Journal of Multivariate Analysis 150 (2016) 245–266

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On the consistency of inversion-free parameter estimation for Gaussian random fields[☆]

Hossein Keshavarz [a,*], Clayton Scott [b,a], XuanLong Nguyen [a,b]

[a] *Department of Statistics, University of Michigan, United States*

[b] *Department of Electrical Engineering and Computer Science, University of Michigan, United States*

ARTICLE INFO

Article history:
Received 15 January 2016
Available online 4 July 2016

AMS subject classification:
primary 62M30, 62M40
secondary 60G15

Keywords:
Inversion-free estimation
Covariance function
Stationary Gaussian process
Asymptotic analysis

ABSTRACT

Gaussian random fields are a powerful tool for modeling environmental processes. For high dimensional samples, classical approaches for estimating the covariance parameters require highly challenging and massive computations, such as the evaluation of the Cholesky factorization or solving linear systems. Recently, Anitescu et al. (2014) proposed a fast and scalable algorithm which does not need such burdensome computations. The main focus of this article is to study the asymptotic behavior of the algorithm of Anitescu et al. (ACS) for regular and irregular grids in the increasing domain setting. Consistency, minimax optimality and asymptotic normality of this algorithm are proved under mild differentiability conditions on the covariance function. Despite the fact that ACS's method entails a non-concave maximization, our results hold for any stationary point of the objective function. A numerical study is presented to evaluate the efficiency of this algorithm for large data sets.

1. Introduction

Gaussian processes have plethora of applications, ranging from the modeling of environmental processes in *geostatistics* (e.g., [9,11]) to supervised regression and classification in *machine learning* [5,8,24], and the simulation of complex computer models [18]. The versatility of the correlation structure of Gaussian processes provides a tractable and powerful tool for the modeling of large and highly dependent environmental variables. As a common approach in the field of spatial statistics, the covariance functions of Gaussian processes are assumed to belong to a parametric family. High precision estimates of the covariance parameters are pivotal for interpolating Gaussian processes which is the ultimate goal in many geostatistical problems [9,19].

In the last two decades, there has been extensive research regarding the statistical and computational facets of the estimation of the Gaussian processes' covariance parameters. Maximum likelihood estimation (MLE) was the earliest favored algorithm in the geostatistics community, e.g., Mardia et al. [14] and Ying [25]. However, solving systems of linear equations is inevitable to evaluate the Gaussian likelihood. Notwithstanding the recent advances toward scalable and efficient solution of the system of linear equation (e.g., iterative *Krylov* subspace method or block preconditioned conjugate gradient algorithm [16]) which moderately reduces the computational and memory costs of the direct evaluation of the

☆ This research is partially supported by NSF grant ACI-1047871. Additionally, CS is partially supported by NSF grants 1422157, 1217880, and 0953135, and LN by NSF CAREER award DMS-1351362, NSF CNS-1409303, and NSF CCF-1115769.

* Corresponding author.
E-mail addresses: hksh@umich.edu (H. Keshavarz), clayscot@umich.edu (C. Scott), xuanlong@umich.edu (X. Nguyen).

http://dx.doi.org/10.1016/j.jmva.2016.06.003

precision matrix, obtaining the MLE of unknown covariance parameters using such linear systems solvers is still a strenuous task, especially for a generic Gaussian spatial process observed at numerous and possibly irregularly spaced locations. Approximating the likelihood function by tapering the covariance matrix is another class of algorithms aiming to reduce the numerical burden of MLE (see Kaufman et al. [12]). Due to the sparsity of the tapered covariance matrix, its inverse can be computed in a faster and more stable way. Recent studies [10,12,22] demonstrate the consistency and asymptotic normality of this algorithm under some mild conditions on the taper function.

Because of the obstacles of solving system of linear equations for massive data, which is necessary for tapered and exact MLE, it is of great interest to develop estimation techniques without requiring such extensive computations. Such class of algorithms, which will be referred to as *inversion-free*, are based upon optimizing a loss function whose form (and its derivatives) is independent of the precision matrix of data. The first attempt toward such a goal has been done by Anitescu, Chen and Stein [2] in 2014 (referred here to as ACS). Their proposed procedure is faster and more stable than likelihood based algorithms. In [2], the covariance parameters are estimated by computing the global maximizer of a non-concave program. Simulation studies verify the efficiency of ACS's approach in the case that the covariance matrix has a bounded condition number. The main purpose of this paper is to appraise the asymptotic properties of the ACS's algorithm such as consistency, minimax optimality and asymptotic normality. The developed theory in this paper shows that ACS's algorithm has the same asymptotic rate of convergence as the MLE. In practice, the solution of ACS's optimization problem may also serve as the starting point (initial guess) of a likelihood maximization procedure.

