Convex optimization and feasible circulant matrix embeddings
in synthesis of stationary Gaussian fields[*][‡][§]

Hannes Helgason[1], Stefanos Kechagias[2], and Vladas Pipiras[2]

[1]University of Iceland, Reykjavik
[2]University of North Carolina, Chapel Hill

October 9, 2014

Abstract

Circulant matrix embedding is one of the most popular and efficient methods for the exact generation of Gaussian stationary univariate series. Although the idea of circulant matrix embedding has also been used for the generation of Gaussian stationary random fields, there are many practical covariance structures of random fields where classical embedding methods break down. In this work, we propose a novel methodology which adaptively constructs feasible circulant embeddings based on convex optimization with an objective function measuring the distance of the covariance embedding to the targeted covariance structure over the domain of interest. The optimal value of the objective function will be zero if and only if there exists a feasible embedding for the a priori chosen embedding size. In cases where the optimum is nonzero, the resulting feasible covariance embedding will be the optimal approximation to the targeted covariance.

1   Introduction

In this work, we are interested in the synthesis of stationary Gaussian two-dimensional random fields. One of the most popular methods for the exact generation of such fields is based on *circulant matrix embedding* (CME). Introduced by Davis and Harte (1987) for simulating stationary Gaussian univariate processes, the method was later extended to random fields

---

by Dietrich and Newsam (1993), Wood and Chan (1994), Chan and Wood (1997), Gneiting et al. (2006), Stein (2012). See also Stein (2002) concerning the case of fractional Brownian surfaces; Chan and Wood (1999), Helgason et al. (2011) for extensions to the multivariate (vector) context; Percival (2006) for the case of complex-valued Gaussian processes; and Craigmile (2003) for some theoretical contributions.

The idea of the method is to use a suitable periodization to embed a covariance matrix of interest into a larger circulant matrix whose eigenvalues can be computed efficiently using the fast Fourier transform (FFT). If all the eigenvalues are nonnegative, a stationary Gaussian random field can be constructed with the embedding circulant matrix as its covariance matrix. Its suitable subfield is then the desired stationary Gaussian field. While in theory all eigenvalues can be shown to be nonnegative in the limit of the increasing sample size, some of the eigenvalues often remain negative for computationally feasible large sample sizes and many covariance structures of practical interest.

To have all the eigenvalues nonnegative, two approaches have been suggested in the literature, namely, the *cutoff circulant embedding* or CCE, for short (Stein (2002), Gneiting et al. (2006)) and the *smoothing windows circulant embedding* or SWCE, for short (Helgason et al. (2014)). In the CCE method, the initial covariance is extended suitably in a parametric fashion, based on the model at hand, to a larger domain, leading to a covariance embedding with nonnegative eigenvalues. While such extensions have been found for several examples of stationary fields, their construction is often nontrivial.

The SWCE approach, on the other hand, does not depend on the model at hand and is easier to implement by suitably modifying the standard embedding. A possible reason that some eigenvalues are negative is that the covariance embedding is not "smooth" at the boundary of periodization. In the SWCE method, the covariance is extended over a transition region where it is smoothed at the boundary, using a smoothing kernel. Helgason et al. (2014) propose two types of smoothing: overlapping and nonoverlapping windows. Numerical studies show that both variants of SWCE work well for a range of covariance structures. Moreover, when using overlapping windows, the SWCE method greatly outperforms CCE in the sense of efficiency considered in Helgason et al. (2014). However, for some covariance structures (for example, as those in Section 4 below), a large transition region is needed for the SWCE method to work.

We propose and study here a new method to simulate stationary Gaussian fields using circulant matrix embedding, called *optimal circulant embedding* or OCE, for short. We find the method interesting and promising for several reasons. First, the method proposes a novel approach based on quadratic constrained optimization. This is quite fitting given the growing integration of optimization tools in statistics. Second, as seen below, several key components of the optimization procedure can be implemented more efficiently using FFT, making it natural to combine the circulant matrix embedding and optimization approaches. Third, we show numerically that the OCE method works better than the SWCE method (and hence the CCE method by the discussion above) for several covariance classes, especially when the transition region is small.

More specifically, the OCE method consists of formulating a quadratic programming problem with linear inequality constraints, to find a *feasible* covariance embedding, meaning

that its eigenvalues are nonnegative. The exact optimization problem has the form

$$\min_{\widetilde{r}(n_1,n_2)} \sum_{n_1,n_2=-N+1}^{N-1} \left(r(n_1,n_2) - \widetilde{r}(n_1,n_2)\right)^2, \tag{1.1}$$
$$\text{subject to} \quad \lambda_{k_1,k_2}(\widetilde{r}) \geq 0, \quad -\widetilde{N}+1 \leq k_1, k_2 \leq \widetilde{N}-1,$$

where $\{r(n_1,n_2), \ -N+1 \leq n_1, n_2 \leq N-1\}$ is a given covariance function of a random field $X_{(n_1,n_2)}$ on a square grid $\{(n_1,n_2) : \ 0 \leq n_1, n_2 \leq N-1\}$, $\{\widetilde{r}(n_1,n_2), \ -\widetilde{N}+1 \leq n_1, n_2 \leq \widetilde{N}-1\}$ is the corresponding covariance embedding on a larger grid $\{(n_1,n_2), \ -\widetilde{N}+1 \leq n_1, n_2 \leq \widetilde{N}-1\}$ with $\widetilde{N} \geq N$, and $\lambda_{k_1,k_2}(\widetilde{r})$ are the eigenvalues of the covariance embedding $\widetilde{r}$. The difference between the larger and smaller square grids is the transition region, referred to above. When the minimum value of the objective function is equal to zero, then the method is exact, namely, the synthesized field will have the targeted covariance. If for some choice of $\widetilde{N}$ this is not true, then the solution of (1.1) can be thought as the best approximation to the targeted covariance, that has no negative eigenvalues. The focus of this paper though is on the situations where the objective function is zero (up to some tolerance error) and the simulation is exact.

To solve (1.1) numerically, we use an interior-point algorithm called the *primal log-barrier method* (see Section 3 for more details). The basic idea of this algorithm is to solve a sequence of unconstrained problems of the form

$$\min_{\widetilde{r}(n_1,n_2)} \quad F_t(\widetilde{r}) := \sum_{n_1,n_2=-N+1}^{N-1} \left(r(n_1,n_2) - \widetilde{r}(n_1,n_2)\right)^2 - \frac{1}{t} \sum_{k_1,k_2=-\widetilde{N}+1}^{\widetilde{N}-1} \log(\lambda_{k_1,k_2}(\widetilde{r})), \tag{1.2}$$

for positive increasing values of t. The log term in (1.2) serves as a "barrier" not allowing the search algorithm starting from a strictly feasible point $\widetilde{r}$ (with all eigenvalues $\lambda_{k_1,k_2}(\widetilde{r})$ positive) to move into the region of non-feasible points $\widetilde{r}$ (with some eigenvalues $\lambda_{k_1,k_2}(\widetilde{r})$ zero or negative). On the other hand, a large t ensures that the minimized function $F_t(\widetilde{r})$ is close to that of interest in (1.1). One can show that the solutions $\widetilde{r}_t(n_1,n_2)$ of (1.2) approach a solution of (1.1) as t increases. Eliminating inequality constraints of (1.1) by inserting them in the objective function in (1.2) is a frequently used practice, since unconstrained problems are far easier to handle.

To solve the problem (1.2) (for a fixed t), we approximate F_t in (1.2) by a quadratic function $\widehat{F}_t$ (around a strictly feasible initial point), and calculate its minimizer $\widetilde{r}_{t,a}(n_1,n_2)$. This reduces the problem to solving a linear system of the form

$$Hv = b, \tag{1.3}$$

where H is the Hessian of F_t in (1.2) and v is the covariance embedding $\widetilde{r}$ indexed as a vector. The symmetry and the positive definiteness of the coefficient matrix H allows solving the systems (1.3) using a popular iterative algorithm called the *conjugate gradient* method. As we show below, key steps of the conjugate gradient method applied to our problem can be carried out more efficiently using FFT.

Though the various components of the optimization procedure discussed above might appear straightforward, there are in fact a number of technical issues that need to be addressed. For example, as indicated above, we need to show how FFT is used in solving (1.3).

Among other issues arising in the procedure are the role and choice of the updating rule for t in (1.2), the effect that the approximation of F_t by $\widehat{F}_t$ has on the convergence of the approximate solutions $\widetilde{r}_{t,a}(n_1, n_2)$ to the solution of (1.1), and so on. No less important is to see how the suggested approach performs on covariance structures of interest and how it compares to other approaches. All these issues are addressed in the paper.

The rest of the paper is organized as follows. In Section 2, we briefly review the existing circulant embedding schemes. Optimal circulant embedding is presented in Section 3. The performance of the method for several covariance structures is studied in Section 4. Technical proofs and figures are moved to Appendices A and B.

2 Available circulant matrix embeddings

Let $X = \{X_n, n = (n_1, n_2) \in \mathbb{Z}^2\}$ be a zero mean, stationary Gaussian two-dimensional random field. The autocovariance function of X is defined as

$$r(n) = r(n_1, n_2) = \mathbb{E}X_0 X_n, \tag{2.1}$$

and satisfies $r(n) = r(-n)$, $n \in \mathbb{Z}^2$. We are interested here in the synthesis of the two dimensional field X on the square grid

$$G(N) = \{n = (n_1, n_2) \in \mathbb{Z}^2 : 0 \leq n_1, n_2 \leq N - 1\}, \tag{2.2}$$

though the methods described below can be extended easily to rectangular grids and likely to higher dimensions. For reference and comparison, the rest of Section 2 contains a short description of the existing circulant matrix embedding methods. We recall the standard CME (Circulant Matrix Embedding) method in Section 2.1. The CCE (Cutoff Circulant Embedding) and SWCE (Smoothing Windows Circulant Embedding) methods are discussed briefly in Sections 2.2 and 2.3, respectively.

2.1 Standard embedding

Set $M = 2N - 1$ to be the (side) size of a larger embedding grid $G(M)$, where an embedding field will be generated. Let also $\widetilde{r}(n)$, $n \in G(M)$, be the covariance embedding defined through the symmetry relation

$$\widetilde{r}(n) = r(\xi_N(n)), \quad n \in G(M), \tag{2.3}$$

where $\xi_L(u) = (\xi_{1,L}(u), \xi_{2,L}(u))$ is defined by

$$\xi_{1,L}(u) = \xi_{2,L}(u) = \begin{cases} u, & \text{if } 0 \leq u \leq L - 1, \\ u - M, & \text{if } L \leq u \leq M - 1. \end{cases} \tag{2.4}$$

(The subscript L in (2.4) differs from N in (2.3) since the function (2.4) will also be used for other indices than N.) Extend $\widetilde{r}$ periodically in both dimensions by period M. Note that $\widetilde{r}$ can also be defined as the function that is M–periodic in both dimensions, and satisfies

$$\widetilde{r}(n) = r(n), \quad n \in \widetilde{G}(N), \tag{2.5}$$

4

where the grid $\widetilde{G}(N)$ is defined as

$$\widetilde{G}(N) = \{n = (n_1, n_2) \in \mathbb{Z}^2 : -(N-1) \leq n_1, n_2 \leq N-1\} \tag{2.6}$$

(see also Remark 2.1 below).

