View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Chalmers Publication Library

Chalmers Publication Library

Copyright Notice

©2005 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This document was downloaded from Chalmers Publication Library (http://publications.lib.chalmers.se/), where it is available in accordance with the IEEE PSPB Operations Manual, amended 19 Nov. 2010, Sec. 8.1.9 (http://www.ieee.org/documents/opsmanual.pdf)

(Article begins on next page)

Optimal lattices for sampling

Hans R. Künsch, Erik Agrell, and Fred A. Hamprecht

Abstract— **The generalization of the sampling theorem to multidimensional signals is considered, with or without bandwidth constraints. The signal is modeled as a stationary random process and sampled on a lattice. Exact expressions for the mean square error of the best linear interpolator are given in the frequency domain. Moreover, asymptotic expansions are derived for the average mean square error when the sampling rate tends to zero and infinity, respectively. This makes it possible to determine the optimal lattices for sampling. In the low-rate sampling case, or equivalently for rough processes, the optimal lattice is the one which solves the packing problem, whereas in the high-rate sampling case, or equivalently for smooth processes, the optimal lattice is the one which solves the dual packing problem. In addition, the best linear interpolation is compared with ideal low-pass filtering (cardinal interpolation).**

Keywords— **Best linear estimator, cardinal interpolation, lattice theory, multidimensional signal processing, packing problem, sampling theorem.**

I. Introduction

IN CLASSICAL sampling and interpolation theory, the objective is to discretize and store a time signal in such a way that the signal can be estimated as accurately as possible, even at instants for which no sample was stored [21]. The classical method is to sample the signal at regular intervals and to interpolate by summation of shifted and scaled $\sin(x)/x$ functions. The interpolation error of this method is zero if the signal is a realization of a stationary, band-limited stochastic process and the sampling frequency is sufficiently high.

In this presentation, we consider the analogous problem in multidimensional signal processing, where a signal with spatial and/or spectral and/or temporal resolution is to be discretized, stored, and reconstructed. Applications include computer vision and image processing [11,20], remote sensing [12], medical imaging [18], and experimental design [9, 10]. From a geometrical point of view, it is intuitively clear that the multidimensional signal should be sampled as uniformly as possible, in order to gain as much information as possible about the signal everywhere in the relevant region. No part of the region should lie very far from the closest sample point, since this would cause a relatively large uncertainty in the estimate of the signal in that part.

Manuscript submitted Oct. 16, 2003. The material in this paper was presented in Res. Rep. 119, Seminar for Statistics, ETH, Switzerland, 2003, and in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June–July 2003.

Hans R. Künsch is with the Seminar for Statistics, ETH Zentrum, CH–8092 Zürich, Switzerland (email: kuensch@stat.math.ethz.ch).

Erik Agrell is with the Department of Signals and Systems, Chalmers University of Technology, SE–412 96 Göteborg, Sweden (e-mail: agrell@s2.chalmers.se).

Fred A. Hamprecht is with the IWR, INF 368, University of Heidelberg, D–69120 Heidelberg, Germany (e-mail: fred.hamprecht@iwr.uni-heidelberg.de).

The problem of placing points uniformly in a multidimensional space has been studied extensively in other applications and the solution is often to use a *lattice*. Which lattice to use depends on which criterion is used to measure uniformity: the *packing problem* aims at maximizing the distance between the closest pair of lattice points, the *covering problem* aims at minimizing the maximum distance between a (nonlattice) point in space and its closest lattice point, the *quantizer problem* aims at minimizing the moment of inertia of the Voronoi region (defined in the next section), etc. In one dimension, the only lattice (disregarding rescaling) is the set of integers and in two dimensions, the hexagonal lattice is most uniform (according to all common optimality criteria). In higher dimensions, the best known lattices for various criteria are listed in [4] and its references. None of these criteria, however, is immediately applicable to sampling and interpolation.

If each dimension is sampled at regular intervals independently of each other, the resulting multidimensional sampling pattern is the *cubic lattice*. It has been recommended for sampling based on complexity considerations [11], but its performance in terms of estimation error is unfortunately poor. The cubic lattice has the property that it contains quite deep "holes" in between the lattice points, from which the distance to any lattice point is much higher than the corresponding distance in other lattices. Hence, the samples would not support an accurate representation of the signal near such "holes." This undesirable property becomes more prominent with increasing dimension [9].

We assume that the multidimensional signal is a realization of a real stationary stochastic process and that its (multidimensional) covariance function is known. It is not required to be band-limited in any direction. If one has to estimate also the covariance function, then uniform sampling schemes perform poorly [19, Sec. 6.6]. The present article complements earlier efforts that have focused on finding an optimal sampling scheme on a finite domain [13, 14, 17]. Johnson *et al.* considered a different kind of asymptotics [13] and also note the "obvious connection" with lattice theory, without investigating it further. Lim *et al.* introduced numerical procedures [14].

In Section II, we introduce notation, define basic concepts, and summarize Fourier analysis on lattices. In Section III, the best linear estimator for interpolation of the signal is derived and its average error variance is calculated, as a function of the covariance function and the lattice. It is concluded that the optimal lattice type depends on the sampling rate. In Sections IV–V, we show that the best sampling lattice for very low rate is the solution of the packing problem and for very high rate, the dual of the same lattice. Finally, in Section VI we give some numerical examples to illustrate our results. The proofs of all

Theorems are given in the Appendix.

II. Preliminaries

A. Stationary random fields

In the following, a multidimensional stochastic process will be called a "random field." We consider a real zero-mean stationary random field $(Z(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{R}^d)$ with finite second moments and denote its covariance function by

$$R(\boldsymbol{x}) = \mathbb{E}[Z(\boldsymbol{y})Z(\boldsymbol{y}+\boldsymbol{x})].$$

If Z is mean square continuous, then, by Bochner's theorem [6, Ch. 4, Sec. 2, Theorem 2], R is the Fourier transform of a finite, positive measure, the spectral measure. In addition, we assume that this spectral measure has a density S. This means that

$$R(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S(\boldsymbol{\omega}) \exp(i\boldsymbol{\omega}^T\boldsymbol{x}) d\boldsymbol{\omega}. \tag{1}$$

A sufficient condition for this is

$$\int_{\mathbb{R}^d} |R(\boldsymbol{x})| d\boldsymbol{x} < \infty, \tag{2}$$

and then the spectral density can be obtained as

$$S(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} R(\boldsymbol{x}) \exp(-i\boldsymbol{\omega}^T\boldsymbol{x}) d\boldsymbol{x}.$$

For estimation problems, it is useful to introduce the Hilbert space $H_Z(D)$ spanned by the random variables $(Z(\boldsymbol{x}); \boldsymbol{x} \in D)$ for any $D \subseteq \mathbb{R}^d$, with inner product $\langle Z(\boldsymbol{x}), Z(\boldsymbol{y})\rangle = R(\boldsymbol{x}-\boldsymbol{y})$, see [2, Ch. 2]. It consists of all finite linear combinations $\sum_{i=1}^n a_i Z(\boldsymbol{x}_i)$ with $\boldsymbol{x}_i \in D$ and all limits (with respect to the inner product) of such linear combinations. Many calculations are simplified if we pass from $H_Z(D)$ to an equivalent Hilbert space of functions $\psi \in L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$ that satisfy $\psi(-\boldsymbol{\omega}) = \overline{\psi(\boldsymbol{\omega})}$. Here overline denotes complex conjugation and $L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$ is the space of complex functions on $\mathbb{R}^d$ with the inner product

$$\langle \psi, \phi\rangle = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \psi(\boldsymbol{\omega})\overline{\phi(\boldsymbol{\omega})} S(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{3}$$

