Covariance Tapering for Anisotropic Nonstationary Gaussian Random Fields with Application to Large Scale Spatial Data Sets

Abolfazl Safikhani and Yimin Xiao

Department of Statistics and Probability,
Michigan State University

Abstract

Estimating the covariance structure of spatial random processes is an important step in spatial data analysis. Maximum likelihood estimation is a popular method in spatial models based on Gaussian random fields. But calculating the likelihood in large scale data sets is computationally infeasible due to the heavy computation of the precision matrix. One way to mitigate this issue, which is due to Furrer et al. (2006), is to "taper" the covariance matrix. While most of the results in the current literature focus on isotropic tapering for stationary Gaussian processes, there are many cases in application that require modeling of anisotropy and/or nonstationarity. In this article, we propose a nonstationary parametric model, in which the underlying Gaussian random field may have different regularities in different directions, thus can be applied to model anisotropy. Using the theory of equivalence of Gaussian measures under nonstationary assumption, strong consistency of the tapered likelihood based estimation of the variance component under fixed domain asymptotics are derived by putting mild conditions on the spectral behavior of the tapering covariance function. The procedure is illustrated with numerical simulation.

Keywords: Anisotropic covariance tapering, Nonstationary Gaussian random fields, Maximum likelihood estimation, Equivalence of Gaussian measures, Large scale spatial data sets

1. Introduction

Spatial Statistics is nowadays a very active research field in Statistics, and has many applications in Geology, Agricultural Science, Environmental Science, Climate data, etc. (Cressie, 1993; Stein, 1999). A common problem in this field is the estimation of the covariance structure in spatial models based on Gaussian random fields. Likelihood-based estimators are the most popular method for estimating the covariance parameters. However, in large scale spatial data sets, calculating the likelihood is computationally infeasible due to the heavy calculation of the precision matrix. There are different ways to overcome this issue. The first idea is to set the off-diagonal entries of the covariance matrix to zero, or to keep the first k subdiagonal entries for some integer k and put the rest of them to be zero, which is called banding. In this way, the resulting matrix becomes sparse, and one can use the existing algorithms dealing with sparse matrices to handle the new covariance matrix efficiently. But, the problem with banding is that the final covariance matrix may not be positive definite, which is a huge drawback since all the theoretical covariance matrices must be positive definite. An alternative approach is to multiply the covariance function by a positive definite compactly supported correlation function. By this way, the resulting covariance matrix is again sparse, but still positive definite. This is called covariance tapering (Furrer *et al.*, 2006). A natural question is that how we can use tapering to construct consistent estimates for the covariance parameters in spatial regression models. Kaufman et al. (2008) proposed two different likelihood-based estimations of the parameters in the Matern covariance function, and proved the strong consistency of the estimates using the results in the equivalence of stationary Gaussian measures (Skorokhod and Yadrenko, 1973). See also Du et al. (2009) for more results in the asymptotics of tapered maximum likelihood estimators. There are two strong assumptions under which the consistency is proved, isotropy, and stationarity of the underlying Gaussian random field. However, there are many cases in application in which the data sets resemble anisotropy and nonstationarity. There have been interests in general (not necessarily in the context of covariance tapering) for dropping these conditions in spatial modellings (See Cressie and Johannesson, 2008, Anderes and Stein, 2011 and Hitczenko and Stein, 2012). In this article, we relax both of these conditions simultaneously. For that purpose, in section 2, we introduce a class of parametric models for spatial modeling which are anisotropic and non-stationary. In section 3, we state the main result of the paper which is deriving strong consistency of the tapered likelihood-based estimates of the variance parameter. In the last section, we demonstrate the procedure proposed here via numerical simulation.

2. Preliminary

In this section, we introduce a class of Gaussian random fields which can be used as spatial models. Any Gaussian field model needs a mean structure and a covariance structure to be uniquely defined. However, finding new models for covariance structure might be hard since one can only choose covariance structures from the family of positive definite functions and it is complicated to verify the positive definiteness. An alternative is to use the spectral representations of positive definite functions or variograms. We consider here specifically Gaussian field models with stationary increments. By applying the results in Yaglom (1957), the covariance functions of Gaussian random fields with stationary increments on $\mathbb{R}^d$ can be determined by a symmetric non-negative measures on $\mathbb{R}^d \setminus \{0\}$ satisfying certain integrability condition (See e.g., Xue and Xiao, 2011 for more details). This measure is called spectral measure. If this measure is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, we will call its Radon-Nikodym derivative the "spectral density".