Next, consider a circular convolution operator $\widetilde{\Sigma}$, whose action on a vector $u(n), n \in G(M)$, is defined by

$$\widetilde{\Sigma}u(n) = \sum_{m \in G(M)} \widetilde{r}(m-n)u(m), \quad n \in G(M). \tag{2.7}$$

The two-dimensional DFT basis $\{e^{-i2\pi k \cdot (n/M)}, \ n \in G(M)\}$, $k \in G(M)$, where $n/M = (n_1/M, n_2/M)$ and $\cdot$ is the usual inner product, diagonalizes the operator $\widetilde{\Sigma}$. Hence, the eigenvalues of $\widetilde{\Sigma}$ are given by

$$\lambda_k = \sum_{n \in G(M)} \widetilde{r}(n)e^{-i2\pi k \cdot (n/M)}, \quad k \in G(M), \tag{2.8}$$

and can be computed efficiently using the two-dimensional FFT. Assuming that

$$\lambda_k \geq 0, \quad k \in G(M), \tag{2.9}$$

consider the complex-valued random variables

$$\widetilde{X}_n = M^{-1} \sum_{k \in G(M)} \lambda_k^{1/2}(Z_k^0 + iZ_k^1)e^{-i2\pi n \cdot (k/M)}, \quad n \in G(M), \tag{2.10}$$

where Z_k^0, Z_k^1, $k \in G(M)$, are independent standard Gaussian random variables. One can show that $\{\Re(\widetilde{X}_n), n \in G(M)\}$ and $\{\Im(\widetilde{X}_n), n \in G(M)\}$, that is, the real and imaginary parts of $\widetilde{X}_n$, are independent random fields with the covariance structure

$$\mathbb{E}\Re(\widetilde{X}_n)\Re(\widetilde{X}_{n+h}) = \mathbb{E}\Im(\widetilde{X}_n)\Im(\widetilde{X}_{n+h}) = \widetilde{r}(h), \quad n, n+h \in G(M). \tag{2.11}$$

By using (2.5) and (2.11), a subfield $X_n = \Re(\widetilde{X}_n)$ or $X_n = \Im(\widetilde{X}_n)$, $n \in G(N)$, is a Gaussian random field with the desired covariance structure r.

The construction of the variables $\widetilde{X}_n$ is possible only under the condition (2.9). This condition, however, often does not hold for some $k \in G(M)$, and hence the above standard circulant matrix embedding (CME) method fails. A common way to make it work is to increase N to some $\widetilde{N}$. (As noted in Section 1, the condition (2.9) holds in the limit $N \to \infty$, under mild assumptions.) It is convenient to think of $\widetilde{G}(\widetilde{N}) \backslash \widetilde{G}(N)$ (or its periodization in $G(M)$ with $M = 2\widetilde{N} - 1$) as a *transition region*. Increasing N to $\widetilde{N}$ can thus be thought as extending the covariance function $r(n)$ over the transition region. Since increasing N to any computationally feasible $\widetilde{N}$ often does not ensure (2.9), it is natural to consider other ways to extend $r(n)$ over the transition region $\widetilde{G}(\widetilde{N}) \backslash \widetilde{G}(N)$.

Remark 2.1 The reader may find the use of both domains $\widetilde{G}(N)$ and $G(M)$ excessive, and wonder why not use only one of them. In particular, for example in view of (2.3) and (2.5), the domain $\widetilde{G}(N)$ may seem to be preferred for simplicity. There are several reasons why we, and others, work on $G(M)$. First, it is natural to think of extending $G(N)$ to $G(M)$, where a subfield is ultimately selected. (In one dimension, a longer series is generated at times $0, 1, \ldots, M-1$, and a shorter series is selected of size N.) Second, the indexing by $G(M)$ is more common in scientific software, e.g. MATLAB. Third and more importantly in our context, we will refer below to the discontinuities of the covariance embedding at the boundary of periodization (see the beginning of Section 2.3). The discontinuities are easier to visualize on $G(M)$, not $\widetilde{G}(N)$.

2.2 Cutoff embedding

The CCE (Cutoff Circulant Embedding) method considers two extension schemes over the transition region, and is used in models with isotropic covariances $r(n) = \psi(||n||_2)$, where ψ is a function and $||\cdot||_2$ denotes the usual Euclidean distance. In one CCE extension (Gneiting et al. (2006)), $\widetilde{r}$ is defined on $\widetilde{G}(\widetilde{N})$ as

$$\widetilde{r}(n) = \begin{cases} r(n), & \text{if } 0 \le ||n||_2 \le \sqrt{2}N, \\ b_1(a_1 - ||n||_2^{1/2}), & \text{if } \sqrt{2}N \le ||n||_2 \le \widetilde{N} - 1, \end{cases} \tag{2.12}$$

and then periodically extended in both dimensions. The constants a_1, b_1 and $\widetilde{N}$ are chosen as

$$a_1 = (\sqrt{2}N)^{1/2} - (\sqrt{2}N)^{-1/2} \frac{1}{2} \frac{\psi(\sqrt{2}N)}{\psi'\sqrt{2}N}, \quad b_1 = -2(\sqrt{2}N)^{1/2}\psi'(\sqrt{2}N), \quad \widetilde{N} = [a_1^2], \tag{2.13}$$

where $[x]$ denotes the integer part of x. The choices (2.13) ensure that $\widetilde{r}(n)$ is once continuously differentiable at $t = \sqrt{2}N$, while the choice of $\sqrt{2}N$ in (2.12) is to have $\widetilde{r}(n) = r(n)$ for $n \in \widetilde{G}(N)$.

In another CCE extension, the embedding $\widetilde{r}$ is defined as

$$\widetilde{r}(n) = \begin{cases} r(n), & \text{if } 0 \le ||n||_2 \le \sqrt{2}N, \\ b_2(a_2 - ||n||_2)^2, & \text{if } \sqrt{2}N \le ||n||_2 \le \widetilde{N} - 1, \end{cases} \tag{2.14}$$

where the constants a_2, b_2 and $\widetilde{N}$ satisfy

$$a_1 = \sqrt{2}N - 2\frac{\psi(\sqrt{2}N)}{\psi'(\sqrt{2}N)}, \qquad b_1 = \frac{(\psi'(\sqrt{2}N))^2}{4\psi(\sqrt{2}N)}, \qquad \widetilde{N} = [a_2^2]. \tag{2.15}$$

After extending r into $\widetilde{r}$ as in (2.12) or (2.14), and periodization, the rest of CCE is the same as the algorithm in (2.7)–(2.10) with $M = 2\widetilde{N} - 1$. As shown in in Theorems 1 and 2 of Gneiting et al. (2006), under the extension schemes (2.12) and (2.14), the condition (2.9) always holds for several classes of covariance models.

2.3 Smoothing windows

The SWCE (Smoothing Windows Circulant Embedding) method is another way of extending the covariance function over the transition region. In the standard CME, the cross sections $\{(n_1, n_2) : n_1 = N - 1,\ 0 \leq n_2 \leq M - 1\}$ and $\{(n_1, n_2) : n_2 = N - 1,\ 0 \leq n_1 \leq M - 1\}$ are the boundary of periodization, when extending r periodically to $\widetilde{r}$ from $\widetilde{G}(N)$ to $\mathbb{Z}^2$ (and, in particular, to $\widetilde{r}$ on $G(M)$). Along this boundary, the covariance embedding $\widetilde{r}$ is often not smooth (see, for example, the top left plots in Figures 1–2). The idea of the SWCE method is to smooth the discontinuities at the boundary of periodization in the standard CME, since the discontinuities can be thought to be the source of negative eigenvalues (of FFT). Helgason et al. (2014) showed that the SWCE method outperforms the CCE and standard CME for a range of covariance functions. In this section, we review briefly the two variants of the SWCE method, called the *overlapping* and *nonoverlapping windows*. For more details, see Helgason et al. (2014).

We begin with the definition of a smoothing window. Let $0 < P < Q$. A smoothing window is defined as

$$w(x) = \begin{cases} 1, & \text{if } x \in \widetilde{G}(P), \\ \rho(x), & \text{if } x \in \widetilde{G}(Q) \setminus \widetilde{G}(P), \\ 0, & \text{if } x \in \mathbb{R}^2 \setminus \widetilde{G}(Q), \end{cases} \tag{2.16}$$

where $\rho(x), x \in \mathbb{R}^2$, is a real-valued function that decays smoothly from 1 to 0 when moving from the boundary of $\widetilde{G}(P)$ toward that of $\widetilde{G}(Q)$.

In the *nonoverlapping* SWCE, one considers a transition region of side length $\widetilde{N} - N$ and an embedding size $M = 2\widetilde{N} - 1$. The covariance function is then extended through the transition region as

$$\widetilde{r}(n) = r(n)w(n), \quad n \in \widetilde{G}(\widetilde{N}), \tag{2.17}$$

where $w(n)$ is the smoothing window in (2.16) with

$$P = N, \quad \text{and} \quad Q = \widetilde{N}. \tag{2.18}$$

Since $w(n) = 1$ for $n \in \widetilde{G}(N)$, we have $\widetilde{r}(n) = r(n)$ for $n \in \widetilde{G}(N)$. The rest of the algorithm (after periodizing $\widetilde{r}$) is the same as CME in Section 2.1. The basic idea behind (2.17) is that $w(n)$ smooths $\widetilde{r}(n)$ at the boundary of periodization by forcing it to zero.

Note that (2.17) is the embedding on $\widetilde{G}(\widetilde{N})$, which is then periodized in both dimensions. The covariance embedding $\widetilde{r}(n)$ can equivalently be defined on $G(M)$ and then periodized by setting

$$\begin{aligned} \widetilde{r}(n) =&\ r(n_1, n_2)w_1(n_1, n_2) + r(n_1, n_2 - M)w_2(n_1, n_2) \\ &+ r(n_1 - M, n_2)w_3(n_1 - M, n_2 - M) + r(n_1, n_2)w_4(n_1, n_2), \quad n \in G(M) \end{aligned} \tag{2.19}$$

where

$$\begin{array}{ll} w_1(n_1, n_2) = w(n_1, n_2), & w_2(n_1, n_2) = w(n_1, n_2 - M), \\ w_3(n_1, n_2) = w(n_1 - M, n_2 - M), & w_4(n_1, n_2) = w(n_1 - M, n_2). \end{array} \tag{2.20}$$

To see the equivalence, note that only one of the $w_i(n)$'s can be different from zero. See Section 3.2 in Helgason et al. (2014) for more details.

In the *overlapping* SWCE, the covariance embedding $\widetilde{r}(n)$ is smoothed without forcing it to be zero as in the nonoverlapping SWCE. The basic idea is to use the embedding given by (2.19)–(2.20) but choosing a smoothing window (2.16) with

$$P = N, \quad \text{and} \quad Q = 2\widetilde{N} - N. \tag{2.21}$$

The fact that $Q > \widetilde{N}$ in (2.21) ensures the overlap of $w_i(n)$'s. One can show that under (2.21), $\widetilde{r}(n) = r(n)$ for $n \in \widetilde{G}(N)$. The rest of the algorithm (after periodizing $\widetilde{r}$) is the same as CME in Section 2.1 using the embedding (2.19)–(2.20) with (2.21). The overlapping SWCE was found to outperform the nonoverlapping SWCE in Helgason et al. (2014).

3   Optimal circulant embedding

We propose here a new type of circulant matrix embedding method to generate a zero mean, stationary Gaussian random field X on the grid $G(N)$. Take $\widetilde{N} > N$ and set $M = 2\widetilde{N} - 1$. We shall also assume that the covariance embedding $\widetilde{r} = \{\widetilde{r}(n), \ n \in G(M)\}$ has the property $\widetilde{r}(n) = \widetilde{r}(-n)$, and satisfies the symmetry condition

$$\widetilde{r}(n) = \widetilde{r}(\xi_{\widetilde{N}}(n)), \quad n \in G(M), \tag{3.1}$$

where the function $\xi_{\widetilde{N}}(u)$ is given in (2.4). The symmetry condition (3.1) ensures, in particular, that the eigenvalues of the covariance embedding $\widetilde{r}$ are real-valued (see also the discussion following Lemma 3.1 below).