From (1), it is clear that the operator T which maps the random variables $Z(\boldsymbol{x})$ to the functions $\exp(i\boldsymbol{x}^T\boldsymbol{\omega})$ preserves inner products. By linearity and completion, this operator can be extended to the whole space $H_Z(\mathbb{R}^d)$ while still preserving inner products. The image of $H_Z(\mathbb{R}^d)$ under T is therefore a subspace of $L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$. By standard results in Fourier theory, this subspace is in fact the whole space $\{\psi \in L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega});\ \psi(-\boldsymbol{\omega}) = \overline{\psi(\boldsymbol{\omega})}\}$. This means that to any such function ψ there corresponds a random variable in $H_Z(\mathbb{R}^d)$ and vice versa, such that all inner products are preserved. For a proof of these basic facts, see, e.g., [6, Ch. 4, Sec. 5, esp. Theorem 3].

The best linear interpolator of $Z(\boldsymbol{x})$ based on observed values of $(Z(\boldsymbol{x}'); \boldsymbol{x}' \in D)$ is defined as the unique element $\widehat{Z}(\boldsymbol{x}) \in H_Z(D)$ such that

$$\mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2] = \inf_{Y \in H_Z(D)} \mathbb{E}[(Z(\boldsymbol{x}) - Y)^2].$$

In some disciplines, $\widehat{Z}(\boldsymbol{x})$ is called the "kriging" estimator of $Z(\boldsymbol{x})$ [9]. By Hilbert space geometry, $\widehat{Z}(\boldsymbol{x})$ is the orthogonal projection of $Z(\boldsymbol{x})$ onto $H_Z(D)$. We can compute $\widehat{Z}(\boldsymbol{x})$ by computing first the orthogonal projection of $\exp(i\boldsymbol{x}^T\boldsymbol{\omega})$ onto the subspace spanned by $(\exp(i\boldsymbol{x}'^T\boldsymbol{\omega});\ \boldsymbol{x}' \in D)$ and then applying the inverse of T.

B. Lattices

The standard reference for lattice theory is the book by Conway and Sloane [4]. A d-dimensional lattice $\Lambda(\boldsymbol{B})$ is a countably infinite subset of $\mathbb{R}^d$ of the form $\{\boldsymbol{u} = \boldsymbol{B}^T\boldsymbol{w};\ \boldsymbol{w} \in \mathbb{Z}^n\}$ where the so-called *generator matrix* $\boldsymbol{B}$ is an $n \times d$ matrix with linearly independent rows. This means that the lattice consists of all integer linear combinations of the row vectors of $\boldsymbol{B}$. $\boldsymbol{B}$ is often square, but in some cases a representation with $n < d$ may be preferable. For $d > 1$, the generator matrix is not unique. For instance, two possible generator matrices for the hexagonal lattice in $d = 2$ dimensions are

$$\boldsymbol{B}_1 = \begin{bmatrix} 2 & 0 \\ 1 & \sqrt{3} \end{bmatrix}, \quad \boldsymbol{B}_2 = \begin{bmatrix} 1 & -\sqrt{3} \\ 1 & \sqrt{3} \end{bmatrix}.$$

In a three-dimensional coordinate system, a rescaled version of the same lattice may be represented without square roots, as with

$$\boldsymbol{B}_3 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

The Voronoi region of a lattice point is the set of all vectors in $\mathbb{R}^d$ that are at least as close to this point as to any other lattice point:

$$\Omega(\boldsymbol{B}, \boldsymbol{u}) \stackrel{\text{def}}{=} \left\{\boldsymbol{x} \in \mathbb{R}^d;\ \|\boldsymbol{x}-\boldsymbol{u}\| \le \|\boldsymbol{x}-\boldsymbol{u}'\| \quad \forall \boldsymbol{u}' \in \Lambda(\boldsymbol{B})\right\}$$

It is easy to see that all Voronoi regions are translations of $\Omega(\boldsymbol{B}) \stackrel{\text{def}}{=} \Omega(\boldsymbol{B}, \boldsymbol{0})$ and that they are convex polytopes that tile the space $\mathbb{R}^d$ (modulo the overlap at the boundaries).

In the frequency domain, an important role is played by the dual lattice of $\Lambda(\boldsymbol{B})$, scaled by 2π. It consists of all points $\boldsymbol{\lambda} \in \mathbb{R}^d$ such that $\boldsymbol{\lambda}^T\boldsymbol{u}$ is an integer multiple of 2π for any $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$. A possible choice of the generator matrix $\boldsymbol{A}$ for the dual lattice is, if $\boldsymbol{B}$ is square, $\boldsymbol{A} = 2\pi(\boldsymbol{B}^{-T})$. We will always use the notation $\boldsymbol{B}$ and $\boldsymbol{A}$ for the generators of two dual lattices scaled by 2π. If $\boldsymbol{u}$ is in $\Lambda(\boldsymbol{B})$, then the function $\boldsymbol{\omega} \to \exp(i\boldsymbol{u}^T\boldsymbol{\omega})$ is periodic with periods $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$:

$$\exp(i\boldsymbol{u}^T(\boldsymbol{\omega}+\boldsymbol{\lambda})) = \exp(i\boldsymbol{u}^T\boldsymbol{\omega})\exp(i\boldsymbol{u}^T\boldsymbol{\lambda}) = \exp(i\boldsymbol{u}^T\boldsymbol{\omega})$$

because $\boldsymbol{u}^T\boldsymbol{\lambda}$ is an integer multiple of 2π. Moreover, these functions are orthonormal in $L_2(\Omega(\boldsymbol{A}), d\boldsymbol{\omega}/\mathrm{vol}(\Omega(\boldsymbol{A})))$ by the following lemma which is proved in the Appendix.

Lemma 1: If $\boldsymbol{B}$ and $\boldsymbol{A}$ are the generators of two lattices that are dual to each other up to a scaling by 2π, then

$$\frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) d\boldsymbol{\omega} = \begin{cases} 1 & (\boldsymbol{u} = \boldsymbol{0}) \\ 0 & (\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}). \end{cases}$$

Finally, like in the case of the cubic lattice, it can be shown that the functions $(\exp(i\boldsymbol{u}^T\boldsymbol{\omega});\ \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$ for any

lattice form an orthonormal basis of the space of periodic functions with periods in $\Lambda(\boldsymbol{A})$, that is, they are complete. In particular, any periodic integrable function g of $\boldsymbol{\omega}$ whose periods belong to a lattice $\Lambda(\boldsymbol{A})$ can be represented as a linear combination (Fourier series) of the functions $(\exp(i\boldsymbol{u}^T\boldsymbol{\omega});\, \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$:

$$g(\boldsymbol{\omega}) = \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} c(\boldsymbol{u}) \exp(i\boldsymbol{u}^T\boldsymbol{\omega}). \tag{4}$$

The sum converges in $L_2(\Omega(\boldsymbol{A}), \mathrm{vol}(\Omega(\boldsymbol{A}))^{-1}d\boldsymbol{\omega})$, and the coefficients are given by

$$c(\boldsymbol{u}) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} g(\boldsymbol{\omega}) \exp(-i\boldsymbol{u}^T\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{5}$$

III. The interpolation error

Let $\widehat{Z}(\boldsymbol{x})$ be the best linear interpolator of $Z(\boldsymbol{x})$ based on observations $(Z(\boldsymbol{u});\, \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$ on a lattice. In this section, we develop some expressions and bounds for the mean square error for a fixed $\boldsymbol{x}$

$$\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) \stackrel{\mathrm{def}}{=} \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2]$$

and the average mean square error over $\boldsymbol{x} \in \Omega(\boldsymbol{B})$

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \stackrel{\mathrm{def}}{=} \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \int_{\Omega(\boldsymbol{B})} \sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) d\boldsymbol{x}.$$

Some of the results (equations (8), (11), and (12) below) were given already by Petersen and Middleton [16, Sec. VI]. Since they form the basis for our main results and our arguments are different, we give proofs here.

