Journal of Multivariate Analysis 149 (2016) 177–191

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotic properties of multivariate tapering for estimation and prediction

Reinhard Furrer [a,b,*], François Bachoc [c], Juan Du [d]

[a] Institute of Mathematics, University of Zurich, Switzerland
[b] Institute for Computational Science, University of Zurich, Switzerland
[c] Toulouse Mathematics Institute, University Paul Sabatier, France
[d] Department of Statistics, Kansas State University, United States

HIGHLIGHTS

- We present a unified asymptotic framework for tapering multivariate spatial fields.
- Based on weak assumptions, the one-taper maximum likelihood estimator preserves the consistency of the untapered one.
- Prediction using tapering preserves asymptotically the mean squared prediction error.
- For prediction, the computationally attractive one-taper approach is sufficient.

ARTICLE INFO

Article history:
Received 4 June 2015
Available online 25 April 2016

AMS subject classifications:
primary 62M30
secondary 62F12

Keywords:
One-taper likelihood
Gaussian random field
Domain increasing
Sparse matrix

ABSTRACT

Parameter estimation for and prediction of spatially or spatio-temporally correlated random processes are used in many areas and often require the solution of a large linear system based on the covariance matrix of the observations. In recent years, the dataset sizes to which these methods are applied have steadily increased such that straightforward statistical tools are computationally too expensive to be used. In the univariate context, tapering, i.e., creating sparse approximate linear systems, has been shown to be an efficient tool in both the estimation and prediction settings. The asymptotic properties are derived under an infill asymptotic setting. In this paper we use a domain increasing framework for estimation and prediction using multivariate tapering. Under this asymptotic regime we prove that tapering (one-tapered form) preserves the consistency of the untapered maximum likelihood estimator and show that tapering has asymptotically the same mean squared prediction error as using the corresponding untapered predictor. The theoretical results are illustrated with simulations.

1. Introduction

Parameter estimation for and smoothing or interpolation of spatially or spatio-temporally correlated random processes are used in many areas and often require the solution of a large linear system based on the covariance matrix of the observations. In recent years, the dataset sizes to which these methods are applied have steadily increased such that straightforward statistical tools are computationally too expensive to be used. For example, a typical Landsat 7 satellite image

* Correspondence to: Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
E-mail addresses: reinhard.furrer@math.uzh.ch (R. Furrer), francois.bachoc@math.univ-toulouse.fr (F. Bachoc), dujuan@ksu.edu (J. Du).

http://dx.doi.org/10.1016/j.jmva.2016.04.006
0047-259X/

consists of more than 34 million pixels (30 m resolution for an approximate scene size of 170 km × 183 km; source landsat.usgs.gov). Hence, classical spatial and spatio-temporal models for such data sizes cannot be handled with typical soft- and hardware. Thus, one typically relies on approximation approaches. In the univariate context, tapering, i.e., creating sparse approximate linear systems through a direct product of the presumed covariance function and a positive definite but compactly supported correlation function, has been shown to be an efficient tool in both the estimation and prediction settings.

The vast majority of the theoretical work on univariate tapering has been placed in an infill-asymptotic setting using the concept of Gaussian equivalent measures and mis-specified covariance functions set forth in a series of papers by M. Stein [32–35]. Subsequently, Furrer et al. [16], Kaufman et al. [21], Du et al. [13] and Wang and Loh [40] have assumed a second-order stationary and isotropic Matérn covariance to show asymptotic optimality for prediction, consistency, and asymptotic efficiency for estimation. Recently, Stein [38] has extended these results to other covariance functions by placing appropriate conditions on the spectral density of the covariance.

In the infill-asymptotic setting, it is essentially sufficient to match the degree of differentiability at the origin of an appropriately chosen taper function with the smoothness of the covariance at the origin. Loosely speaking, for prediction, the predictor based on tapered covariances has the same convergence rate as the optimal predictor and the naive formula for the prediction kriging variance has the correct convergence rate as well (Theorem 2.1 of [16], Theorem 1 of [38]).