3.1   Formulation of the constrained optimization problem

The basic idea of the OCE (Optimal Circulant Embedding) method is to obtain the covariance embedding $\widetilde{r} = \{\widetilde{r}(n), \ n \in G(M)\}$, or by periodization $\widetilde{r} = \{\widetilde{r}(n), \ n \in \widetilde{G}(\widetilde{N})\}$, by solving the following optimization problem:

$$\min_{\widetilde{r}} \quad f^*(\widetilde{r}) = \sum_{n \in \widetilde{G}(N)} \left(r(n) - \widetilde{r}(n) \right)^2,$$

$$\text{subject to} \quad g_k(\widetilde{r}) \geq 0, \quad k \in \widetilde{G}(\widetilde{N}), \tag{3.2}$$

where $r = \{r(n), n \in \widetilde{G}(N)\}$ is a given covariance structure and $g_k(\widetilde{r})$ are the eigenvalues of the covariance embedding $\widetilde{r}$, given by (after replacing $G(M)$ by $\widetilde{G}(\widetilde{N})$ in (2.8) for convenience and using $\widetilde{G}(\widetilde{N})$ for indexing)

$$g_k(\widetilde{r}) = \sum_{n \in \widetilde{G}(\widetilde{N})} \widetilde{r}(n) e^{-i2\pi k \cdot (n/M)}, \quad k \in \widetilde{G}(\widetilde{N}). \tag{3.3}$$

Note that the sum in the first relation in (3.2) is taken over a smaller grid $\widetilde{G}(N)$ than $\widetilde{G}(\widetilde{N})$ of the sum in (3.3). Minimization of f forces the covariance embedding $\widetilde{r}$ to be as close as possible to the targeted covariance r over the domain of interest $\widetilde{G}(N)$. Hence, when the minimum value of f is zero, the solution of the problem can be used for the exact synthesis

of the desired field. The inequality constraints in (3.2) rule out candidate (non-feasible) minimizers that lead to covariance embeddings with negative eigenvalues.

To write the problem (3.2) on the domain $G(M)$, introduce the weights

$$\beta(n) = \begin{cases} 1, & n \in \widetilde{G}(N), \\ 0, & n \in \widetilde{G}(\widetilde{N}) \backslash \widetilde{G}(N), \end{cases} \tag{3.4}$$

on $\widetilde{G}(\widetilde{N})$ and extend them periodically in both dimensions. With r extended periodically in the same way, the problem (3.2) can be written as

$$\begin{aligned} \min_{\widetilde{r}} \quad & f^*(\widetilde{r}) = \sum_{n \in G(M)} \beta(n) \big(r(n) - \widetilde{r}(n) \big)^2, \\ \text{subject to} \quad & g_k(\widetilde{r}) \geq 0, \quad k \in G(M). \end{aligned} \tag{3.5}$$

The optimization problem (3.5) is a quadratic programming problem with linear inequality constraints. In the optimization literature, this problem is treated by regarding the underlying variable as a vector. In the case considered here, the underlying variable $\widetilde{r}$ is indexed by the square grid $G(M)$. Indexing is just a question of perspective and we will continue using parameters indexed by grids (though also sometimes switch to vectors, as for example, in one of the proofs).

An issue requiring immediate attention, however, is the symmetry of the covariance embedding $\widetilde{r}$, expressed through the relation (3.1). We need to keep track only of the non-symmetric half of the covariance, which we choose to be over the smaller rectangular grid

$$G_+(M) = \{n \in \mathbb{Z}^2,\ 0 \leq n_1 \leq \widetilde{N} - 1,\ 0 \leq n_2 \leq M - 1\}. \tag{3.6}$$

Note that the symmetry relation (3.1) implies that, for $(n_1, n_2) \in G(M) \backslash G_+(M)$, $\widetilde{r}(n_1, n_2) = \widetilde{r}(M - n_1, M - n_2)$, so that the values of $\widetilde{r}(n)$ for $n \in G(M) \backslash G_+(M)$ are indeed determined from those on $G_+(M)$.

When working on the smaller rectangle $G_+(M)$, we need to have a representation of the eigenvalues $g_k(\widetilde{r})$ where the sum $\sum_{n \in G(M)}$ is replaced by $\sum_{n \in G_+(M)}$, that is, a representation of the form

$$g_k(\widetilde{r}) = \sum_{n \in G_+(M)} c_k(n) \widetilde{r}(n) =: [A\widetilde{r}](k), \tag{3.7}$$

where $c_k(n)$ are suitable weights and A is viewed as a linear operator acting on a field defined on $G_+(M)$. The next lemma provides an expression for the weights $c_k(n)$. The proof of the lemma can be found in Appendix A.

Lemma 3.1 *Suppose (3.1) holds, and the eigenvalues $g_k(\widetilde{r})$ are given by (2.8). Then, the relation (3.7) holds with the weights $c_k(n)$, $k \in G(M)$, $n \in G_+(M)$, given by*

$$c_k(n) = c_{(k_1, k_2)}(n_1, n_2) = \begin{cases} 2 \cos(2\pi k \cdot (n/M)), & \text{if } n_1 \neq 0, \\ \cos(2\pi k_2 n_2 / M), & \text{if } n_1 = 0. \end{cases} \tag{3.8}$$

A number of comments are in place concerning Lemma 3.1. First, note that the relation (3.8) implies that $c_{(M,M)-k}(n) = c_k(n)$. Therefore,

$$g_k(\widetilde{r}) \geq 0, \quad k \in G_+(M), \tag{3.9}$$

yields nonnegative $g_k(\widetilde{r})$ for general k in the larger grid $G(M)$. Thus, the number of inequality constraints in (3.5) can be reduced by half. Second, it also follows from (3.8) that g_k are real-valued.

We will not use Lemma 3.1 to calculate the eigenvalues $g_k(\widetilde{r}) = [A\widetilde{r}](k)$ – these can be computed efficiently by using the two-dimensional FFT. The relation (3.7), however, plays a significant role in the practical implementation of the OCE method in the following sense. As can be seen below in Section 3.2, the algorithm we use to solve the problem (3.5) requires multiple evaluations of both $[A\widetilde{r}](k)$ and $[A^T\widetilde{r}](k)$, where A^T refers to the adjoint operator of A (if A is viewed as a matrix, A^T is its transpose). We show in the next lemma, that $[A^T\widetilde{r}](k)$ can be computed using $[A\widetilde{r}](k)$ and hence FFT. The proof of the lemma is based on (3.7) and can be found in Appendix A.

Lemma 3.2 *The operators A and A^T are related as*

$$[A^T y](k) = [Ay](k) + E_k, \quad k \in G_+(M), \tag{3.10}$$

where

$$E_k = \begin{cases} -\sum_{n_2=0}^{M-1} \cos(2\pi k_2 n_2/M) \left[\sum_{n_1=1}^{\widetilde{N}-1} y(n_1, n_2) \right], & \text{if } k_1 = 0, \\ \sum_{n_2=0}^{M-1} \cos(2\pi k_2 n_2/M) y(0, n_2), & \text{if } k_1 \neq 0. \end{cases}$$

Note that E_k in the lemma above can also be calculated efficiently using FFT. We are now ready to present the method to solve the optimization problem (3.5), which by the discussion around (3.9) becomes

$$\min_{\widetilde{r}} \quad f(\widetilde{r}) = \sum_{n \in G_+(M)} \beta(n) \big(r(n) - \widetilde{r}(n) \big)^2,$$
$$\text{subject to} \quad g_k(\widetilde{r}) \geq 0, \quad k \in G_+(M). \tag{3.11}$$

3.2 Primal log-barrier method

To solve (3.11), we use the *primal log-barrier* (PLB) method. We outline the method below in order to refer to its parts that will be specialized in our problem. For a more detailed account, see Chapter 11 in Boyd and Vandenberghe (2004), and Section 19.6 in Nocedal and Wright (2006).[1]

We start by recalling some convex optimization notions from Boyd and Vandenberghe (2004), p. 128, adjusting the notation suitably for two-dimensional fields. The *optimal value* z^* of the problem (3.11) is defined as $z^* = \inf\{f(\widetilde{r}) \mid g_k(\widetilde{r}) \geq 0, \ k \in G_+(M)\}$. We will say that a (field) point $\widetilde{r}$ is *feasible* if $g_k(\widetilde{r}) \geq 0, \ k \in G_+(M)$ (if $g_k(\widetilde{r}) > 0, \ k \in G_+(M)$, the point $\widetilde{r}$ is called strictly feasible). The set of all feasible points will be called *feasible*

---

[1]The reader familiar with optimization methods may skip to the end of the section where the use of FFT in the CG method is discussed.

region. A (field) point $\widetilde{r}^*$ is said to be an *optimal point*, or to solve the problem (3.11) if it is feasible and $f(\widetilde{r}^*) = z^*$. Moreover, a feasible point $\widetilde{r}$ with $f(\widetilde{r}) \leq z^* + \epsilon$ ($\epsilon > 0$) is called $\epsilon-suboptimal$.

To ease the presentation of the PLB method, we break it down into 3 parts described next. The actual steps of the PLB algorithm are outlined at the end of this section.

Part 1: Eliminating the constraints

Define the *logarithmic barrier* function

$$\phi(\widetilde{r}) = - \sum_{k \in G_+(M)} \log(g_k(\widetilde{r})), \tag{3.12}$$

with the domain $\{\widetilde{r} = \widetilde{r}(n) \in G_+(M) \mid g_k(\widetilde{r}) > 0, k \in G_+(M)\}$. To eliminate the constraints in (3.11), the logarithmic barrier ϕ is introduced as a penalty term in the objective function of (3.11). More specifically, let $t > 0$ and consider the unconstrained problem

$$\min_{\widetilde{r}} \quad f_t(\widetilde{r}) := tf(\widetilde{r}) + \phi(\widetilde{r}). \tag{3.13}$$

If $g_k(\widetilde{r}) < 0$, then by convention the value of f_t in (3.13) is ∞.

One can show that the minimizers of (3.13), which we denote by $\widetilde{r}_t^*(n)$, approach a solution of (3.11) as t grows, under certain conditions (see, for example, Theorem 3.10 in Forsgren et al. (2002), p. 548, for a detailed proof and Boyd and Vandenberghe (2004), pp. 562-563, for an intuitive discussion). The points $\widetilde{r}_t^*(n)$ form the so-called *central path* $\{\widetilde{r}_t^*(n), t > 0\}$, and are at most m/t–*suboptimal* for the problem (3.11), i.e.

$$f(\widetilde{r}_t^*(n)) - z^* \leq m/t, \tag{3.14}$$

where $m = \widetilde{N}M$ is the number of inequality constraints in (3.11). See Boyd and Vandenberghe (2004), pp. 565-566, for a proof.

Part 2: Quadratic approximation

We now turn to solving the unconstrained optimization problem (3.13). It is convenient here to view each f_t as a function whose argument is a vector $x \in \mathbb{R}^m$, where m is the number of points in the grid $G_+(M)$. Consider the second-order Taylor approximation $\widehat{f}_t$ of f_t around some given $x = x_0$ (for a fixed t)

$$\widehat{f}_t(x + v) = f_t(x) + \nabla f_t(x)v + \frac{1}{2}v^T \nabla^2 f_t(x)v, \tag{3.15}$$

where ∇f_t and $\nabla^2 f_t$ denote the gradient and Hessian of f_t, respectively. Instead of minimizing the function $f(x)$, we will choose a point x_0 and find the direction v that minimizes the Taylor approximation of f_t around x_0. In other words, we will solve the quadratic problem

$$\min_{v} \quad \nabla f_t(x)v + \frac{1}{2}v^T \nabla^2 f_t(x)v \tag{3.16}$$

for $x = x_0$. In the optimization literature, the vector v is called the *Newton direction* and the process of calculating v is called the *Newton step*. Carrying out multiple Newton steps yields a sequence of points that converges to the solution of the exact problem (3.13).

11

For our problem, however, since t is increasing, it is not necessary to find exact minimizers of the problem (3.13). In fact, it is sufficient to calculate *only one* Newton direction for each value of t and initial value $x = x_0$. This fact is argued in Bertsimas and Tsitsiklis (1997) for linear objective functions (see the discussion under relation (9.19) on p. 424 and figure 9.6 on p. 425) and in Boyd and Vandenberghe (2004), p. 570, for general convex functions.