A. Expressions for the best linear interpolator

We use the isometry between Hilbert spaces discussed in Section II and give in the following theorem the element $\psi_{\boldsymbol{x}}$ in $L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$ that corresponds to $\widehat{Z}(\boldsymbol{x})$. This generalizes a result by Stein [19, pp. 98–99] for cubic lattices.

Proposition 1: Under the isometry $Z(\boldsymbol{x}) \leftrightarrow \exp(i\boldsymbol{\omega}^T\boldsymbol{x})$ between H_Z and $L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$, the best linear estimator $\widehat{Z}(\boldsymbol{x})$ based on observations $(Z(\boldsymbol{u});\, \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$ corresponds to the function

$$\psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \frac{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(i\boldsymbol{x}^T(\boldsymbol{\omega}+\boldsymbol{\lambda}))S(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})}. \tag{6}$$

The proof is given in the Appendix.

Under additional assumptions, we can obtain a more explicit representation of $\widehat{Z}(\boldsymbol{x})$ in the space domain. The function $\psi_{\boldsymbol{x}}$ is periodic with period belonging to $\Lambda(\boldsymbol{A})$ and can thus be expanded into a Fourier series, cf. (4). Moreover, it is easily seen that the Fourier coefficients (5) of $\psi_{\boldsymbol{x}}$ are of the form $c(\boldsymbol{x}-\boldsymbol{u})$ where

$$c(\boldsymbol{x}-\boldsymbol{u}) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \cdot \int_{\Omega(\boldsymbol{A})} \frac{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(i(\boldsymbol{x}-\boldsymbol{u})^T(\boldsymbol{\omega}+\boldsymbol{\lambda}))S(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})} d\boldsymbol{\omega}.$$

Introducing

$$S^*(\boldsymbol{\omega}) = \frac{S(\boldsymbol{\omega})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})} \tag{7}$$

and using the periodicity of the denominator, we can also write

$$c(\boldsymbol{x}-\boldsymbol{u}) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\mathbb{R}^d} \exp(i(\boldsymbol{x}-\boldsymbol{u})^T\boldsymbol{\omega}) S^*(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{8}$$

Since $\exp(i\boldsymbol{u}^T\boldsymbol{\omega})$ corresponds to $Z(\boldsymbol{u})$ under the isometry between H_Z and $L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$, one expects from the Fourier series

$$\psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \sum_{u\in\Lambda(\boldsymbol{B})} c(\boldsymbol{x}-\boldsymbol{u})\, \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \tag{9}$$

that also

$$\widehat{Z}(\boldsymbol{x}) = \sum_{u\in\Lambda(\boldsymbol{B})} c(\boldsymbol{x}-\boldsymbol{u})\, Z(\boldsymbol{u}). \tag{10}$$

The weight function $c(\boldsymbol{x})$ is sometimes called an "interpolation function", especially in the spline community [21].

However, (9) converges in $L_2(\Omega(\boldsymbol{A}), d\boldsymbol{\omega})$ and not necessarily in $L_2(\mathbb{R}^d, (2\pi)^{-d}S(\boldsymbol{\omega})d\boldsymbol{\omega})$. A sufficient condition for this to hold is for instance that $\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})$ is bounded. The difference between the two L_2-spaces also shows up in cases where the set $\{\boldsymbol{\omega} \in \Omega(\boldsymbol{A});\, \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda}) = 0\}$ is not empty. For Proposition 1, it is irrelevant how we define $\psi_{\boldsymbol{x}}$ on this set. However, for the Fourier coefficients in (8) this can make a difference: These coefficients and the representation (10) are then not unique.

B. Expressions for the mean square error

As a consequence of Proposition 1, we obtain in the next theorem several equivalent expressions for the mean square error. For numerical evaluation or asymptotic analysis, one can choose whichever is most convenient in a given situation. In order to state the result, we introduce a short notation for the continuous convolution of the covariance function

$$R^{*2}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \int_{\mathbb{R}^d} R(\boldsymbol{y})R(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} = \int_{\mathbb{R}^d} R(\boldsymbol{y})R(\boldsymbol{x}+\boldsymbol{y})d\boldsymbol{y}.$$

Theorem 1: The following expressions hold for the mean square error of the best linear interpolator:

$$\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S^*(\boldsymbol{\omega}) \cdot \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} (1 - \exp(i\boldsymbol{x}^T\boldsymbol{\lambda}))S(\boldsymbol{\omega}+\boldsymbol{\lambda})d\boldsymbol{\omega}, \tag{11}$$

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S(\boldsymbol{\omega})(1 - S^*(\boldsymbol{\omega}))d\boldsymbol{\omega} \tag{12}$$

$$= R(\boldsymbol{0}) - \frac{1}{(2\pi)^d} \int_{\Omega(\boldsymbol{A})} \frac{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S^2(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})} d\boldsymbol{\omega} \tag{13}$$

$$= R(\boldsymbol{0}) - \frac{1}{(2\pi)^d} \int_{\Omega(\boldsymbol{A})} \frac{\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega})R^{*2}(\boldsymbol{u})}{\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega})R(\boldsymbol{u})} d\boldsymbol{\omega}. \tag{14}$$

Here S^* is defined in (7) and a value "0/0" should be interpreted as zero. In particular, we have

$$\sigma^2(\text{ave}, \Lambda(\boldsymbol{B})) \le \sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) \le 2\,\sigma^2(\text{ave}, \Lambda(\boldsymbol{B})). \tag{15}$$

The proof is given in the Appendix. Eq. (12) is similar to a general one-dimensional sampling error formula [1].

From Theorem 1, we can deduce following bounds for the average mean square error which are new to our knowledge. Again, the proof is given in the Appendix.