Now, we propose the following spectral density:

$$f(\lambda) = \frac{\sigma^2}{\left(1 + \sum_j |\lambda_j|^{H_j}\right)^{Q+\nu}} \tag{1}$$

where $\lambda = (\lambda_1, ..., \lambda_d) \in \mathbb{R}^d$, $Q = \sum_j^d 1/H_j$, and σ^2, ν and H_j's are all positive parameters. σ^2 is the variance component, and H_j's are related to the smoothness of the model in different directions. Figures 1 and 2 are realizations of such a Gaussian field over the two-dimensional grid $[0,1]^2$ with increments of 0.02 with parameters $H_1 = 1.5$, $H_2 = 3$, and $\nu = 0.4$.

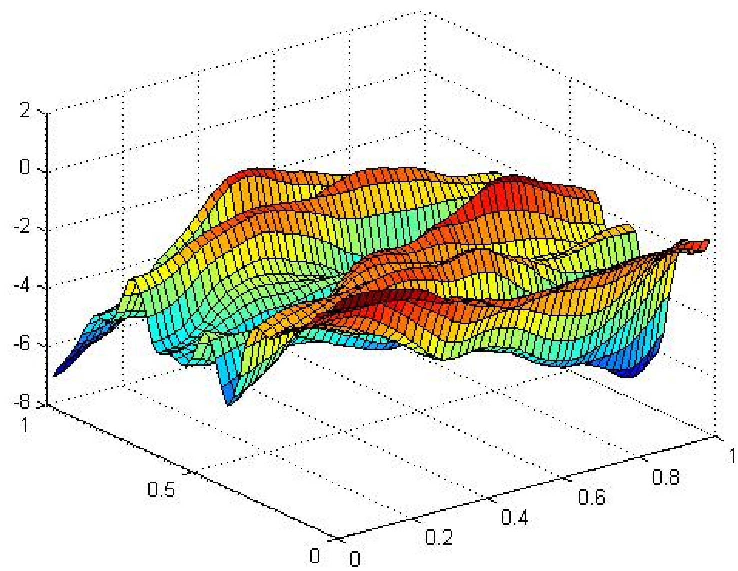

Figure 1: 3D Simulated Surface on $[0, 1]^2$ with increments 0.02

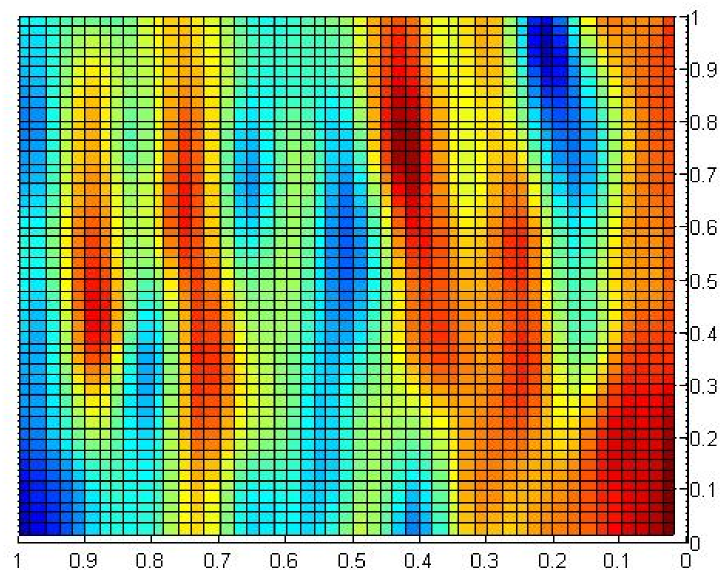

Figure 2: 2D Projection of the Simulated Surface on $[0, 1]^2$ with increments 0.02

Having different parameters, H_j, $j = 1, ..., d$, for different components of $\lambda \in \mathbb{R}^d$, allows the underlying random field to have different regularities in different directions, and therefore, this model is able to capture anisotropy appearing in spatial data sets. (See more about fractal and smoothness properties of these models in Xue and Xiao, 2011. See also Stein, 2005).

3. Main Result

In this section, we state the main result of the article, which is the consistency of the tapered likelihood estimation of the variance parameter σ^2 in the model introduced in the last section under the fixed domain asymptotics. The following theorem, shows how to get equivalent Gaussian measures in tapering set up, which is necessary in proving strong consistency, by putting mild conditions on the spectral behavior of the tapering covariance function.