As expected, the approximate problem (3.16) is much easier to solve. The first-order condition linear system is

$$Hv = b, \tag{3.17}$$

where H and b are given by

$$H = \nabla^2 f_t(x), \qquad b = -\nabla f_t(x). \tag{3.18}$$

Thus, the minimizer of $\widehat{f}_t$, for a fixed t and initial point $x = x_0$, is given by

$$\widehat{x}^* = x_0 + v^*, \tag{3.19}$$

where v^* is the solution of (3.17).

Part 3: Conjugate gradient algorithm

The *conjugate gradient* (CG) algorithm is an iterative procedure for solving linear systems of the form (3.17) with symmetric and positive definite matrices H. It is particularly appropriate for large problems, which is the case with (3.11) even for moderate M. Moreover, in our case, the key steps of the algorithm can be implemented more efficiently through FFT. We give next a description of the CG method following Nocedal and Wright (2006), p. 112.

Outline of the CG algorithm

Given v_0; Set $\epsilon_0 = Hv_0 - b$, $p_0 = \epsilon_0$, $k = 0$;

while $\epsilon_k \neq 0$

$$
\begin{aligned}
\alpha_k &\leftarrow \frac{\epsilon_k^T \epsilon_k}{p_k^T H p_k}; \\
v_{k+1} &\leftarrow v_k + \alpha_k p_k; \\
\epsilon_{k+1} &\leftarrow \epsilon_k + \alpha_k H p_k; \\
s_{k+1} &\leftarrow \frac{\epsilon_{k+1}^T \epsilon_{k+1}}{\epsilon_k^T \epsilon_k}; \\
p_{k+1} &\leftarrow -\epsilon_{k+1} + s_{k+1} p_k; \\
k &\leftarrow k + 1;
\end{aligned}
$$

end (**while**)

The major computational tasks to be performed at each iteration of the CG algorithm, are calculating the products Hp_k, $p_k^T H p_k$ and $\epsilon_{k+1}^T \epsilon_k$. In our case, however, these calculations can be done more efficiently using FFT. This is possible because, as shown in Lemma 3.3

below, both the Hessian matrix H and the negative gradient b can be expressed as functions of the linear operators A and A^T, which can be implemented using FFT by Lemmas 3.1 and 3.2. Indeed, note that a direct matrix-vector calculation of Hp is of the order m^2 for an $m \times m$ matrix H and an $m \times 1$ vector p. In our problem, m is of the order N^2 and hence m^2 of the order N^4. On the other hand, if H refers to FFT and p to a field, the order of calculation becomes $N^2 \log(N)$.

In the practical implementation of the CG algorithm, we considered the stopping rule $\epsilon_k < \text{TOL}$ for a tolerance $\text{TOL} = 0.1$. For all covariance functions considered in Section 4, we found that calculating the Newton step with higher precision (taking smaller tolerance) led to the same optimal value for f, requiring however a larger number of steps.

Before we state Lemma 3.3, some notation is necessary. Given a field $r = \{r(n), n \in G_+(M)\}$, define $\{d(k), k \in G_+(M)\}$ to be the field whose kth element is given by

$$d(k) = -([Ar](k))^{-1} = -1/([Ar](k)). \tag{3.20}$$

Next, let $D := \text{diag}(d)$ be a linear operator whose action on the field r is defined as

$$[Dr](k) = d(k) \cdot r(k), \quad k \in G_+(M). \tag{3.21}$$

We are now ready to state Lemma 3.3. The proof of the lemma can be found in Appendix A.

Lemma 3.3 *Let $\beta(n)$ be the weights defined in (3.4). Then, the Hessian matrix H and the negative gradient b in (3.17) can be viewed as a linear operator and a two-dimensional field given by, respectively,*

$$H = tW + A^T D^2 A, \quad and \quad b(n) = t(W\widetilde{r}(n) - s(n)) + [A^T d](n), \tag{3.22}$$

where $W = \text{diag}(\beta(n))$ and $s(n) = \beta(n)r(n)$.

We conclude with a general outline of the primal log-barrier method. Further comments can be found in Section 3.3. Let $\widetilde{r}^*_{t,a}$ denote the field analogue of the approximate minimizer $\widehat{x}^*$ in (3.19), i.e. the solution of the optimization problem

$$\min_{\widetilde{r}} \widehat{f}_t(\widetilde{r}). \tag{3.23}$$

Steps of the PLB method

1. Find an initial strictly feasible point $\widetilde{r}^0$ for the problem (3.11). Pick constants $t_0 > 0$, $\mu > 1$, and accuracy $\epsilon > 0$. Set $t = t_0$ and $\widetilde{r} = \widetilde{r}^0$.

2. Compute $\widetilde{r}^*_{t,a}$ by solving the problem (3.23) or equivalently (3.17) with initial point $\widetilde{r}$, using the CG algorithm.

3. Update $\widetilde{r} = \widetilde{r}^*_{t,a}$. If $m/t < \epsilon$, stop and return $\widetilde{r}$; since $f(\widetilde{r}^*_{t,a})$ $(\approx f(\widetilde{r}^*_t))$ does not differ from the objective at the optimum by more than m/t, we can ensure that $f(\widetilde{r})$ is less than ϵ away from the optimal value.

4. Increase t to μt and start again from Step 2.

3.3 Further discussion on the PLB method

In this section, we discuss the choice of the parameters $\widetilde{r}_0$, t_0 and μ above. The initial covariance embedding $\widetilde{r}_0$ needs to be strictly feasible (i.e. $[A\widetilde{r}_0](k) > 0$ for all k). We describe below a straightforward way to choose such a point. The roles of t_0 and μ, on the other hand, are less straightforward. Note that the domain of the function $\widehat{f}_t$ is different from the feasible region of the constrained problem (3.11). This means that the CG algorithm used to find approximate minimizers of f_t will possibly result in minimizers of $\widehat{f}_t$ which are not feasible points for (3.11). The parameters t_0 and μ are selected to mitigate this effect.

Choice of $\widetilde{r}_0$: To find a strictly feasible initial point $\widetilde{r}^0$ in Step 1 of the PLB method, we first calculate the standard covariance embedding $\widetilde{r}(n)$, $n \in G(M)$, as defined through the relation (2.3). Let $F(k) := [A\widetilde{r}](k)$, $k \in G(M)$, be the corresponding spectrum (which can be calculated efficiently through FFT). We then shift the spectrum by a positive constant c so that all of its elements become positive (taking $c = 2|\min_{k \in G(M)} F(k)|$ will ensure that). Denote the resulting spectrum by $F_+(k)$. Finally, let $y(n), n \in G(M)$, denote the inverse FFT of $F_+(k)$. Taking $\widetilde{r}^0(n) = r(0)y(n)/y(0)$ yields a strictly feasible initial point with the targeted scale $\widetilde{r}^0(0) = r(0)$.

Choice of t_0: Observe that $\widehat{f}_t$ can be written as

$$\widehat{f}_t(\widetilde{r}) = tf(\widetilde{r}) + \widehat{\phi}(\widetilde{r}), \tag{3.24}$$

where $\widehat{\phi}$ denotes the second-order Taylor approximation of the logarithmic barrier ϕ (defined in a similar manner as $\widehat{f}_t$ in (3.15)). Selecting a small t_0 essentially diminishes the contribution of f to $\widehat{f}_t$ and enhances the role of $\widehat{\phi}$, in the early steps of the PLB method. For points outside the feasible region, the value of ϕ is (by our convention) infinite, and thus we expect the value of $\widehat{\phi}$ to be very large. As a result, these points will not be favored by the CG algorithm in the initial iterations of the PLB method. For our simulations (see Section 4), we used a small $t_0 = 0.0001$, which proved to work well for all the covariance structures considered. See Boyd and Vandenberghe (2004), pp. 570-571, for more elaborate choices for t_0.

Choice of μ: When μ is small (near 1), two consecutive t's in the PLB method are not very different. As a result, the problem (3.23) is perturbed only slightly from one iteration to another. Moreover, the update rule in Step 3 of the algorithm implies that the initial point used when solving the problem (3.23) is the solution of the previous iteration. Since the two problems are only slightly different, we expect this choice of initial point to be a good one, and therefore the conjugate gradient algorithm will require less time to find the solution. The downside of this strategy is that a large number of iterations will be required to reach the suboptimality property $m/t < \epsilon$ in Step 3 of the method.

On the other hand, when μ is large, t is increased rapidly at each PLB iteration and thus the convergence to the suboptimality property is much faster. This aggressive update of t, however, means that there will be a large difference between the problems (3.23) at two consecutive iterations, and therefore the solution of one iteration might not be a good initial

point for the next. This might decrease the quality of the Taylor approximation, potentially leading to a solution outside the feasible region.

When t is updated aggressively (large μ), note also that the first term in (3.24) dominates the objective function and thus the CG algorithm focuses more on iterates that minimize f in (3.11) and less on iterates that satisfy the constraints in (3.11). In other words, large values of μ are more likely to lead to an exact circulant embedding, which however might have some negative eigenvalues. Taking smaller values of μ will ensure that the eigenvalues are nonnegative, however leading to approximate embeddings. We discuss these situations further in Section 4.

4   Simulations

In this section, we provide a Monte Carlo study comparing the OCE and SWCE methods, as well as numerical illustrations of all the circulant embedding methods discussed above. For our comparisons, we consider the *anisotropic* covariance function of the powered exponential form

$$r(n) = e^{-(\theta||n||_W)^\alpha}, \quad 0 < \alpha \le 2, \quad \theta > 0, \tag{4.1}$$

where $||t||_W := \sqrt{t'Wt}$ for column vectors t and W is a symmetric positive definite 2×2 matrix, and the *isotropic* Cauchy covariance function of the form

$$r(n) = (1 + (\theta||n||_2)^\alpha)^{-\beta/\alpha}, \quad 0 < \alpha \le 2, \quad \beta > 0, \quad \theta > 0. \tag{4.2}$$

For the powered exponential covariance structure (4.1), we consider two configurations, with $\alpha = 0.5$, $\theta = 0.01$ and $W = W_1$ or W_2, where

$$W_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 1.6388 & -1.489 \\ -1.489 & 1.3712 \end{pmatrix}. \tag{4.3}$$

(While there is nothing particularly special about the choice of W_1, the matrix W_2 is chosen to be nearly singular as discussed in Section 4.1 below.) For the Cauchy covariance structure (4.2), we consider the following parameter values

$$\alpha = 1.3, \quad \beta = 0.01, \quad \theta = 0.01. \tag{4.4}$$

For the SWCE method, overlapping windows are used throughout.

4.1   Powered exponential covariance

Figure 1 shows the covariance embeddings resulting from the standard CME, SWCE and OCE procedures for $N = 200$ and $\widetilde{N} = 240$ (transition region length 40) when the matrix W_1 is used in (4.1). In the standard CME (top left), the covariance embedding is not smooth at the boundary of periodization resulting in many negative eigenvalues (77,300). The SWCE method (top right), as expected, smooths the covariance embedding at the boundary. However, for this transition region length, some of the eigenvalues (50) are still negative. Finally, the covariance embeddings of the OCE method (middle left and middle

15

right plots) have no negative eigenvalues, while their values within the corner areas marked with broken lines are those of the desired covariance (up to the indicated error min f). These plots visualize the goal of the optimization problem (3.2). The two bottom plots show 2 cross sections of the covariance embeddings from all 3 methods.

Figure 2 shows the covariance embeddings resulting from the standard CME, SWCE and OCE procedures for $N = 100$ and $\widetilde{N} = 359$ (transition region length 80) when the matrix W_2 is used in (4.1). The matrix W_2 is nearly singular having eigenvalues 0.01 and 3. This singularity causes more pronounced discontinuities at the boundary of periodization – c.f. the standard CME plots in Figures 1–2 (upper left). The SWCE method smooths the discontinuities. However, the resulting embedding still has some negative eigenvalues (116). Moreover, even though the length of the transition region here is larger compared to Figure 1, the number of negative eigenvalues has increased. This is due to the ill behavior of the weight matrix W. On the other hand, the two middle plots of Figure 2 depict the covariance embedding of the OCE method with no negative eigenvalues. The two bottom plots show 2 cross sections of the covariance embeddings from all 3 methods.