Proposition 2: For any spectral density we have

$$\sigma^2(\text{ave}, \Lambda(\boldsymbol{B})) \le 2\,\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus \Omega(\boldsymbol{A})} S(\boldsymbol{\omega}) d\boldsymbol{\omega}.$$

If the spectral density is isotropic and decreasing in $\|\boldsymbol{\omega}\|$, then in addition

$$\sigma^2(\text{ave}, \Lambda(\boldsymbol{B})) \ge \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus \Omega(\boldsymbol{A})} S(\boldsymbol{\omega}) d\boldsymbol{\omega}.$$

As a simple example for Theorem 1 we consider a spectral density S which is constant on a region $D \subset \mathbb{R}^d$ and zero outside of D. Then the integrand on the right hand side of (13) is equal to the nonzero value of S if $\boldsymbol{\omega} + \boldsymbol{\lambda} \in D$ for some $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$ and zero otherwise. This implies that

$$\frac{\sigma^2(\text{ave}, \Lambda(\boldsymbol{B}))}{R(\boldsymbol{0})} = 1 - \frac{\text{vol}(\Omega_D(\boldsymbol{A}))}{\text{vol}(D)}$$

where

$$\Omega_D(\boldsymbol{A}) \stackrel{\text{def}}{=} \{\boldsymbol{\omega} \in \Omega(\boldsymbol{A});\, (\boldsymbol{\omega} + \boldsymbol{\lambda}) \in D \text{ for some } \boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})\}.$$

C. Cardinal interpolation

For a similar example, we look at the case where S is zero outside of $\Omega(\boldsymbol{A})$. Then by Proposition 2, $\sigma^2(\text{ave}, \Lambda(\boldsymbol{B}))$ is zero, and therefore $\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B}))$ is also zero for any $\boldsymbol{x}$ (this can also be seen directly from (11)). In other words, we can recover all values $Z(\boldsymbol{x})$ without error from the values of Z on the lattice $\Lambda(\boldsymbol{B})$. This is the well-known spatial version of Nyquist's sampling theorem [15] due to Peterson and Middleton [16]. Moreover, we can compute the coefficients (8) explicitly. By Proposition 1, the function $\psi_{\boldsymbol{x}}$ corresponding to $\widehat{Z}(\boldsymbol{x})$ (or more precisely, one possible choice of this function) is

$$\psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \exp(i\boldsymbol{x}^T(\boldsymbol{\omega} \bmod \Lambda(\boldsymbol{A})))$$

where we define $\boldsymbol{\omega} \bmod \Lambda(\boldsymbol{A})$ to be $\boldsymbol{\omega} - \boldsymbol{\lambda}$ with $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$ such that $\boldsymbol{\omega} - \boldsymbol{\lambda} \in \Omega(\boldsymbol{A})$. Thus the coefficients (8) in the representation (10) are

$$c(\boldsymbol{x} - \boldsymbol{u}) = \frac{1}{\text{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \exp(i(\boldsymbol{x} - \boldsymbol{u})^T \boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{16}$$

independently of S, which corresponds to ideal low-pass filtering. For the cubic lattice, $c(\boldsymbol{x} - \boldsymbol{u})$ is of course the product of the well known $\sin(x)/x$ functions. We call interpolation with these coefficients *cardinal interpolation.*

Often, cardinal interpolation is applied even for random fields whose spectra do not vanish outside $\Omega(\boldsymbol{A})$. The advantage is that in contrast to the best linear estimator, it does not require the knowledge (or estimation) of the covariance function or the spectrum. The disadvantages are that its coefficients decay slowly and that it is less precise than the best linear estimator. Denote by $\sigma_c^2(\boldsymbol{x}, \Lambda(\boldsymbol{B}))$ the mean square error for cardinal interpolation. Then we have for a general spectral density (which need not vanish outside $\Omega(\boldsymbol{A})$)

$$\begin{aligned}
\sigma_c^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) &= \frac{1}{(2\pi)^d} \\
&\quad \cdot \int_{\mathbb{R}^d} |\exp(i\boldsymbol{x}^T\boldsymbol{\omega}) - \exp(i\boldsymbol{x}^T(\boldsymbol{\omega} \bmod \Lambda(\boldsymbol{A})))|^2 S(\boldsymbol{\omega}) d\boldsymbol{\omega} \\
&= \frac{1}{(2\pi)^d} \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} \\
&\quad \int_{\Omega(\boldsymbol{A})} |\exp(i\boldsymbol{x}^T(\boldsymbol{\omega} + \boldsymbol{\lambda})) - \exp(i\boldsymbol{x}^T\boldsymbol{\omega})|^2 S(\boldsymbol{\omega} + \boldsymbol{\lambda}) d\boldsymbol{\omega} \\
&= \frac{1}{(2\pi)^d} \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} |1 - \exp(i\boldsymbol{x}^T\boldsymbol{\lambda})|^2 \int_{\Omega(\boldsymbol{A})} S(\boldsymbol{\omega} + \boldsymbol{\lambda}) d\boldsymbol{\omega}.
\end{aligned}$$

In one dimension, these expressions have been derived by Brown [3]. Because $|1 - \exp(ix)|^2 = 2(1 - \cos(x))$, by Lemma 1 the average interpolation error with cardinal interpolation is

$$\sigma_c^2(\text{ave}, \Lambda(\boldsymbol{B})) = 2\,\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus \Omega(\boldsymbol{A})} S(\boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{17}$$

which is the upper bound from Proposition 2. By the lower bound of the same theorem, for isotropic and decreasing spectral densities, the average mean square error with cardinal interpolation is larger by at most a factor of two compared with the optimal interpolation.

D. Average versus worst case error

In the next two sections we determine the lattice that minimizes $\sigma^2(\text{ave}, \Lambda(\boldsymbol{B}))$ among all lattices with equal volume $\text{vol}(\Omega(\boldsymbol{B}))$. Note that $1/\text{vol}(\Omega(\boldsymbol{B}))$ is the sampling rate, that is, the limit of the number of points in $\Lambda(\boldsymbol{B}) \cap D$ divided by $\text{vol}(D)$ as the domain D is extended in all directions. We are not able to solve this problem in full generality, but we will derive the solution for the two limiting cases where the sampling rate tends to zero and to infinity respectively for certain classes of random fields.

Alternatively, we could try to minimize the worst case mean square error $\sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B}))$, but this is an even more difficult problem. Note, however, that if Λ_0 minimizes the average mean square error, then by (15) for any other lattice Λ with the same sampling rate

$$\sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda_0) \le 2 \sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda).$$

Hence if we choose the lattice with minimal average mean square error, the loss we will incur with respect to worst case mean square error is bounded.

IV. The optimal lattice for low-rate sampling

In this section, we study the case where the sampling rate tends to zero. More precisely, we look at the behavior of $\sigma^2(\mathrm{ave}, \Lambda(\beta\boldsymbol{B}))$ as β tends to infinity for a fixed lattice $\Lambda(\boldsymbol{B})$ and a fixed covariance $R_0(\boldsymbol{x})$. The sampling rate is then $1/(\beta^d \mathrm{vol}(\Omega(\boldsymbol{B})))$. Instead of rescaling the lattice, we can equivalently rescale the covariance function, that is, we will consider $\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))$ for covariance functions of the form $R(\boldsymbol{x}) = R_0(\beta\boldsymbol{x})$ with β increasing to infinity. For large β, the dependence between any two observed values is small and thus the sampled realizations of Z look rough.

A. An exact expression

Without loss of generality, we will take $R(\boldsymbol{0}) = 1$. Moreover we assume that β is large enough so that

$$\epsilon \stackrel{\mathrm{def}}{=} \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} |R(\boldsymbol{u})| < 1. \tag{18}$$

Let us introduce the notation

$$\Delta(\boldsymbol{u}) \stackrel{\mathrm{def}}{=} \begin{cases} 0 & (\boldsymbol{u} = \boldsymbol{0}) \\ -R(\boldsymbol{u}) & (\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}) \end{cases}$$

By expanding the inverse of the denominator of the integrand in (14) into a Taylor series, we can then write

$$\left(\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) R(\boldsymbol{u}) \right)^{-1} = \sum_{k=0}^{\infty} \left(\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \Delta(\boldsymbol{u}) \right)^k. \tag{19}$$

This leads to another exact expression for $\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))$. In order to formulate it, we denote by $\Delta^k(\boldsymbol{u})$ the k-fold discrete convolution of Δ: We set $\Delta^1(\boldsymbol{u}) = \Delta(\boldsymbol{u})$ and for $k > 1$ we define recursively

$$\Delta^k(\boldsymbol{u}) \stackrel{\mathrm{def}}{=} \sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})} \Delta^{k-1}(\boldsymbol{u} - \boldsymbol{u}') \Delta(\boldsymbol{u}') \tag{20}$$

By assumption (18) and an induction argument, we see that

$$\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} |\Delta^k(\boldsymbol{u})| \le \epsilon^k. \tag{21}$$

In particular, Δ^k is well defined. Now we can state the following result which is proved in the Appendix.