Theorem 1 *Suppose $\{Z(t), t \in T\}$, where $T \subseteq \mathbb{R}^d$ is a mean zero Gaussian random field with stationary increments, and the spectral density of the form 1 for some positive constants σ^2, ν, and H_j, $j = 1, ..., d$. Let's denote the covariance function of the process by $K_0(x)$. If $K_1(x) = K_0(x)K_t(x)$ where $K_t(x)$ is a correlation function, and its Fourier transform exits and satisfies the following condition:*

$$f_t(\lambda) \leq \frac{M}{\left(1 + \sum_j |\lambda_j|^{H_j}\right)^{Q+\nu+\epsilon}} \tag{2}$$

for some $\epsilon > max\{Q/2, 1/min\{H_j/2\}\}$, and $M > 0$. Now, if $H_j > 1, j = 1, ..., d$, then the Gaussian measures induced by K_0 and K_1 will be equivalent on the paths of $\{Z(t), t \in T\}$ for any bounded subset $T \subset \mathbb{R}^d$.

Remark 1 *Proof of this theorem is similar to that of Theorem 1 in (Kaufman et al., 2008). However, due to the nonstationarity assumption, it relies on the study of the equivalence of Gaussian random fields with stationary increments, which is discussed in (Safikhani and Xiao, 2014).*

Suppose one observes the random process at locations s_j's $j = 1, ..., n$ on some bounded domain in $\mathbb{R}^d$, and $K_t(x)$ is the tapering covariance function. Fix a tapering parameter $\gamma > 0$. Define the tapering matrix $T(\gamma)$ to be $T(\gamma)_{ij} = K_t(\|s_i - s_j\|; \gamma)$. This means that if $\|s_i - s_j\| > \gamma$, then $T(\gamma)_{ij} = 0$. Denote the true covariance matrix of the process by $\Sigma(\theta)$. Now, the idea of tapering is to use $\Sigma(\theta) \circ T(\gamma)$ as the new covariance matrix in the likelihood of the process. Here, $\circ$ means the element wise matrix product. Therefore, the Tapered Maximum likelihood estimator (Tapered MLE) of σ^2 will be:

$$\begin{aligned}\widehat{\sigma_t^2} &= argmax_{\sigma^2} \ \text{Tapered likelihood} \\ &= argmax_{\sigma^2} \ \left(-\frac{1}{2}\log|\Sigma(\theta) \circ T(\gamma)| - \frac{1}{2}Z'(\Sigma(\theta) \circ T(\gamma))^{-1}Z\right),\end{aligned}$$

where Z is the vector of observed values of the random field at the specified locations. Evaluating the inverse of $\Sigma(\theta) \circ T(\gamma)$ is much faster than $\Sigma(\theta)$ in many cases, especially with large n. This is the benefit of using Tapered MLE. However, this estimator might be biased in general. In (Kaufman *et al.*, 2008), another estimator based on the same idea of tapering is presented which is unbiased. We call it here the Adjusted Tapered MLE, and it is defined as follows:

$$\begin{aligned}\widehat{\sigma_{adj.}^2} &= argmax_{\sigma^2} \ \text{Adjusted tapered likelihood} \\ &= argmax_{\sigma^2} \ \left(-\frac{1}{2}\log|\Sigma(\theta) \circ T(\gamma)| - \frac{1}{2}Z'\left((\Sigma(\theta) \circ T(\gamma))^{-1} \circ T(\gamma)\right)Z\right).\end{aligned}$$

We will see in section 4 that the Adjusted Tapered MLE performs better comparing to the Tapered MLE due to its unbiasedness. Theorem 1 is the key part in proving the consistency of the tapered maximum likelihood estimators for the variance component (σ^2). The outline of the proof is similar to Theorem 2 in (Kaufman *et al.*, 2008), and also it relies on the above theorem. We omit the proof here.

4. Simulation Results

In this section, we illustrate the methods discussed in previous parts by applying them into some simulated spatial data sets. For this purpose, we simulated 1000 data sets, each consisting of a multivariate Gaussian vector of length 14×14. The locations are two-dimensional grid over $[0, 1]^2$ with increments 0.07. We added a random noise in each coordinate, uniformly distributed on $[-0.01, 0.01]$. We used the spectral density 1 to generate the normal vector with $H_1 = 1.5$, $H_2 = 2$, $\nu = 0.5$, and $\sigma^2 = 1$. Figure 3 shows a simulated surface using the above parameters over the specified grid.