4.2 Cauchy covariance

Figure 3 shows the covariance embeddings resulting from the CME, SWCE and OCE procedures for the Cauchy covariance function (4.2) with parameters (4.4) and $N = 200$. The transition region length for the upper right (SWCE) and middle left (OCE) covariance embeddings is 40, while for the middle left (SWCE) is 400. As in Figures 1–2, the covariance embedding resulting from the standard CME method (upper left) is not smooth at the boundary of periodization (2,836 negative eigenvalues). Note that non-smoothness here is in the derivative (and not in the covariance itself as in Figures 1–2, top left) as indicated through a highlighted contour plot which is spiky at the boundary.

In the top-right plot where the transition region length is 40, the SWCE method results in an embedding with 24,720 negative eigenvalues. One possible explanation for this is a very small range of the values of the given covariance function (the range of approximate size 0.05 as seen from the vertical scale). That is, when overlapping windows are used in the SWCE, the values of the covariance function are superimposed according to (2.19) and as in the case considered here, can result in larger values over the transition region (the white region in the middle of the top-right plot). In fact, the SWCE method still fails to produce a covariance embedding with nonnegative eigenvalues even when $\widetilde{N} = 600$ (for example, the smoothing covariance embedding with $\widetilde{N} = 400$ shown in the middle right plot has 28 negative eigenvalues). The OCE method, on the other hand, finds a feasible covariance embedding even when $\widetilde{N} = 240$. The two bottom plots show 2 cross sections of the covariance embeddings from all 3 methods.

4.3 Efficiency and related issues

We compare here the performance of the SWCE and OCE methods. Following Helgason et al. (2014), define the efficiency of the embedding as

$$\gamma = \gamma(M) = \frac{2N^* - 1}{M}, \tag{4.5}$$

16

where $M = 2\widetilde{N} - 1$ is the embedding size. For the SWCE method, $N^*(\leq N)$ is the largest size for which the covariance embedding has nonnegative eigenvalues (and hence for which the field can be generated exactly on the grid $G(N^*)$). The efficiency γ satisfies $0 < \gamma \leq 1$ and the closer γ is to 1, the larger the grid $G(N^*)$ one can simulate the field on.

For the OCE, the value N^* can be defined in several ways. One possibility is to take N^* as the largest value for which the covariance embedding yields the objective function smaller than some small value, for example, 10^{-5}. Another possibility which we find more informative, is to take N^* as the largest value for which the OCE method results in a feasible embedding (i.e. an embedding with nonnegative eigenvalues) – see Section 3.3 for a related discussion. The value of N^*, however, needs to be supplemented by the value of the objective function since the latter is not necessarily small. This choice of N^* is used below, though we also comment on what N^* is expected under the first possibility above.

In Figures 4–5, we compare the embedding efficiency of the OCE (solid line) and the SWCE (dashed line) methods. Figure 4 compares the embedding efficiency for the powered exponential covariance (4.1) as a function of $M\theta$ (this choice, rather than θ, allows a nicer scale on the $x-$axis). The comparisons were done for 3 different values of $\alpha = 0.5, 1, 1.5$ (corresponding to the three rows), $\theta = 0.05k$, $k = 1, \ldots, 10$, and embedding size $M = 101$. The OCE method was implemented for $\mu = 2$ (left column) and $\mu = 3$ (right column), and the optimum values of f in (3.11) are given below each point on the solid lines. The OCE method performs significantly better than the SWCE in all cases considered (the solid lines are located above the dashed ones). Note also that as α decreases, the performance of the SWCE method gets worse, while the OCE method is only slightly affected.

Focusing on the OCE method, the plots illustrate nicely the tradeoff in the choice of μ discussed in Section 3. Observe, for example, in the second (middle) row of Figure 4, that the embedding efficiency γ has higher values for $\mu = 2$ (left) than for $\mu = 3$ (right). This means that, for the given covariance $r(n), n \in \widetilde{G}(N)$, and the given size M, taking $\mu = 2$ allows the synthesis of the desired field on larger grids, than taking $\mu = 3$. The synthesized fields, however, will not necessarily be exact, since the optimum value of the objective function is often non-negligible (e.g. for $\mu = 2$ and $M\theta = 20$ the optimum value is 1). On the other hand, the optimum values of the objective function f in (3.11) are lower for $\mu = 3$ (right plots) illustrating the fact that larger values of μ are more likely to yield exact embeddings.

Note also that, for example, all the values of the objective function in the top-right plot $(\mu = 3)$ of Figure 4 are smaller than 10^{-5}. If we defined the efficiency for the OCE method requiring the objective function to be smaller than 10^{-5} at N^*, the resulting top-left plot $(\mu = 2)$ of Figure 4 would naturally have a lower efficiency for the OCE method. We find, however, that this efficiency would still be higher than the one for $\mu = 3$ in the top-right plot. The same observation applies to other plots concerning the cases $\mu = 2$ and $\mu = 3$.

Remark 4.1 A field $X = \{X_n, n \in G(N)\}$ on grids of small size ($N < 100$) can be synthesized exactly using Cholesky decomposition. However, even a fast variant of the method developed by Dietrich (1993) has complexity $O(N^5)$, and thus is not suitable for the synthesis of fields on grids of size larger than $N = 100$ (see Table 2 in Gneiting et al. (2006), p. 485, for a detailed comparison of the computational requirements of Cholesky decomposition and standard CME methods). On the other hand, the OCE method works well for grids of size up to $N = 1,000$. Note also that the speed of the OCE method varies depending on how

strong the discontinuities of the covariance embedding are. For the Cauchy covariance of Figure 3 for example, where $N = 200$ and $M = 479$, the OCE method for $\mu = 2.5$ needs 150 seconds to produce a feasible covariance embedding.[2] On the other hand, for the powered exponential covariance function as in Figure 2 but for $N = 60$ and $M = 215$, the time needed is almost the same (140 secs) even though the grid size is much smaller.

Remark 4.2 Regarding the computational requirements of the SWCE and OCE methods, at first glance, the SWCE method seems to be much faster. As discussed above, however, the SWCE method can break down for small transition regions. In fact, the minimum transition region length needed for the method to work is not known in advance. This means that to produce a feasible covariance embedding, it may be necessary to perform the SWCE method multiple times, each time increasing the transition region length until a feasible covariance embedding is produced. Although this trial and error approach can be optimized, it is still likely that a large number of transition region lengths will need to be tested. In other words, the OCE and SWCE methods should not be compared in speed solely for fixed transition regions leading to feasible embeddings, where the SWCE method is likely to be faster (even if the corresponding transition region is larger). One should also take into account the uncertainty in the size of the transition region, where the OCE method is the favorite (in allowing for smaller regions).

Finally, note again that the approach taken in this work stresses nonnegative eigenvalues of covariance embeddings and focuses on the values of the objective function. Related to this, it is natural to ask how our approach compares to the following simple procedure. If the standard CME fails, a straightforward way to obtain a feasible covariance embedding $\widetilde{r}_N$ is by setting all negative eigenvalues of the CME covariance $\widetilde{r}_I$ equal to 0. Variations of this are suggested, for example, in Chan and Wood (1997), p. 167. Moreover, by using Parseval's identity, note that

$$\underset{\widetilde{r}}{\operatorname{argmin}} \, ||\widetilde{r}_I - \widetilde{r}||_2 = \underset{\widetilde{r}}{\operatorname{argmin}} \, ||g(\widetilde{r}_I) - g(\widetilde{r})||_2 = \widetilde{r}_N,$$

where $||\cdot||_2$ denotes the Frobenius norm, $g(\widetilde{r}) = \{g_k(\widetilde{r}), k \in \widetilde{G}(N)\}$ are the eigenvalues of the covariance embedding $\widetilde{r}$ and the minimum is taken over all covariance embeddings $\widetilde{r}$ having nonnegative eigenvalues. In other words, $\widetilde{r}_N$ is the solution to our optimization problem (3.2) when $\widetilde{N} = N$.

Indeed, when the transition region is empty ($\widetilde{N} = N$), the OCE method results in the covariance embedding $\widetilde{r}_O$ very close to $\widetilde{r}_N$. On the other hand, when $\widetilde{N} > N$, the OCE method can result in significantly smaller optimal values of the objective function (which moreover can be made smaller if desired). For example, in the case of the powered exponential covariance function of Figure 1, the optimal value in the OCE method is less than 10^{-6}, while the above modification of the CME method leads to the Frobenius distance between $\widetilde{r}_I$ and $\widetilde{r}_N$ equal to 24.

---

[2]All simulations were performed in an Intel Core i5 processor with 4GB of RAM.

A Appendix

We first prove Lemma 3.1.

PROOF OF LEMMA 3.1: First we write (3.1) as

$$
\widetilde{r}(n_1, n_2) = \begin{cases} \widetilde{r}(n_1, n_2), & \text{if} \quad (n_1, n_2) \in Q_1, \\ \widetilde{r}(n_1 - M, n_2 - M), & \text{if} \quad (n_1, n_2) \in Q_3, \\ \widetilde{r}(n_1, n_2 - M), & \text{if} \quad (n_1, n_2) \in Q_2, \\ \widetilde{r}(n_1 - M, n_2), & \text{if} \quad (n_1, n_2) \in Q_4, \end{cases} \tag{A.1}
$$

where the regions Q_1, Q_2, Q_3 and Q_4 are given by

$$
\begin{aligned}
Q_1 &= \{(n_1, n_2) : 0 \le n_1 \le \widetilde{N} - 1, \ 0 \le n_2 \le \widetilde{N} - 1\}, \\
Q_2 &= \{(n_1, n_2) : 0 \le n_1 \le \widetilde{N} - 1, \ \widetilde{N} \le n_2 \le M - 1\}, \\
Q_3 &= \{(n_1, n_2) : \widetilde{N} \le n_1 \le M - 1, \ \widetilde{N} \le n_2 \le M - 1\}, \\
Q_4 &= \{(n_1, n_2) : \widetilde{N} \le n_1 \le M - 1, \ 0 \le n_2 \le \widetilde{N} - 1\}
\end{aligned} \tag{A.2}
$$

and partition the grid $G(M)$. By using this partition, we rewrite the eigenvalues $g_k(\widetilde{r})$ in (2.8) as

$$
g_k(\widetilde{r}) = \left(\sum_{n \in Q_1} + \sum_{n \in Q_2} + \sum_{n \in Q_3} + \sum_{n \in Q_4} \right) \widetilde{r}(n) e^{-i2\pi k \cdot n / M}. \tag{A.3}
$$

The first sum of (A.3) can be written as

$$
\sum_{n \in Q_1} \widetilde{r}(n_1, n_2) e^{-i2\pi k \cdot n / M} = \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1, n_2) e^{-i2\pi k \cdot n / M} + \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(0, n_2) e^{-i2\pi k_2 n_2 / M}
$$

$$
+ \sum_{n_1=1}^{\widetilde{N}-1} \widetilde{r}(n_1, 0) e^{-i2\pi k_1 n_1 / M} + \widetilde{r}(0, 0). \tag{A.4}
$$

Using the symmetry condition (3.1) and the property $\widetilde{r}(n) = \widetilde{r}(-n)$, we have for the third

sum of (A.3),

$$
\begin{aligned}
\sum_{n \in Q_3} \widetilde{r}(n_1, n_2) e^{-i2\pi k \cdot n/M} &= \sum_{n_1=\widetilde{N}}^{M-1} \sum_{n_2=\widetilde{N}}^{M-1} \widetilde{r}(n_1, n_2) e^{-i2\pi k \cdot n/M} \\
&= \sum_{n_1=-\widetilde{N}+1}^{-1} \sum_{n_2=-\widetilde{N}+1}^{-1} \widetilde{r}(n_1 + M, n_2 + M) e^{-i2\pi k \cdot (n+M)/M} \\
&= \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(M - n_1, M - n_2) e^{i2\pi k \cdot n/M} \\
&= \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1 - M, n_2 - M) e^{i2\pi k \cdot n/M} \\
&= \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1, n_2) e^{i2\pi k \cdot n/M}.
\end{aligned}
$$
$$(A.5)$$