Proposition 3: If $R(\boldsymbol{0}) = 1$ and assumption (18) holds,

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) = 1 - \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \left(R^{*2}(\boldsymbol{0}) + \sum_{k=1}^{\infty} \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \Delta^k(\boldsymbol{u}) R^{*2}(\boldsymbol{u}) \right). \tag{22}$$

B. Asymptotic approximations

From Proposition 3 we now derive a series of approximations of $\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))$ in equations (23), (24), (27), (28) and (29) below. Heuristically, the contribution of the terms in the expression (22) becomes smaller as k increases, cf. (21). If we ignore all terms with $k \ge 1$, then we obtain an approximation that is independent of the lattice:

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\boldsymbol{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))}. \tag{23}$$

In order to compare different lattices, we thus have to look at the next order approximation. In addition to the terms with $k = 1$ we need to include also the term with $k = 2$ and $\boldsymbol{u} = \boldsymbol{0}$, because in the sparse sampling case $R^{*2}(\boldsymbol{0})$ is much larger than $R^{*2}(\boldsymbol{u})$ for $\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}$. This gives

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\boldsymbol{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \left(1 + \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} R^2(\boldsymbol{u}) - \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} R(\boldsymbol{u}) \frac{R^{*2}(\boldsymbol{u})}{R^{*2}(\boldsymbol{0})} \right). \tag{24}$$

By (21) and the Cauchy-Schwarz inequality for $R^{*2}(\boldsymbol{u})$, the error in this approximation can be bounded by

$$\frac{R^{*2}(\boldsymbol{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \left(\epsilon^2 \max_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} \frac{R^{*2}(\boldsymbol{u})}{R^{*2}(\boldsymbol{0})} + \frac{\epsilon^3}{1-\epsilon} \right). \tag{25}$$

The approximation (24) is however too complicated for optimization. Hence we need to simplify it, using additional assumptions. If we assume R to be isotropic and monotonically decreasing with distance, the largest summands are those where $\|\boldsymbol{u}\|$ is minimal, and if R decays quickly these largest summands dominate the sum of all other terms. The following theorem contains a precise statement. We denote by $\rho = \rho(\boldsymbol{B})$ half the minimum distance between two points of the lattice, i.e., the *packing radius,* by $\tau = \tau(\boldsymbol{B})$ the number of lattice points at distance 2ρ from the origin, i.e., the *kissing number,* and by $\boldsymbol{e}$ an arbitrary unit vector.

Theorem 2: Consider a sequence of isotropic covariance functions $R(\boldsymbol{x}) = R_0(\beta\|\boldsymbol{x}\|)$ depending on a parameter $\beta \ge 1$ and assume that R_0 is monotonically decreasing and satisfies

$$\frac{R_0(r)}{C\exp(-r^p)} \to 1 \quad (r \to \infty) \tag{26}$$

for a constant $0 < C < \infty$ and $p > 0$. Then up to an error that is of lower order as $\beta \to \infty$,

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\boldsymbol{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \cdot \left(1 + \tau R^2(2\rho\boldsymbol{e}) - \tau R(2\rho\boldsymbol{e}) \frac{R^{*2}(2\rho\boldsymbol{e})}{R^{*2}(\boldsymbol{0})} \right). \tag{27}$$

Moreover, for $p \geq 1$,

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\mathbf{0}) - \tau R(2\rho \boldsymbol{e}) R^{*2}(2\rho \boldsymbol{e})}{\mathrm{vol}(\Omega(\boldsymbol{B}))}, \quad (28)$$

and for $p < 1$

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\mathbf{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} + \frac{\tau R^2(2\rho \boldsymbol{e})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \cdot \left(2 \int_{\mathbb{R}^d} R(\boldsymbol{x}) d\boldsymbol{x} - R^{*2}(\mathbf{0}) \right), \quad (29)$$

where in both cases the error is of lower order as $\beta \to \infty$.

As a corollary, the lattice $\Lambda(\boldsymbol{B})$ minimizing $\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))$ for given $\mathrm{vol}(\Omega(\boldsymbol{B}))$ is in all cases considered the one maximizing the packing radius $\rho(\boldsymbol{B})$. For $p \geq 1$, this follows from the monotonicity of R^{*2} and for $p < 1$ it follows because

$$2 \int_{\mathbb{R}^d} R(\boldsymbol{x}) d\boldsymbol{x} - R^{*2}(\mathbf{0}) = \beta^{-d} \int_{\mathbb{R}^d} (2R_0(\|\boldsymbol{x}\|) - R_0^2(\|\boldsymbol{x}\|)) d\boldsymbol{x}$$

and the last integral is strictly positive. When there are several lattices with the same maximal packing radius, we should take the one with minimal value of $\tau(\boldsymbol{B})$.

V. The optimal lattice for high-rate sampling

In this section, we study the case where the sampling rate tends to infinity, that is the behavior of $\sigma^2(\mathrm{ave}, \Lambda(\beta \boldsymbol{B}))$ as β tends to zero. This means that the dependence between sampled values is strong and thus the sampled realizations of Z look smooth. As in the previous section, we will fix the lattice and rescale the covariance function as $R(\boldsymbol{x}) = R_0(\beta \boldsymbol{x})$. Equivalently, in the frequency domain the spectral density function takes the form $S(\boldsymbol{\omega}) = \beta^{-d} S_0(\|\boldsymbol{\omega}\|/\beta)$ where S_0 and R_0 are a Fourier pair. In order to simplify notation, we will use the parameter $\alpha = 1/\beta$ which tends to infinity.

A. Optimal interpolation

We rewrite (12) in the equivalent form

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (30)$$

where

$$r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) = \frac{S(\boldsymbol{\omega}) \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\mathbf{0}\}} S(\boldsymbol{\omega} + \boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} S(\boldsymbol{\omega} + \boldsymbol{\lambda})}. \quad (31)$$

