp(i\pi\theta) - \Delta\}\{x_1(\mathbf{u}) + ix_2(\mathbf{u})\} = \mathcal{W}_1(\mathbf{u}) + i\mathcal{W}_2(\mathbf{u}), \qquad (7)$$

where $\mathcal{W}_1(\mathbf{u})$, $\mathcal{W}_2(\mathbf{u})$ are two independent white noise fields, θ is the oscillation parameter, $0 \leq \theta < 1$.

Non-isotropic models and spatial deformations

Consider SPDEs with a non-isotropic Laplacian and a directional derivative term. Use the deformation method for non-stationary covariances (Sampson and Guttorp, 1992).

Non-separable space-time models

References

- Besag, J. and C. Kooperberg, 1995. On conditional and intrinsic autoregressions. Biometrika 82(4), 733-746.
- Lindgren, F., Rue, H., Lindstrm, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society, Series B: Statistical Methodology 73 (4), 423498 (with discussion).
- Cameletti 2012 Michela Cameletti, Finn Lindgren, Daniel Simpson, Hvard Rue, 2012. Spatio Temporal modelling of particulate matter concentration through the SPDE approach. Advances in Statistical Analysis, 97 (2), 109-131.
- Rue, H. and H. Tjelmeland, 2002.Fitting Gaussian Markov random fields to Gaussian fields, Scandinavian Journal of Statistics 29(1), 31-50.

Outline

Why you should like your INLAs

Introduced by Rue, Martino, and Chopin (2009) in JRSSB.

- "to provide accurate and fast deterministic approximations to all, or some of,the n posterior marginals for x_i [latent variables] . . . plus possibly the posterior marginals for θ or some of its components . . . "
- ". . . MCMC sampling remains painfully slow from the end user's point of view."
- "We argue, however, that for a given computational cost, the deterministic approach that is developed in this paper outperforms MCMC algorithms to such an extent that, for latent Gaussian models, resorting to MCMC sampling rarely makes sense in practice."
- "It is our experience that INLA outperforms without comparison any MCMC alternative, in terms of both accuracy and computational speed."

Setting where INLA is Appropriate

Recall the Bayesian Setting:

- Data: $\mathbf{y} = (y_i) \quad i \in \mathcal{I}$.
- Likelihood: $\prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta})$
  where $\mathbf{x}$ are latent variables and $\boldsymbol{\theta}$ are hyper-parameters.
- Priors: $\pi(\mathbf{x}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$.

Assumptions for INLA:

- $\dim(\mathbf{x}) = n$ is around 10^2 to 10^5 (large) and
  $\dim(\boldsymbol{\theta})$ is 2 to 6 (small)
- $\pi(\mathbf{x}|\boldsymbol{\theta}) =$ GMRF with sparse precision matrix $\mathbf{Q}(\boldsymbol{\theta})$.

Posterior

The posterior is:

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta})$$

$$\propto \pi(\boldsymbol{\theta})|Q(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_i \log(\pi(y_i|x_i, \boldsymbol{\theta}))\right)$$

Our goal is to find posterior marginals $\pi(x_i|\mathbf{y})$, $\pi(\theta_j|\mathbf{y})$.

We will make a series of approximations in order to obtain these posterior marginal distributions, including an approximation for $\pi(\boldsymbol{\theta}|\mathbf{y})$ as an intermediate step.

Step 1: Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$

Begin with a Laplace approximation to the marginal posterior of $\pi(\boldsymbol{\theta}|\mathbf{y})$ (see Tierney and Kadane, 1986):

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left.\frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}\right|_{x=x^*(\boldsymbol{\theta}),} \tag{3}$$

where $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to the full conditional of $\mathbf{x}$ and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\mathbf{x}$ for a given $\boldsymbol{\theta}$.

Step 1: Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$

To find $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, we are attempting to approximate

$$\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_i g_i(x_i)\right),$$

where in our case

$$g_i(x_i) = \log(\pi(y_i|x_i, \boldsymbol{\theta})),$$

with a Gaussian distribution that has the same mode and curvature at the mode.