By combining the relations (A.4) and (A.5), we get

$$
\left(\sum_{n \in Q_1} + \sum_{n \in Q_3} \right) \widetilde{r}(n) e^{-i2\pi k \cdot (n/M)} = 2 \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1, n_2) \cos\left(2\pi \left(\frac{k_1 n_1}{M} + \frac{k_2 n_2}{M} \right) \right)
$$
$$
+ \sum_{n_1=1}^{\widetilde{N}-1} \widetilde{r}(n_1, 0) e^{-i2\pi k_1 n_1/M} + \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(0, n_2) e^{-i2\pi k_2 n_2/M} + \widetilde{r}(0, 0). \quad (A.6)
$$

Similar calculations for the second and fourth sums of (A.3) yield

$$
\left(\sum_{n \in Q_2} + \sum_{n \in Q_4} \right) \widetilde{r}(n) e^{-i2\pi k \cdot (n/M)} = 2 \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=\widetilde{N}}^{M-1} \widetilde{r}(n_1, n_2) \cos\left(2\pi \left(\frac{k_1 n_1}{M} + \frac{k_2 n_2}{M} \right) \right)
$$
$$
+ \sum_{n_1=1}^{\widetilde{N}-1} \widetilde{r}(n_1, 0) e^{i2\pi k_1 n_1/M} + \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(0, n_2) e^{i2\pi k_2 n_2/M}. \quad (A.7)
$$

Combining (A.6) and (A.7) yields

$$
\begin{aligned}
g_k(\widetilde{r}) &= 2 \sum_{n_1=1}^{\widetilde{N}-1} \sum_{n_2=1}^{M-1} \widetilde{r}(n_1, n_2) \cos\left(2\pi \left(\frac{k_1 n_1}{M} + \frac{k_2 n_2}{M} \right) \right) \\
&\quad + 2 \sum_{n_1=1}^{\widetilde{N}-1} \widetilde{r}(n_1, 0) \cos\left(\frac{2\pi k_1 n_1}{M} \right) + 2 \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(0, n_2) \cos\left(\frac{2\pi k_2 n_2}{M} \right) + \widetilde{r}(0, 0). \quad (A.8)
\end{aligned}
$$

The first two terms on the right-hand side of (A.8) give $c_k(n)$ in (3.8) for $n_1 \neq 0$, and the last two terms give $c_k(n)$ in (3.8) for $n_1 = 0$. $\quad \square$

20

For the proofs of Lemmas 3.2 and 3.3 we first obtain an expression for the adjoint operator A^T of A similar to (3.7). More specifically, we will show that A^T satisfies

$$[A^T \widetilde{r}](k) = \sum_{n \in G_+(M)} c_n(k) \widetilde{r}(n), \quad k \in G_+(M). \tag{A.9}$$

Note that the only difference between the two operators A and A^T is that $c_k(n)$ in the expression (3.7) of A is interchanged with $c_n(k)$ in the expression (A.9) of A^T.

We will show (A.9) by viewing A and A^T as matrices. To make the transition to a matrix point of view, let m be the number of points in the grid $G_+(M)$ and consider two arbitrary fixed bijective mappings $\phi_i(u) : \{1, \ldots, m\} \to G_+(M)$, $i = 1, 2$, for rearranging the values of the fields $\widetilde{r}(n)$ and $c_k(n)$ in the relation (3.7) into vectors. More specifically, let

$$n = \phi_1(j) \quad \text{and} \quad k = \phi_2(l), \quad j, l = 1, \ldots, m,$$

for $n, k \in G_+(M)$. Then, we can interpret the two-dimensional field $\widetilde{r}(n), n \in G_+(M)$, as a column vector $\widetilde{r}_v$ whose jth entry is $\widetilde{r}(\phi_1(j))$. Similarly, for each k, the coefficients $c_k(n), n \in G_+(M)$, can be viewed as a row vector a_l^T whose jth entry is $c_{\phi_2(l)}(\phi_1(j))$. This allows us to rewrite the relation (3.7) as

$$[A\widetilde{r}](k) = \sum_{n \in G_+(M)} c_k(n) \widetilde{r}(n) = \sum_{j=1}^m a_l^T(j) \widetilde{r}(\phi_1(j)) = a_l^T \widetilde{r}_v = (A\widetilde{r}_v)_l, \tag{A.10}$$

where A in the last equation is viewed as a matrix with rows a_l^T, $l = 1, \ldots, m$, and $(\cdot)_l$ denotes the lth element of a vector. The subscript v in $\widetilde{r}_v$ is to avoid a possible confusion regarding which point of view is adopted, as $A\widetilde{r}$ will denote the action of the linear operator A on a field $\widetilde{r}$, whereas $A\widetilde{r}_v$ is the usual matrix-vector product.

Next, let b_l^T, $l = 1, \ldots, m$, denote the rows of the transpose matrix A^T of A. The jth entry of b_l^T satisfies

$$b_l^T(j) = a_j^T(l) = c_n(k), \quad \text{for } n = \phi_1(j) \text{ and } k = \phi_2(l).$$

Then, arguing as for (A.10) but in reverse order, we have

$$(A^T \widetilde{r}_v)_l = b_l^T \widetilde{r}_v = \sum_{j=1}^m b_l^T(j) \widetilde{r}(\phi_1(j)) = \sum_{n \in G_+(M)} c_n(k) \widetilde{r}(n), \tag{A.11}$$

which yields (A.9).

We are now ready to prove Lemmas 3.2 and 3.3.

PROOF OF LEMMA 3.2: In view of the relations (3.7) and (A.9), it is enough to show that the weights $c_n(k)$ satisfy

$$c_n(k) = c_k(n) - \cos(2\pi k_2 n_2 / M) 1_{\{k_1 = 0, n_1 \neq 0\}} + \cos(2\pi k_2 n_2 / M) 1_{\{k_1 \neq 0, n_1 = 0\}}. \tag{A.12}$$

To show that (A.12) holds, we simply use (3.8) and compare the values of $c_k(n)$ and $c_n(k)$ for the four cases of the values of n and k shown in Tables 1–2. $\quad \square$

$c_k(n)$	$n_1 \neq 0$	$n_1 = 0$
$k_1 \neq 0$	$2\cos(2\pi k \cdot (n/M))$	$\cos(2\pi k_2 n_2/M)$
$k_1 = 0$	$2\cos(2\pi k_2 n_2/M)$	$\cos(2\pi k_2 n_2/M)$

Table 1: The values of $c_k(n)$

$c_n(k)$	$n_1 \neq 0$	$n_1 = 0$
$k_1 \neq 0$	$2\cos(2\pi k \cdot (n/M))$	$2\cos(2\pi k_2 n_2/M)$
$k_1 = 0$	$\cos(2\pi k_2 n_2/M)$	$\cos(2\pi k_2 n_2/M)$

Table 2: The values of $c_n(k)$

PROOF OF LEMMA 3.3: Recall from the relation (3.18), that H and b are the Hessian and negative gradient of the function $f(x) = tf(x) + \phi(x)$, where f and ϕ are given in (3.11) and (3.12), respectively. To show that H and b satisfy the relations (3.22), we will consider the functions f and ϕ separately.

By using (3.7) and (A.10), we can express the function $\phi(\widetilde{r})$ in (3.12) from the vector perspective as

$$\phi(\widetilde{r}_v) = -\sum_{l=1}^{m} \log(a_l^T \widetilde{r}_v). \tag{A.13}$$

By using (A.10), we can also write the field $d(k)$ in (3.20) as a vector d_v whose lth entry $d_v(l)$ is $d(\phi_2(l)) = -(a_l^T \widetilde{r}_v)^{-1}$. Then, the gradient and Hessian of ϕ are given by

$$\nabla \phi(\widetilde{r}_v) = \sum_{l=1}^{m} \frac{1}{-a_l^T \widetilde{r}_v} a_l = A^T d_v, \tag{A.14}$$

$$\nabla^2 \phi(\widetilde{r}_v) = \sum_{l=1}^{m} \frac{1}{(a_l^T \widetilde{r}_v)^2} a_l a_l^T = A^T D^2 A, \tag{A.15}$$

where $D = \mathrm{diag}(d_v)$. As in the case of the operator/matrix A, the diagonal matrix $D = \mathrm{diag}(d_v)$ is the matrix analogue of the operator D defined in (3.21). Indeed, let $y = [Au](k)$, for some two dimensional field $u = \{u(n), n \in G_+(M)\}$. Let also u_v and y_v be the vectors whose jth elements are $u(\phi_1(j))$ and $a_j^T u_v$, respectively. Then, the action of D on y yields

$$
\begin{aligned}
[Dy](k) &:= d(k) \cdot y(k) \\
&= d(\phi_2(l)) \cdot [Au](\phi_2(l)) \\
&= d_v(l) a_l^T u_v,
\end{aligned}
$$

where the last term is the lth element of the matrix Dy_v.

Next, we calculate the gradient and Hessian of f in (3.11). Letting s_v be a vector whose jth entry is $s(\phi_1(j))$, we can write f in a quadratic form as

$$f(\widetilde{r}_v) = \widetilde{r}_v^T W \widetilde{r}_v - 2s_v^T \widetilde{r}_v + r_v^T W r_v. \tag{A.16}$$

22

Since the last term in the relation (A.16) is a constant, minimizing f is equivalent to minimizing the function

$$\tilde{f}_0(\widetilde{r}_v) = \frac{1}{2}\widetilde{r}_v^T W \widetilde{r}_v - s_v^T \widetilde{r}_v. \tag{A.17}$$