We first explain our approximation heuristically. Because the square has maximal area among all rectangles with common perimeter, it is intuitively clear that $r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega})$ is maximal if the two factors in the numerator are equal. If S_0 is monotonic and decreases quickly, then for α tending to infinity, the infinite sum $\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\mathbf{0}\}} S(\boldsymbol{\omega} + \boldsymbol{\lambda})$ is approximately equal to the largest summand which is the one where $\|\boldsymbol{\omega} + \boldsymbol{\lambda}\|$ is minimal. Together, this implies that $r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega})$ is maximal for $\boldsymbol{\omega} = \hat{\boldsymbol{\lambda}}/2$ for any $\hat{\boldsymbol{\lambda}}$ that belongs to the set $\Psi(\boldsymbol{A})$ of shortest nonzero vectors in $\Lambda(\boldsymbol{A})$. Moreover, near such a point $r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega})$ can be approximated as

$$r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) \approx q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega}) \stackrel{\mathrm{def}}{=} \frac{S(\boldsymbol{\omega}) S(\hat{\boldsymbol{\lambda}} - \boldsymbol{\omega})}{S(\boldsymbol{\omega}) + S(\hat{\boldsymbol{\lambda}} - \boldsymbol{\omega})}$$

and the contribution from other points to the integral is negligible. This suggests that

$$\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \sum_{\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega}) d\boldsymbol{\omega} = \frac{\tau}{(2\pi)^d} \int_{\mathbb{R}^d} q(2\rho \boldsymbol{e}, \boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (32)$$

where $\rho = \rho(\boldsymbol{A})$ and $\tau = \tau(\boldsymbol{A})$ are the packing radius and the kissing number respectively of the dual lattice, and $\boldsymbol{e}$ is an arbitrary unit vector in $\mathbb{R}^d$. The integral on the right hand side of (32) depends essentially only on the values of $S(\boldsymbol{\omega})$ for $\boldsymbol{\omega}$ near $\rho \boldsymbol{e}$. Since we assume S to be isotropic and monotonic, $\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))$ will be minimal if ρ is maximal, that is, the optimal lattice $\boldsymbol{B}$ for high-rate sampling is the dual of the one solving the packing problem.

We now state a rigorous result which is proved in the Appendix.

Theorem 3: Consider a sequence of isotropic spectral density functions $S(\boldsymbol{\omega}) = \alpha^d S_0(\alpha \|\boldsymbol{\omega}\|)$ depending on a parameter $\alpha \geq 1$ and assume that for some $p > 0$ and some $0 < C < \infty$

$$\frac{S_0(r)}{C \exp(-r^p)} \to 1 \quad (r \to \infty). \quad (33)$$

Then for $\alpha \to \infty$, the error in the approximation (32) is of lower order and

$$\frac{\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))}{\exp(-(\alpha\rho)^p)(\alpha\rho)^{d - p(d+1)/2}} \to \tau(\boldsymbol{A}) \frac{C\pi}{2} (2\pi p)^{-(d+1)/2}.$$

B. Cardinal interpolation

It is interesting to compare cardinal interpolation (ideal low-pass filtering) with optimal interpolation in the high-rate sampling case. If the sampling rate goes to infinity, the mean square error of cardinal interpolation also converges to zero, and one might conjecture that in this situation the two interpolation methods are actually equivalent, meaning that the ratio of the mean square errors converges to one. Cardinal interpolation is optimal for band-limited fields, and if the spectral mass accumulates at the origin, the field is almost band-limited. However, the results of Stein [19] point out that the high frequency behavior of S is crucial for the interpolation error in the high rate sampling case, and the conjecture is actually false. To show this, we first approximate (17) with

$$\sigma_c^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \approx 2\tau(\boldsymbol{A}) \frac{1}{(2\pi)^d} \int_{\omega_1 > \rho} S(\boldsymbol{\omega}) d\boldsymbol{\omega},$$

which is the union-bound approximation for high signal-to-noise ratios. By analogous arguments as used in the proof of Theorem 3 we can show the following result.

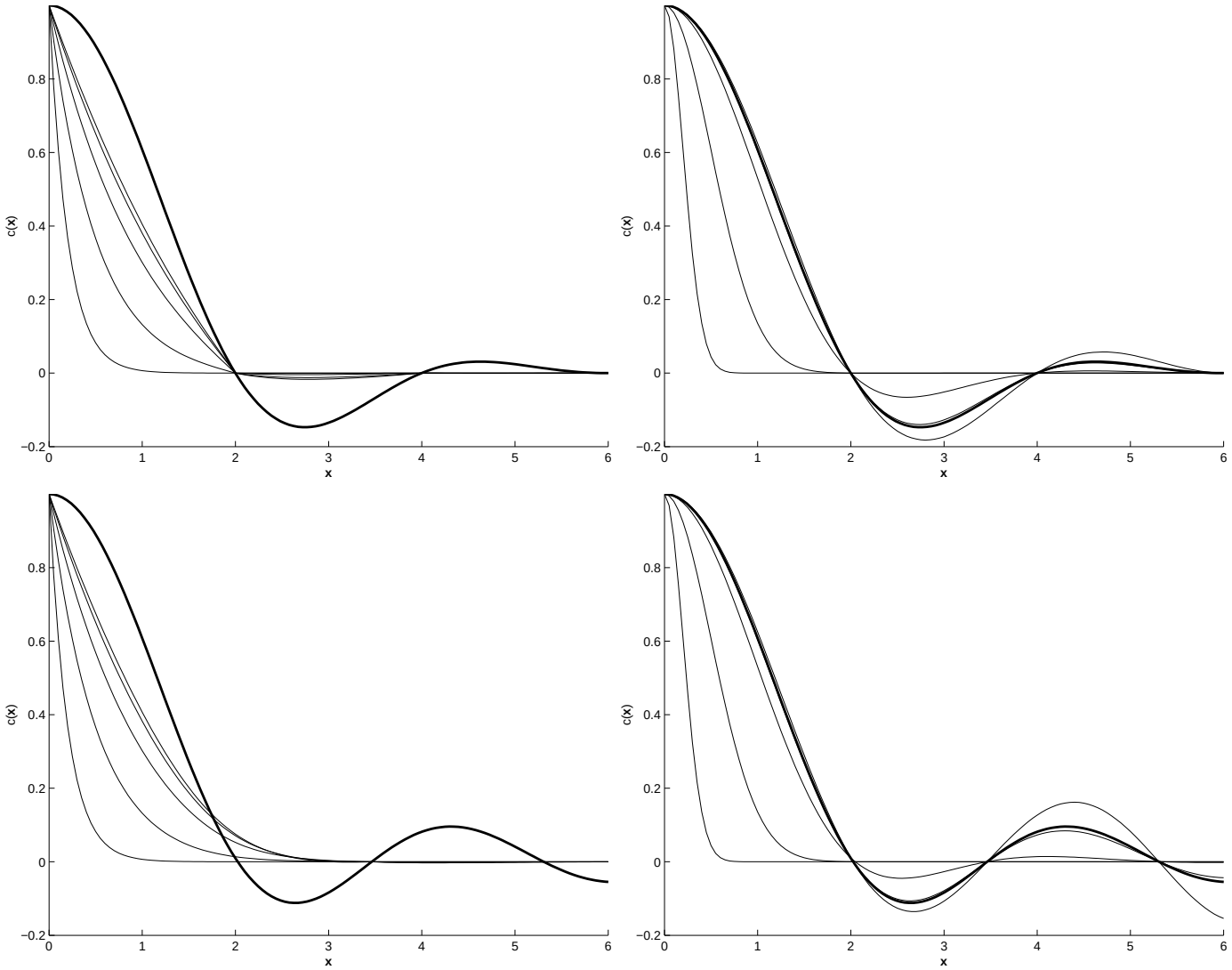

Fig. 1. Cardinal (thick line) and optimal interpolation functions c (thin lines, for $\beta \in \{1/5, 1/2, 1, 2, 5\}$) for the 2D hexagonal lattice. Top: along an axis through a lattice point and one of its nearest neighbors, situated at a distance of 2. Bottom: along an axis through a lattice point and one of its second nearest neighbors, situated at a distance of $2\sqrt{3}$. Left: Exponential covariance $\exp(-\beta\|\boldsymbol{x}\|)$. Right: Gaussian covariance $\exp(-\|\beta\boldsymbol{x}\|^2/2)$.