In geostatistics, there are two common asymptotic regimes: *increasing-domain* and *fixed-domain*, the latter sometimes referred to as infill asymptotics (see [19], Section 3.3 or [9], p. 480). In the former setting, the minimum distance among the sampling points is bounded away from zero and more samples are collected by increasing the diameter of the spatial domain. In the latter regime, the data are sampled in a fixed and bounded domain, and the observations get denser as the sample size n increases.

Zhang [26] showed that not all the covariance parameters are consistently estimable in the fixed domain regime. Strictly speaking, there is no asymptotically consistent algorithm for estimating the *non-micro ergodic* covariance parameters, which do not asymptotically affect the interpolation mean square error (see [19] for a precise definition). On the other hand, it is known in the literature that subject to some mild regularity conditions, maximizing the likelihood provides a strongly consistent and asymptotically normal estimate for all the covariance parameters in the increasing domain setting [4,14].

Increasing domain asymptotic analysis of covariance estimation has two significant benefits. First, unlike the infill asymptotic setting, the geometry of the spatial sampling has a crucial impact on the asymptotic distribution of the parameter estimate. Thus, this regime is a natural asymptotic framework for assessing the role of irregularity of spatial sampling on the covariance parameter estimation [4]. This claim can be verified by a deeper look at the asymptotic distribution of the microergodic parameter estimates in the fixed domain (see e.g., [19,25] for MLE and [10,12,22] for tapered MLE). Another significant characteristic of increasing domain regime is that the covariance matrix has a universally bounded condition number as n grows under some mild regularity conditions. This feature of the covariance matrix plays a major role in our asymptotic analysis. Although in many geostatistical applications in a fixed bounded domain the condition number of the covariance matrix increases at least linearly with respect to n, preconditioning filters is commonly used to uniformly control the condition number independent of n [7,20]. Therefore, we believe that our developed increasing domain asymptotics can be useful for the fixed domain analysis of preconditioned inversion-free algorithms.

Outline of main results. This paper studies the increasing domain asymptotic behavior of ACS's estimation algorithm introduced in [2]. Specifically, suppose that $\mathfrak{G}$ is a zero mean stationary Gaussian process in $\mathbb{R}^d$ with covariance function $\operatorname{cov}\left(\mathfrak{G}(s), \mathfrak{G}(s')\right) = R\left(s - s', \eta\right)$ in which $\eta \in \Omega$ denotes the vector of unknown covariance parameters. One realization of $\mathfrak{G}$ has been observed on a *d-dimensional perturbed regular lattice* of $n = N^d$ points, which will be formally defined in Section 2. The specific contributions of this work are given as follows:

(a) Assuming the polynomial decay of $R(s, \eta)$ and its gradient (with respect to η) in terms of the Euclidean norm of s, and under some mild identifiability condition on R, we prove that the global maximizer of ACS's method consistently estimates η. Furthermore, the estimation error is of order $\sqrt{n^{-1}\ln n}$ which is shown to be minimax optimal up to some $\sqrt{\ln n}$ term.
(b) As the proposed loss function in Anitescu et al. [2] is not jointly concave in η, finding its global maximizer is challenging. For a large enough sample size and under an additional condition regarding the polynomial decay of the second derivative of $R(s, \eta)$ with respect to η, we show that any *stationary point* of this non-concave program is concentrated around the true η with radius of order $\sqrt{n^{-1}\ln n}$.
(c) The asymptotic normality of the stationary points of the aforementioned algorithm will be substantiated under some mild restriction on the third derivative of R with respect to η.

Furthermore, the Appendix contains several easy-to-reference nonasymptotic results on the global and local behavior of the quadratic forms of Gaussian processes which may come in hand for the analysis of certain problems in statistics and machine learning.

Plan of the paper. In Section 2, we formulate ACS's inversion-free estimation method and precisely introduce the geometry of the sampling points. Section 3 expresses the necessary assumptions and studies the asymptotic properties of the estimation algorithm. Section 3.1 presents the convergence rate of the global and local maximizers of the optimization problem