Linear Algebra and its Applications 466 (2015) 512–526

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

# On the condition number anomaly of Gaussian correlation matrices

R. Zimmermann

*Institute of Computational Mathematics, TU Braunschweig, 38100, Germany*

ARTICLE INFO

*Article history:*
Received 25 October 2013
Accepted 22 October 2014
Available online 8 November 2014
Submitted by Z. Strakoš

*MSC:*
15B99

*Keywords:*
Gaussian process
Condition number
Design and analysis of computer experiments
Euclidean distance matrix
Correlation matrix
Kriging
Radial basis functions
Data assimilation
Spatial linear model

ABSTRACT

Spatial correlation matrices appear in a large variety of applications. For example, they are an essential component of spatial Gaussian processes, also known as spatial linear models or Kriging estimators, which are powerful and well-established tools for a multitude of engineering applications such as the design and analysis of computer experiments, geostatistical problems and meteorological tasks.

In radial basis function interpolation, Gaussian correlation matrices arise frequently as interpolation matrices from the Gaussian radial kernel function. In the field of data assimilation in numerical weather prediction, such matrices arise as background error covariances.

Over the past thirty years, it was observed by several authors from several fields that the Gaussian correlation model is exceptionally prone to suffer from ill-conditioning, but a quantitative theoretical explanation for this anomaly was lacking. In this paper, a proof for the special position of the Gaussian correlation matrix is given. The theoretical findings are illustrated by numerical experiment.



*E-mail address:* ralf.zimmermann@tu-bs.de.

http://dx.doi.org/10.1016/j.laa.2014.10.038

## 1. Introduction

The spatial linear model in the classical setting (see [1–3]) is defined as follows. Consider a real-valued covariance stationary Gaussian process in $d \in \mathbb{N}$ spatial dimensions

$$y(x) = f(x)^T\beta + \epsilon(x); \quad \epsilon(x) \sim N(0, \sigma^2), \quad x \in \mathbb{R}^d,$$

where $f(x) = (f_1(x), ..., f_p(x))^T$ is the regressor vector and $\beta = (\beta_1, ..., \beta_p)$ is the vector of regression coefficients. Let $\{x^1, \ldots, x^n\} \subset \mathbb{R}^d$ be a set of mutually distinct sample points. Suppose that the stationary covariance structure is modeled via a positive definite covariance function $\operatorname{cov}(y(x^i), y(x^j)) = \sigma^2\rho(\theta, (x^i - x^j))$, by convention parametrized such that for $v \in \mathbb{R}^d \setminus \{0\}$,

$$\rho(\theta, v) \to \begin{cases} 1, & \text{for } \|\theta\| \to 0 \\ 0, & \text{for } \|\theta\| \to \infty, \end{cases}$$

with coordinate-wise range-parameters $\theta = (\theta_1, \ldots, \theta_d)$, also referred to as the model's hyper-parameters. The reciprocal values $1/\theta_k$ are called the correlation lengths. For a vector $Y = (y(x^1), \ldots, y(x^n))^T \in \mathbb{R}^n$ of $n$ observations, let $R := (\rho(\theta, (x^i - x^j)))_{ij} \in \mathbb{R}^{n\times n}$ be the corresponding correlation matrix and $r(x) := (\rho(\theta, (x^i - x)))_i \in \mathbb{R}^n$. The best linear unbiased predictor is

$$\hat{y}(x) = f(x)^T\beta + r(x)^T R^{-1}(Y - F\beta),$$

where the matrix $F$ features the regressor vectors $f(x_i)^T$ $(i = 1, \ldots, n)$ as rows and $\beta = (F^T R^{-1} F)^{-1} F^T R^{-1} Y$ is the generalized least-squares solution to the regression problem. Up to an additive constant, the profile log likelihood is

$$L(Y, \theta) = -\frac{1}{2}\left(n \log\left(\sigma^2(\theta)\right) + \log\left(\det\left(R(\theta)\right)\right)\right), \tag{1}$$

where $\sigma^2(\theta) = 1/n(Y - F\beta)^T R^{-1}(\theta)(Y - F\beta)$. Both the predictor and the likelihood function require to compute the inverse of the correlation matrix, which may be replaced by solving linear systems of the form $Rv = b$.

Closely related to Kriging is radial basis function (RBF) interpolation [4,5]. Here, Gaussian correlation matrices arise frequently as interpolation matrices, also referred to as distance matrices, and, as in Kriging, it is required to solve linear systems featuring such matrices as operators.

In the field of data assimilation for numerical weather prediction [6,7], the inverses of spatial correlation matrices appear in the area-defining optimization problem, see e.g. [7, Eq. (1)].

As a consequence, the accuracy and numerical robustness in all of the aforementioned applications depend crucially on the correlation matrices' condition numbers, a fact that has been acknowledged by several authors and is still subject to ongoing investigations.