Proposition 4: Under the assumptions of Theorem 3,

$$\frac{\sigma_c^2(\text{ave}, \Lambda(\boldsymbol{B}))}{\exp(-(\alpha\rho)^p)(\alpha\rho)^{d-p(d+1)/2}} \to \tau(\boldsymbol{A}) 2C(2\pi p)^{-(d+1)/2}$$

as $\alpha \to \infty$, and hence

$$\frac{\sigma^2(\text{ave}, \Lambda(\boldsymbol{B}))}{\sigma_c^2(\text{ave}, \Lambda(\boldsymbol{B}))} \to \frac{\pi}{4} < 1.$$

The optimal lattice for cardinal interpolation is again the lattice solving the dual packing problem. It is somewhat surprising that the asymptotic loss of cardinal over optimal interpolation is independent of the dimension and of the parameter p, that is, of the shape of the spectral density.

VI. Some numerical results

First, we illustrate the difference between the cardinal and the optimal interpolator. Figure 1 shows the the weight functions c of (8) and (16) in specific directions for $d = 2$ and the hexagonal lattice. For (8), we use both the exponential covariance $R(\boldsymbol{x}) = \exp(-\beta\|\boldsymbol{x}\|)$ and the Gaussian covariance $R(\boldsymbol{x}) = \exp(-\|\beta\boldsymbol{x}\|^2/2)$. Note that all weight functions must be equal to one at the origin and zero at all other lattice points, but they may be zero also at non-lattice points. The figures show that in general the optimal weight function decays more quickly than the cardinal weight function, with a faster decay for higher β. For small values of β, such as $\beta = 1/5$ for the Gaussian covariance in Figure 1, the decay is slower than cardinal, since the spectral density $S(\boldsymbol{\omega}) = 2\pi\beta^{-2}\exp(-\|\boldsymbol{\omega}\|^2/(2\beta^2))$ is effectively concentrated on a set that is smaller than $\Omega(\boldsymbol{A})$.

Next, we compare efficiencies of specific lattices with our asymptotic results. If we have two lattices $\Lambda(\boldsymbol{B}_1)$ and $\Lambda(\boldsymbol{B}_2)$ such that $\text{vol}(\Omega(\boldsymbol{B}_1)) = \text{vol}(\Omega(\boldsymbol{B}_2)) = 1$ then we can define the asymptotic relative efficiency of $\Lambda(\boldsymbol{B}_2)$

with respect to $\Lambda(\boldsymbol{B}_1)$ as follows: For any $\beta > 0$, define $\beta' = \beta'(\beta)$ by the equation

$$\sigma^2(\text{ave}, \beta'\Lambda(\boldsymbol{B}_2)) = \sigma^2(\text{ave}, \beta\Lambda(\boldsymbol{B}_1)),$$

assuming that a solution exists. In words, we adjust the sampling rate for the second lattice such that the average interpolation error is the same. The ratio of the sampling rates is then $(\beta'/\beta)^{-1/d}$. The high-rate asymptotic relative efficiency of $\Lambda(\boldsymbol{B}_2)$ with respect to $\Lambda(\boldsymbol{B}_1)$ is now defined as the limit of $(\beta'/\beta)^{1/d}$ as $\beta \to 0$. Similarly, the low-rate asymptotic relative efficiency is defined as the limit of the same expression as $\beta \to \infty$. It is easily seen that under the condition (33) of Theorem 3, the high-rate efficiency is equal to $(\rho(\boldsymbol{A}_1)/\rho(\boldsymbol{A}_2))^d$ and under the condition (26) of Theorem 2, the low-rate efficiency is equal to $(\rho(\boldsymbol{B}_1)/\rho(\boldsymbol{B}_2))^d$. For $d = 2$, the asymptotic relative efficiency of the hexagonal with respect to the square lattice is equal to 1.15 both in the high- and the low-rate sampling limit since both lattices are self dual. For $d = 3$ the packing radius is maximized for the face-centered cubic lattice. The dual of the face-centered cubic lattice is the body-centered cubic lattice which therefore maximizes the dual packing radius. Hence in $d = 3$, the optimal lattice depends on the sampling rate. The relative efficiency of these lattices is equal to 1.09 and so the gains are not tremendous. However, the asymptotic relative efficiency of the optimal lattice in $d = 3$ over the cubic lattice is 1.41 in both the low- and high-rate case, which is more substantial. In $d = 8$, where the so-called lattice E_8 has a number of optimality properties [4], both the high- and low-rate asymptotic relative efficiency of E_8 over the cubic lattice is as high as 16.

Figure 2 illustrates our duality result. We have evaluated numerically the average mean square error for interpolation of a process with a Gaussian covariance which is sampled on two rectangular lattices with generator matrices $\boldsymbol{B}_1 = \text{diag}(1/2, 1/2, 4)$ and $\boldsymbol{B}_2 = \text{diag}(2, 2, 1/4)$. The two lattices have the same point density, but their performance in sampling depends on the sampling rate: $\boldsymbol{B}_1$ is better for low-rate sampling and $\boldsymbol{B}_2$ for high-rate sampling. This curious behavior corresponds to the fact that $\boldsymbol{B}_1$ has a higher packing radius than $\boldsymbol{B}_2$, whereas the reverse is true for their duals. The relative efficiency between these two lattices is 8 in both the low- and high-rate regime. The two optimal lattices for $d = 3$ have a similar relation, although their low relative efficiency (1.09) would make the two curves almost indistinguishable in a graph like Fig. 2.

Finally, we give some examples to illustrate the quality of our asymptotic approximations for the average mean square error. In Figure 3, we show the average mean square error (computed numerically) as a function of β for a process with Gaussian covariance sampled on the rectangular lattice generated by $\boldsymbol{B}_1 = \text{diag}(1/2, 1/2, 4)$, along with five approximations. We see that the approximations agree closely in the appropriate ranges of β, and that (32) and (27) are better approximations than those given by Theorem 3 and (28), respectively. The extra term in (28) compared with (23) does not improve the convergence much in this example.

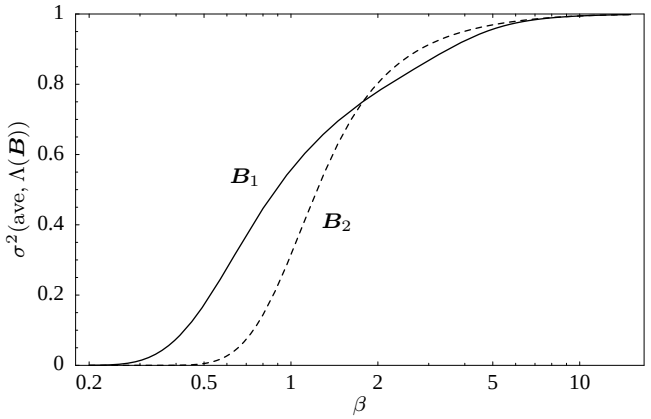

Fig. 2. The dependency on the packing radius is illustrated by two rectangular lattices with generator matrices $\boldsymbol{B}_1 = \text{diag}(1/2, 1/2, 4)$ and $\boldsymbol{B}_2 = \text{diag}(2, 2, 1/4)$. The process has a Gaussian covariance.