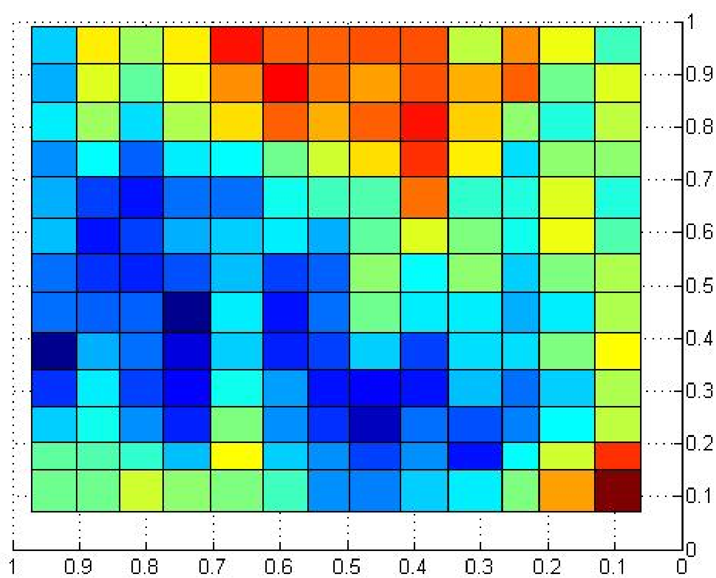

Figure 3: Simulated Surface on $[0, 1]^2$ with increments 0.07

We used the same tapering function used in (Kaufman *et al.*, 2008) with different tapering parameter $\gamma = 0.7$, 0.4, 0.2, respectively. We applied three diferent estimators for estimating σ^2: MLE, Tapered MLE and Adjusted Tapered MLE (See (Kaufman *et al.*, 2008) for details.). Table 1 shows the results for this procedure for different tapering parameters. The values in the table are the averages of the estimated value over the 1000 repetitions with their standard errors in the bracket. As it was expected, by decreasing the tapering parameter, the Tapered MLE becomes more and more biased. However, the Adjusted Tapered MLE is almost unbiased regardless of the changes in the tapering parameter.

Simulations show that comparing to MLE, Adjusted Tapered MLE is a good alternative for estimating the covariance parameters in spatial data

Table 1: Results for estimation of σ^2

$\widehat{\sigma^2}$	$\gamma = 0.7$	$\gamma = 0.4$	$\gamma = 0.2$
MLE	0.999(0.10)	0.999(0.10)	0.999(0.10)
Tapered MLE	0.437(0.06)	0.409(0.07)	0.400(0.17)
Adjusted Tapered MLE	1.025(0.53)	1.019(0.47)	1.031(0.61)

analysis, and it is computationally feasible. Further, this method has the potential to be generalized to more complicated models which have nonstationarity and anisotropy.

There are other methods dealing with covariance estimation in large spatial data sets. For example, in Cressie and Johannesson, 2008, nonstationary spatial models are defined through some fixed basis functions, and then weighted least squares method (rather than the MLE approach) is chosen for covariance parameter estimation with the emphasis of finding the best linear unbiased prediction (BLUP).

References

Anderes, E. B., Stein, M. L. (2011), "Local likelihood estimation for nonstationary random fields". *Journal of Multivariate Analysis*, Vol. 102(3):506-520.

Cressie, N. (1993), *Statistics for Spatial Data*, Wiley Series in Probability and Statistics, USA.

Cressie, N., Johannesson, G. (2008), "Fixed rank kriging for very large spatial data sets". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 70(1):209-226.

Du, J., Zhang, H., Mandrekar, V. S. (2009), "Fixed-domain asymptotic properties of tapered maximum likelihood estimators". *The Annals of Statistics*, Vol. 37(6A), 3330-3361.

Furrer, R., Genton M., and Nychka D. (2006), "Covariance tapering for interpolation of large spatial datasets". *Journal of Computational and Graphical Statistics*, Vol. 15(3).

Hitczenko, M., Stein, M. L. (2012), "Some theory for anisotropic processes on the sphere". *Statistical Methodology*, Vol. 9(1): 211-227.

Kaufman C., Schervish M., Nychka D. (2008), "Covariance tapering for likelihood-based estimation in large spatial data sets". *Journal of The American Statistical Association*, Vol. 103(484).

Safikhani A., Xiao Y. (2014), "Spectral conditions for equivalence of Gaussian random fields with stationary increments". *In Preparation.*

Skorokhod A., Yadrenko M. (1973), "On absolute Continuity of measures corresponding to homogeneous Gaussian fields". *Theory of Probability* *and Its Application*, Vol. 18(1): 27-40.

Stein M. L. (2005), "Space-time covariance functions". *Journal of the American Statistical Association*, Vol. 100 (469): 310-321.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

Xue Y., Xiao Y. (2011), "Fractal and smoothness properties of space-time Gaussian models". *Frontiers of Mathematics in China*, Vol. 6(6): 1217-1248.

Yaglom, A. M. (1957), Some classes of random fields in n-dimensional space, related to stationary random processes. *Th. Probab. Appl.* **2**, 273–320.