For estimation, Kaufman et al. [21] introduced the concept of one-taper and two-taper likelihood equations. In a one-taper setting only the covariance is tapered while for two-tapered both the covariance and empirical covariance are affected. The one-taper equation results in biased estimates while the two-taper equation is an estimating equation approach and is thus unbiased. The price of unbiased estimates is a severe loss of the computational efficiency intended through tapering (see, e.g., Table 2 of [21] or Fig. 2 of [31]).

Extending the idea of tapering to a multivariate setting is not straightforward. The infill-asymptotic setting does not allow one to 'embed' the multivariate framework in a univariate one (e.g., as in [30] for Gaussian Markov random fields). Ruiz-Medina and Porcu [29] introduced the concept of multivariate Gaussian equivalent measures, but the conditions are difficult to verify and their practical applicability is not entirely convincing. Several authors have recently approached the problem using a increasing-domain setting [31,9]. The main advantage of this alternative sampling scheme is that we are not bound to Matérn type covariance functions nor to tapers that satisfy the taper condition (i.e., sufficiently differentiable at the origin and at the taper length). More so, we will show that for collocated data, other practical tapers can be described. The main disadvantage is the somewhat less-intuitive conceptual framework. For example, in the case of heavy metal contents in sediments of a lake, infill-asymptotics can be mimicked by taking more and more measurements. In a increasing-domain setting, this is not possible. On the other hand asymptotics is a theoretical concept and in practice only a finite number of observations are available.

The main contributions of this paper are as follows: (i) under weak conditions on the covariance matrix function and the taper matrix function form we show that in a increasing-domain framework the tapered maximum likelihood estimator preserves the consistency of the untapered likelihood estimator; (ii) the difference between the (integrated) mean squared prediction error of the tapered and the untapered converges in probability to zero, even when prediction is based on estimated parameters. Note that although we require that the taper range increases, no rate assumption is necessary; (iii) numerical simulations illustrate that the approach has very appealing finite sample properties, especially for prediction with plugin estimates we find only a very small loss in efficiency.

This paper is structured as follows: Section 2 introduces basic notation and relevant definitions. The main results are given in Section 3. Section 4 illustrates the methodology using an extensive simulation study. Concluding remarks are given in Section 5. Proofs and technical results are presented in the Appendix.

Note that compared with directly using compactly supported covariance functions, tapering has several advantages. Our modeling experience has shown that the practical dependence structure is often larger or much larger than what can be handled computationally and additional approximations would be needed anyway. We see tapering as a computational approximation that does not alter the statistical model. The taper range, i.e., degree of tapering, depends on the availability of memory and computing power and thus changes when the analysis is carried out on different computers or at some later time with improved hardware.

2. Notation and setting

We denote vectors and matrices with bold lower and upper case symbols. Random variables and processes are denoted with upper case symbols and random vectors and vector processes are denoted with bold upper case symbols. For $\boldsymbol{x} \in \mathbb{R}^m$, we let $|\boldsymbol{x}| = \max_{i=1,\ldots,m} |x_i|$ and $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^m x_i^2}$.

The singular values of a $n \times n$ real matrix $\mathbf{A} = (a_{ij})$ are denoted by $\rho_1(\mathbf{A}) \geq \cdots \geq \rho_n(\mathbf{A}) \geq 0$ and, in the case when $\mathbf{A}$ is symmetric, the eigenvalues are denoted by $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$. The spectral norm is given by $\rho_1(\mathbf{A})$ and $\|\mathbf{A}\|_F^2 = \sum_{i,j} |a_{ij}|^2$ denotes the Frobenius norm.

For a sequence of random variables X_n, we write $X_n = o_p(1)$ when X_n converges to 0 in probability as $n \to \infty$ and we write $X_n = O_p(1)$ when X_n is bounded in probability as $n \to \infty$.