Step 1: Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$

The following steps allow us to do this:

- Initial Guess: $\boldsymbol{\mu}^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_n^{(0)})$.
- Approximate $g_i(x_i) = g_i(\mu_i^{(0)}) + b_i x_i - \frac{1}{2} c_i x_i^2$, where b_i and c_i depend on $\boldsymbol{\mu}^{(0)}$.
- Obtain a Gaussian approximation with precision matrix $\mathbf{Q} + \text{diag}(\boldsymbol{c})$ and mode $\boldsymbol{\mu}^{(1)}$ given by the solution $(\mathbf{Q} + \text{diag}(\boldsymbol{c}))\boldsymbol{\mu}^{(1)} = \boldsymbol{b}$.
- Repeat with $\boldsymbol{\mu}^{(1)}$ in place of $\boldsymbol{\mu}^{(0)}$ and repeat until convergence.

Let this converged Gaussian be $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$.
The mode of $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is $x^*(\boldsymbol{\theta})$ from (3).

Step 1: Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$

We then have

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}^*(\boldsymbol{\theta})|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x}^*(\boldsymbol{\theta}), \boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}^*(\boldsymbol{\theta})|\boldsymbol{\theta}, \mathbf{y})},$$

all of which can be evaluated exactly (though it is still an approximation) at several different $\boldsymbol{\theta}$ values (a grid).

Use this approximation to approximate each $\pi(\theta_i|\mathbf{y})$ by summing out the other dimensions from $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$.

Step 1: Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$

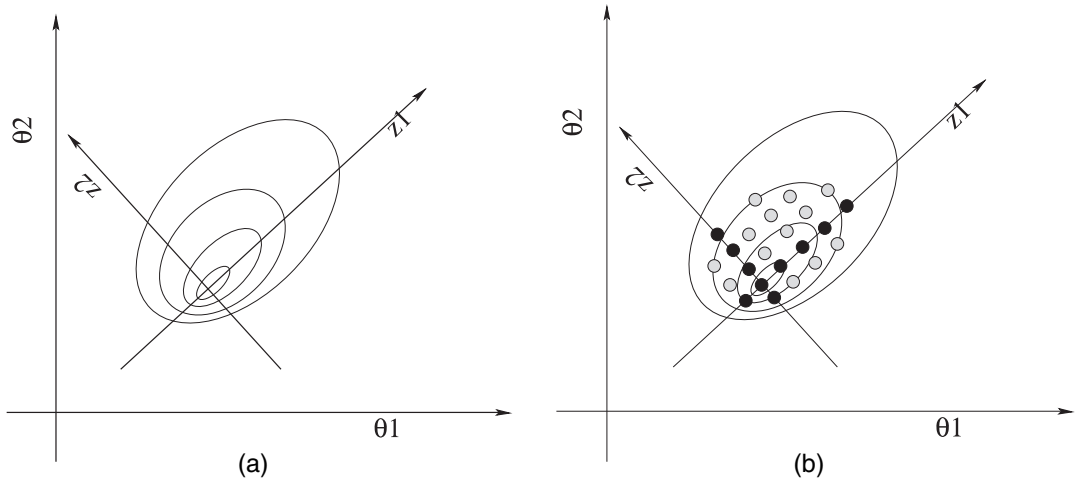


Fig. 1. Illustration of the exploration of the posterior marginal for θ: in (a) the mode is located and the Hessian and the co-ordinate system for $\mathbf{z}$ are computed; in (b) each co-ordinate direction is explored (●) until the log-density drops below a certain limit; finally the new points (●) are explored

How to find a good grid:

- Find the mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$ with numerical maximization techniques.
- Find important directions to explore with the Hessian matrix.
- Explore $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ along the important directions and fill in the grid with points Δ_k apart.

Step 2: Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$

To approximate $\pi(x_i|\mathbf{y})$, we will use a sum over a fine grid of points:

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k. \quad (5)$$

Note that $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ has already been established. We then must find the approximation $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$.

Step 2: Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$

There are three different approximations we could consider for $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$:

- Using a Gaussian approximation (using the recursions we used before)
- Using a Laplace approximation (see paper with their modifications)
- Using a simplified Laplace approximation (again, see paper)

Discussion

A large discussion of examples, approximation errors, and applications is presented in the paper.

There are is also a large amount of discussion from other statisticians at the end of the paper.

References:

- Rue H., Martino S., and Chopin N., (2009) "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society B*, Part 2, p. 319-392.
- Radu's INLA slides for Stat 8540.