The gradient and Hessian of $\tilde{f}_0$ are given by

$$\nabla \tilde{f}_0(\widetilde{r}_v) = W\widetilde{r}_v - s_v, \tag{A.18}$$

$$\nabla^2 \tilde{f}_0(\widetilde{r}_v) = W. \tag{A.19}$$

Finally, by combining the relations (A.14)–(A.15) and (A.18)–(A.19), we get

$$\nabla f(\widetilde{r}_v) = t(W\widetilde{r}_v - s_v) + A^T d_v,$$

$$\nabla^2 f(\widetilde{r}_v) = tW + A^T D^2 A,$$

which are the vector equivalents of the relations (3.22). $\quad\square$
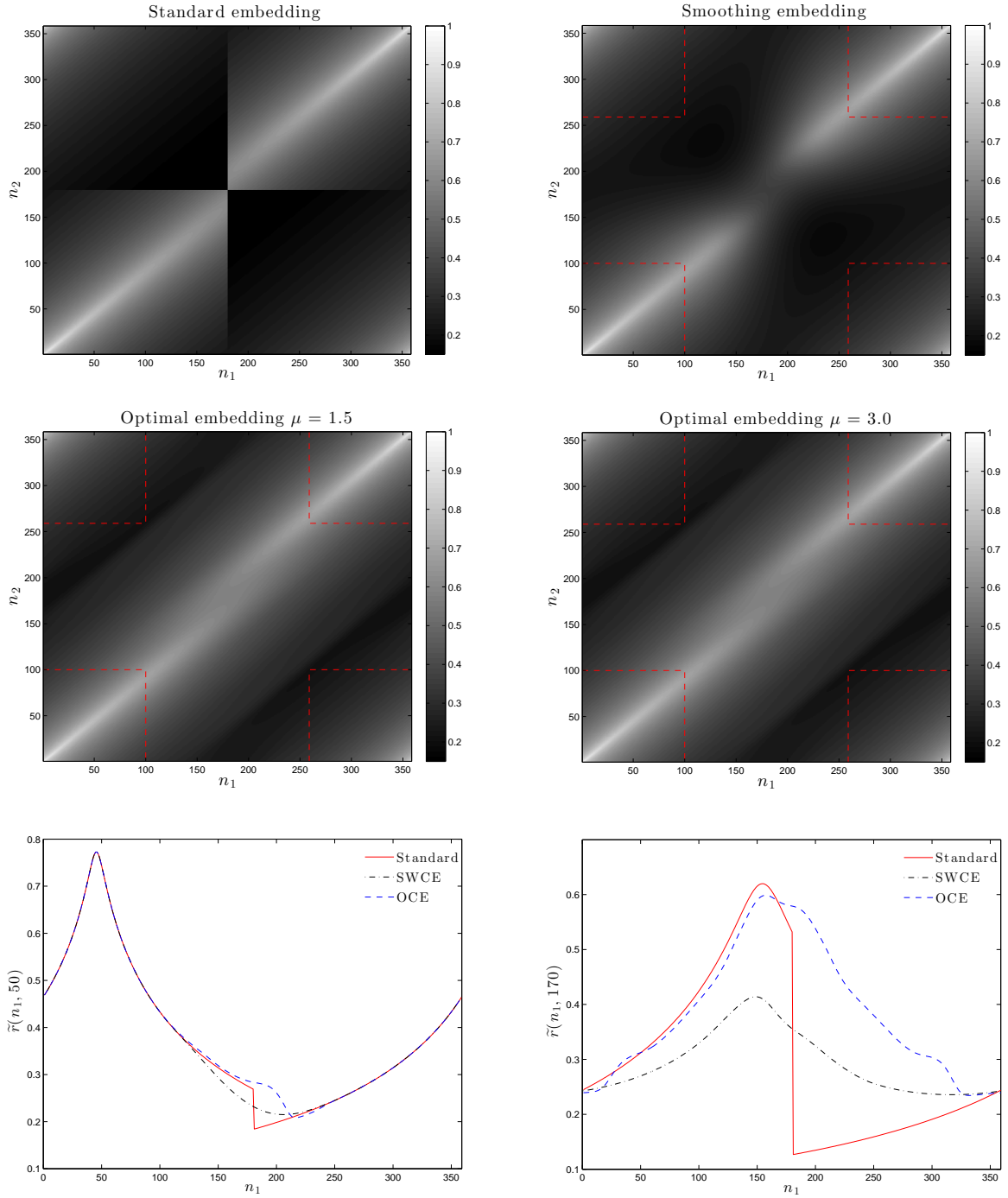
B   Figures and tables
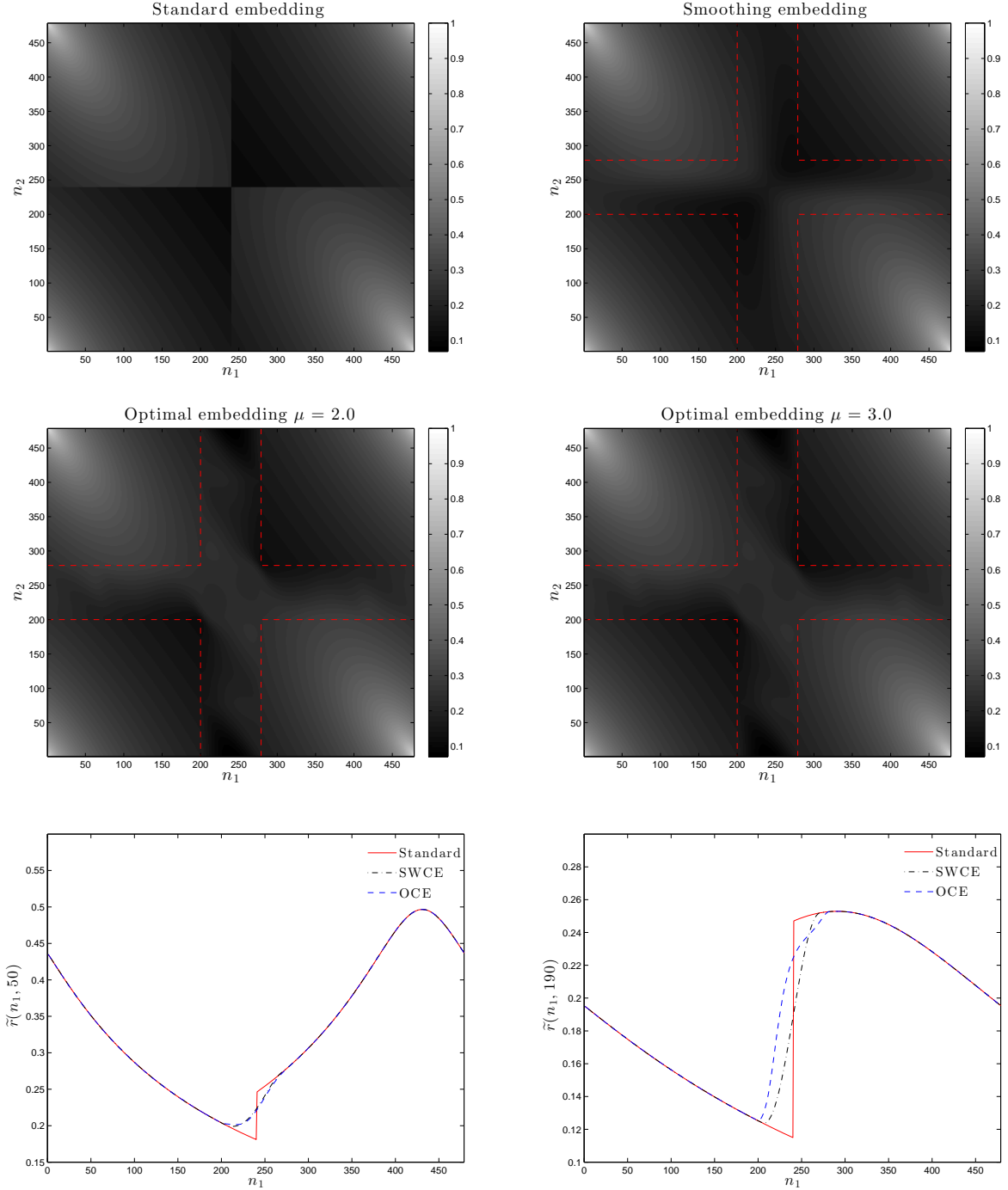
Figure 1: Plots for the anisotropic covariance of the powered exponential form in (4.1) with $W = W_1$ in (4.3), $N = 200$ and $M = 479$. Top left: Standard embedding. Top right: Smoothing windows embedding with 50 negative eigenvalues. Middle left: Optimal embedding with $\mu = 3$ and $\min f = 6.7 \cdot 10^{-7}$. Middle right: Optimal embedding with $\mu = 1.5$ and $\min f = 8 \cdot 10^{-7}$. Bottom plots: Cross sections of standard CME, SWCE and OCE ($\mu = 1.5$) for $n_2 = 50, 190$.
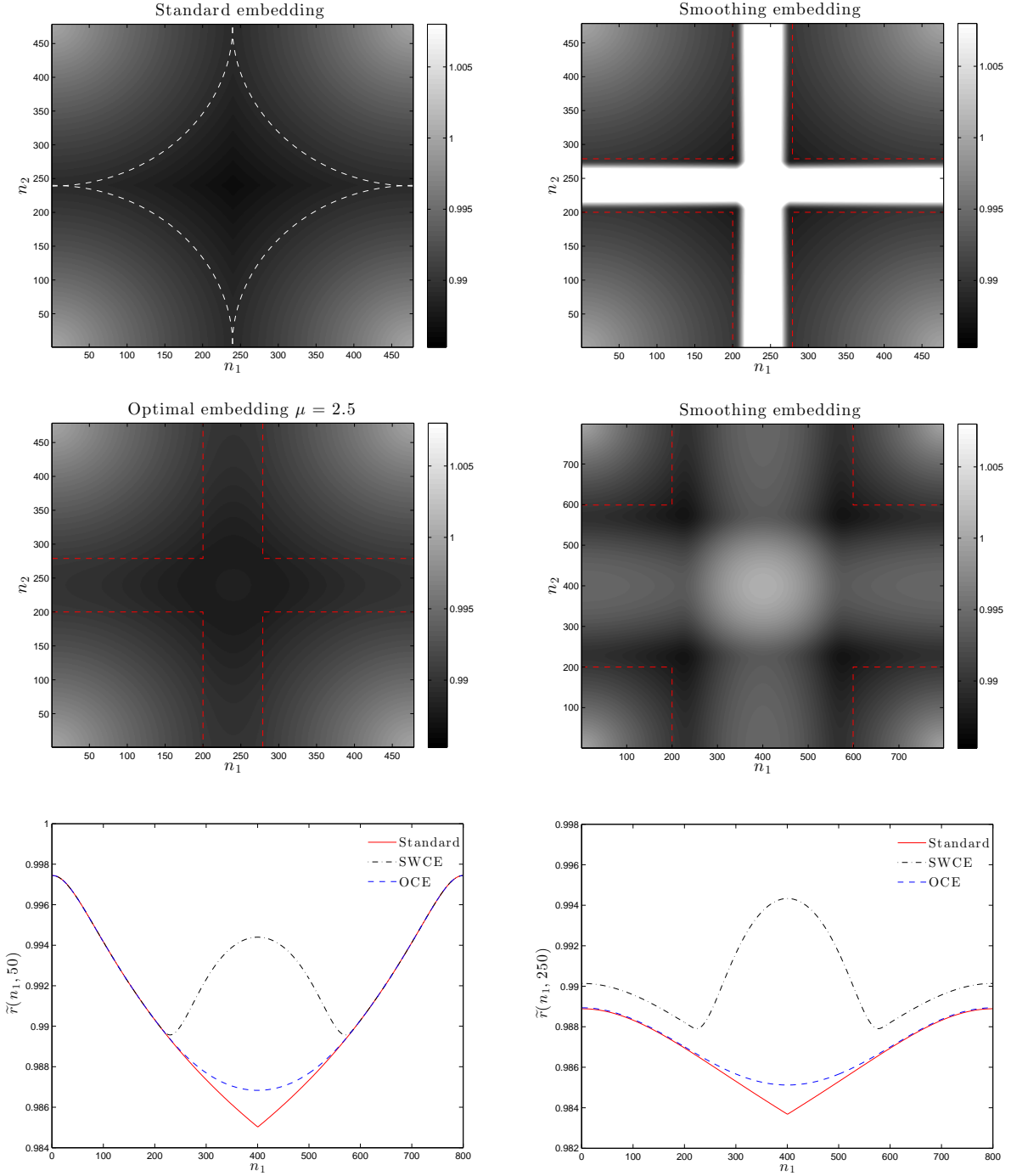
Figure 2: Plots for the anisotropic covariance of the powered exponential form in (4.1) with $W = W_2$ in (4.3) and $N = 100$ and $M = 359$. Top left: Standard embedding. Top right: Smoothing windows embedding with 116 negative eigenvalues. Middle left: Optimal embedding with $\mu = 1.5$, $\min f = 5 \cdot 10^{-6}$, and no negative eigenvalues. Middle right: Optimal embedding with $\mu = 3$, $\min f = 4 \cdot 10^{-6}$, and no negative eigenvalues. Bottom plots: Cross sections of standard CME, SWCE and OCE ($\mu = 1.5$) for $n_2 = 50, 170$.

Figure 3: Plots for the isotropic covariance of the Cauchy form (4.2) with parameters (4.4) $N = 200$. Top left: Standard embedding (2,836 negative eigenvalues). Top right: Smoothing windows embedding with transition region $\widetilde{N} - N = 40$ and 24,720 negative eigenvalues. Middle left: Optimal embedding with $\mu = 2.5$, $\min f = 7 \cdot 10^{-5}$, and no negative eigenvalues. Middle right: Smoothing embedding with transition region $\widetilde{N} - N = 200$ and 136 negative eigenvalues. Bottom plots: Cross sections of standard CME, SWCE and OCE ($\mu = 1.5$) for $n_2 = 50, \ 250$ and $\widetilde{N} - N = 200$.
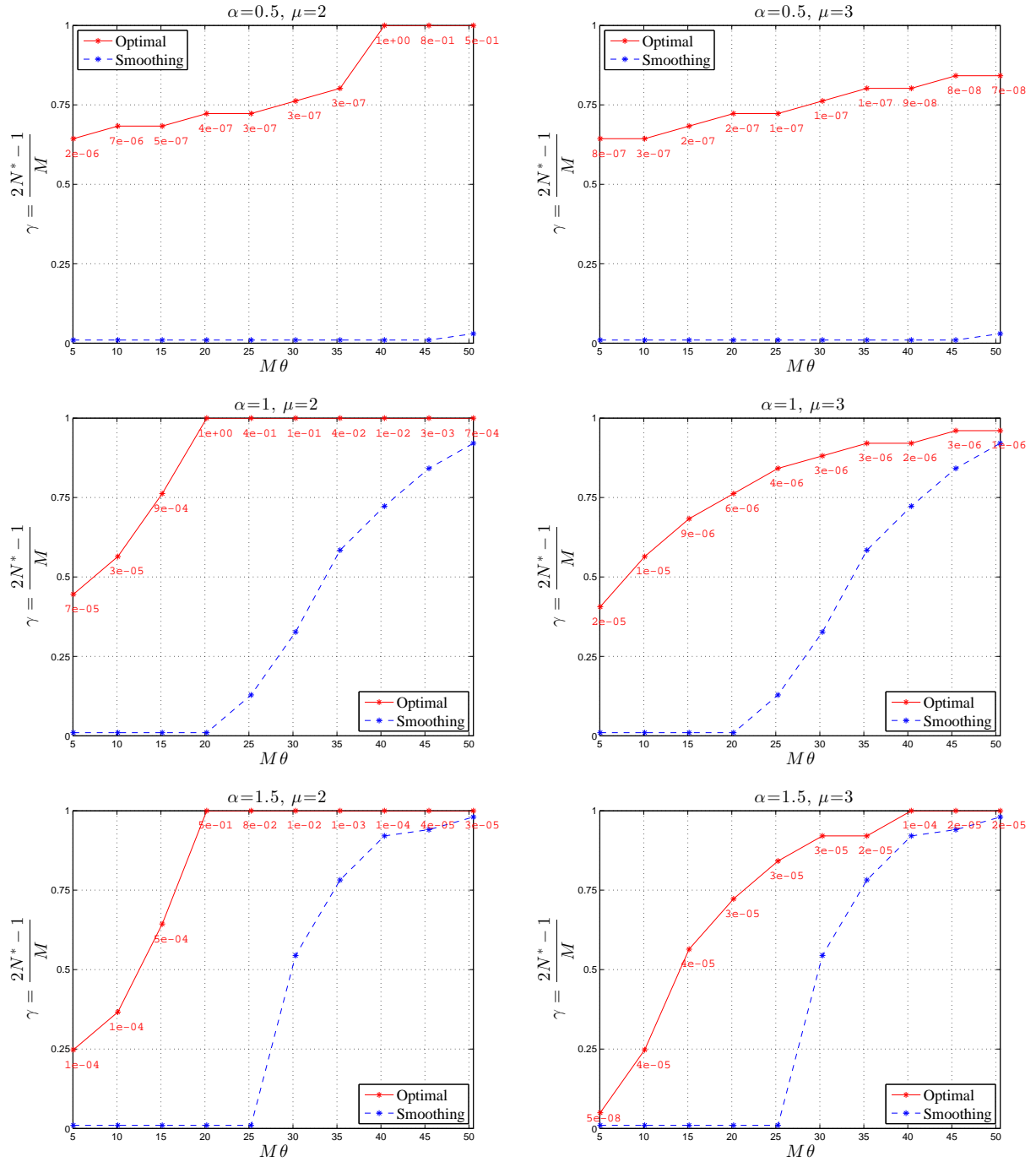
Figure 4: Efficiency values of SWCE (dashed lines) and OCE (solid lines) methods for the powered exponential covariance function (4.1) with $\theta = 0.05k$, $k = 1 \ldots, 10$ and $\alpha = 0.5, 1, 1.5$. Left: OCE implemented with $\mu = 2$. Right: OCE implemented with $\mu = 3$. The tolerance ϵ in the PLB method is equal to 10^{-5} in all cases.
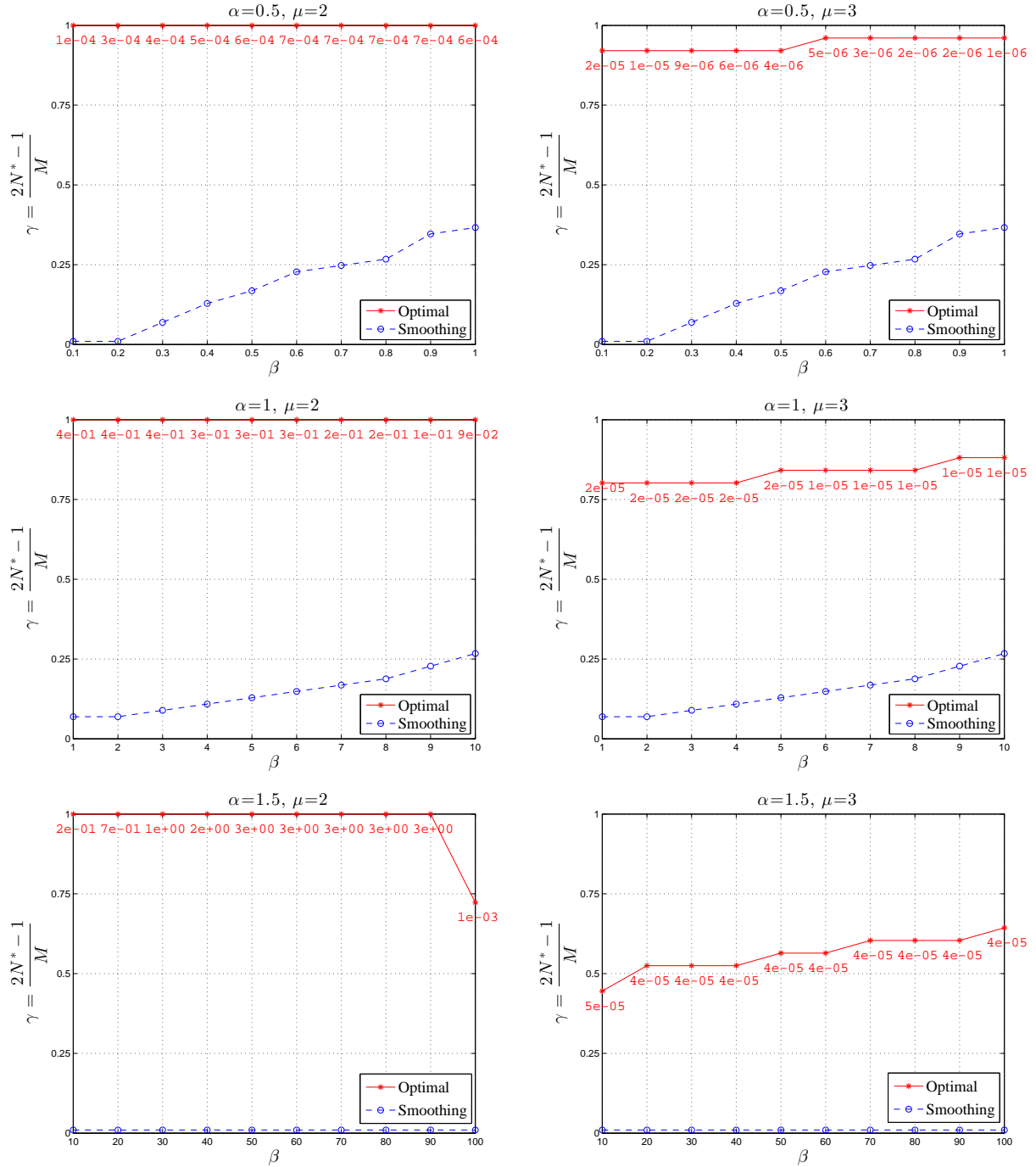
Figure 5: Efficiency values of SWCE (dashed lines) and OCE (solid lines) methods for the Cauchy covariance (4.2) with $\theta = 0.001$ and $\alpha = 0.5, 1, 1.5$. For the first row $\beta = 0.1, 0.2, \ldots, 1$, for the second row $\beta = 1, 2, \ldots, 10$ and for the third row $\beta = 10, 20, \ldots, 100$. Left: OCE implemented with $\mu = 2$. Right: OCE implemented with $\mu = 3$. The tolerance ϵ in the PLB method is equal to 10^{-5} in all cases.

References

Bertsimas, D. & Tsitsiklis, J. N. (1997), *Introduction to Linear Optimization*, Vol. 6, Athena Scientific Belmont, MA.

Boyd, S. & Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press, Cambridge.

Chan, G. & Wood, A. T. (1997), 'An algorithm for simulating stationary Gaussian random fields', *Applied Statistics, Algorithm Section* **46**(1), 171–181.

Chan, G. & Wood, A. T. (1999), 'Simulation of stationary Gaussian vector fields', *Statistics and computing* **9**(4), 265–268.

Craigmile, P. F. (2003), 'Simulating a class of stationary Gaussian processes using the Davies-Harte algorithm, with application to long memory processes', *Journal of Time Series Analysis* **24**(5), 505–511.

Davies, R. B. & Harte, D. S. (1987), 'Tests for Hurst effect', *Biometrika* **74**(1), 95–101.

Dietrich, C. & Newsam, G. (1993), 'A fast and exact method for multidimensional Gaussian stochastic simulations', *Water Resources Research* **29**(8), 2861–2869.

Dietrich, C. R. (1993), 'Computationally efficient Cholesky factorization of a covariance matrix with block Toeplitz structure', *Journal of Statistical Computation and Simulation* **45**(3-4), 203–218.

Forsgren, A., Gill, P. E. & Wright, M. H. (2002), 'Interior methods for nonlinear optimization', *SIAM Review* **44**(4), 525–597 (2003).

Gneiting, T., Ševčíková, H., Percival, D. B., Schlather, M. & Jiang, Y. (2006), 'Fast and exact simulation of large Gaussian lattice systems in $\mathbb{R}^2$: exploring the limits', *Journal of Computational and Graphical Statistics* **15**(3), 483–501.

Helgason, H., Pipiras, V. & Abry, P. (2011), 'Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding', *Signal Processing* **91**(5), 1123–1133.

Helgason, H., Pipiras, V. & Abry, P. (2014), 'Smoothing windows for the synthesis of Gaussian stationary random fields using circulant matrix embedding', *Journal of Computational and Graphical Statistics* **23**(3), 616–635.

Nocedal, J. & Wright, S. J. (2006), *Numerical Optimization*, second edn, Springer, New York.

Percival, D. B. (2006), 'Exact simulation of complex-valued Gaussian stationary processes via circulant embedding', *Signal Processing* **86**(7), 1470–1476.

Stein, M. L. (2002), 'Fast and exact simulation of fractional Brownian surfaces', *Journal of Computational and Graphical Statistics* **11**(3), 587–599.

Stein, M. L. (2012), 'Simulation of Gaussian random fields with one derivative', *Journal of Computational and Graphical Statistics* **21**(1), 155–173.

Wood, A. T. A. & Chan, G. (1994), 'Simulation of stationary Gaussian processes in $[0,1]^{d}$', *Journal of Computational and Graphical Statistics* **3**(4), 409–432.

Hannes Helgason
School of Engineering and Natural Sciences
University of Iceland
Dunhagi 5
107 Reykjavk, Iceland
hannes.helgason@gmail.com

Stefanos Kechagias, Vladas Pipiras
Dept. of Statistics and Operations Research
University of North Carolina at Chapel Hill
CB#3260, Hanes Hall
Chapel Hill, NC 27599, USA
kechagia@email.unc.edu, pipiras@email.unc.edu