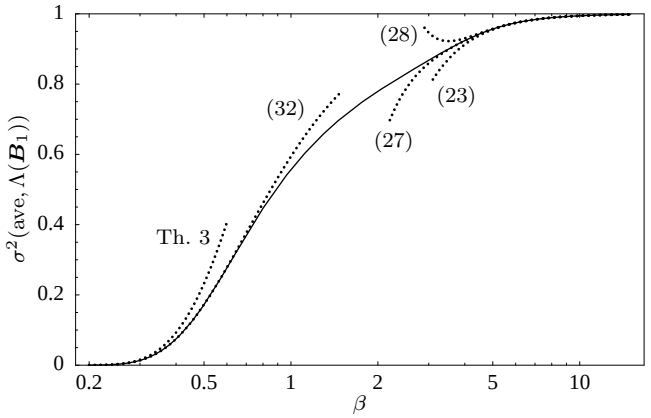

Fig. 3. The average mean square error for a process with Gaussian covariance sampled on the rectangular lattice generated by $\boldsymbol{B}_1 = \text{diag}(1/2, 1/2, 4)$ (solid), along with five approximations thereof.

These features are confirmed also for the case $d = 2$ with the cubic and the hexagonal lattice. Tables I and II deal with low rate sampling. We use numbers instead of graphs in order to give more precise information on the errors. The approximations are excellent and they cover a range of sampling rates where interpolation is still reasonable. The second approximation (27) is always better than (23) and the hexagonal lattice is always better than the cubic lattice. We conjecture that the hexagonal lattice is optimal for all twodimensional, isotropic, monotonic covariance functions and all sampling rates, because it is self-dual. Note that for the exponential covariance, the superiority of the hexagonal over the cubic lattice becomes apparent with the approximation (27) only for relatively high values of β. The reason is that the kissing number $\tau(\boldsymbol{B})$, which appears as a factor in the difference between (23) and (27), is larger for the hexagonal lattice. Generally, the approximations are better for the Gaussian covariance, which was to be expected since the covariance decays faster.

In the high-rate sampling case, we compare the func-

TABLE I

Average mean square interpolation errors and their approximations for the exponential covariance.

value	lattice	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$	$\beta = 3.5$	$\beta = 4$
exact	cubic	.2137	.4074	.5670	.6880	.7743	.8338	.8745	.9028
exact	hexagonal	.2123	.4052	.5649	.6864	.7732	.8331	.8741	.9026
(23)	any	–5.283	–.5708	.3019	.6073	.7487	.8255	.8718	.9018
(27)	cubic	–.1423	.4567	.5720	.6864	.7732	.8334	.8744	.9027
(27)	hexagonal	2.460	.8719	.6549	.7035	.7764	.8338	.8743	.9026

TABLE II

Average mean square interpolation errors and their approximations for the Gaussian covariance.

value	lattice	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$	$\beta = 5$	$\beta = 6$
exact	cubic	.00518	.3147	.6524	.803655	.874336	.9127335
exact	hexagonal	.00329	.3039	.6517	.803652	.874336	.9127335
(23)	any	–2.142	.2146	.6509	.803651	.874336	.9127335
(27)	cubic	–.829	.3135	.6524	.803655	.874336	.9127335
(27)	hexagonal	–.154	.3156	.6517	.803652	.874336	.9127335

tion $r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega})$ (see (31)) whose integral is equal to $\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B}))$ with two approximations: The first approximation is $\sum_{\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega})$ (see (32)). The second approximation is the function we obtain when we replace each term $q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega})$ by its Laplace approximation, cf. (47). Figures 4 and 5 show the results for the spectral density $S(\boldsymbol{\omega}) = \exp(-\alpha\|\boldsymbol{\omega}\|)$ for two values of α. This illustrates the appearance of the peaks at the points $\boldsymbol{\omega} = \hat{\boldsymbol{\lambda}}/2$ with $\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})$ with increasing α. A similar behavior is observed for other spectral densities that satisfy the assumption (33) although the range of values α where the approximations is reasonable depends strongly on p. For $S(\boldsymbol{\omega}) = \exp(-\|\alpha\boldsymbol{\omega}\|^2/2)$, that is for the Gaussian covariance, this is the case for $\alpha \geq 2.5$ or equivalently $\beta \leq 0.4$. For $S(\boldsymbol{\omega}) = \exp(-\|\alpha\boldsymbol{\omega}\|^{1/2})$ we need $\alpha \geq 50$.

VII. Summary and conclusions

If we consider sampling lattices with the same density and a class of signals characterized by a fixed covariance function R, then we can ask for any sampling rate β^{-d}, how large is the average interpolation error $\sigma^2(\mathrm{ave}, \Lambda(\beta\boldsymbol{B}))$ and which lattice minimizes this average interpolation error. The results obtained in this paper can be summarized in the Table III.

Unfortunately, for these statements we need additional conditions on the decay of the covariances or spectral densities respectively, and it would be interesting to formulate and prove more general results. The approximations for the average interpolation error are quite accurate for a large range of sampling rates, and the duality between low- and high-rate sampling that we found is surprising.

TABLE III

Summary of main results.

sampling rate	$\sigma^2(\mathrm{ave}, \Lambda(\beta\boldsymbol{B}))$	Optimality criterion
∞	0	none
large	Theorem 3	dual packing radius
≈ 1	Theorem 1	numerical
small	Theorem 2	packing radius
very small	equation (23)	none
0	$R(\mathbf{0})$	none

Acknowledgment

We thank Tilmann Gneiting for showing us how to prove that the convolution of an isotropic decreasing covariance is decreasing, and to an anonymous referee for showing us an alternative proof of this result. Moreover, the same referee also pointed out formula (14) to us which led to a simplified proof of Proposition 3.

Appendix
Proof of Theorems

Proof of Lemma 1: We introduce the fundamental parallelotope of the lattice $\Lambda(\boldsymbol{A})$:

$$\widetilde{\Omega}(\boldsymbol{A}) \stackrel{\mathrm{def}}{=} \left\{ \boldsymbol{\omega} \in \mathbb{R}^d;\, \boldsymbol{\omega} = \boldsymbol{A}^T\boldsymbol{w} \text{ with } \boldsymbol{w} \in [0,1)^d \right\}.$$

Like the Voronoi regions, the translates of $\widetilde{\Omega}(\boldsymbol{A})$ by elements of the lattice $\Lambda(\boldsymbol{A})$ form a tiling of the space $\mathbb{R}^d$. Moreover, $\mathrm{vol}(\Omega(\boldsymbol{A})) = \mathrm{vol}(\widetilde{\Omega}(\boldsymbol{A})) = \sqrt{(\det(\boldsymbol{A}\boldsymbol{A}^T))}$ or, for square $\boldsymbol{A}$, $\mathrm{vol}(\Omega(\boldsymbol{A})) = |\det(\boldsymbol{A})|$. The integral over $\Omega(\boldsymbol{A})$ of periodic functions is the same as the integral over $\widetilde{\Omega}(\boldsymbol{A})$. Hence by a change of variables from $\boldsymbol{\omega} = \boldsymbol{A}^T\boldsymbol{w}$ to $\boldsymbol{w}$ we obtain

$$\frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) d\boldsymbol{\omega} = \int_{[0,1)^d} \exp(i\boldsymbol{u}^T\boldsymbol{A}^T\boldsymbol{w}) d\boldsymbol{w}.$$

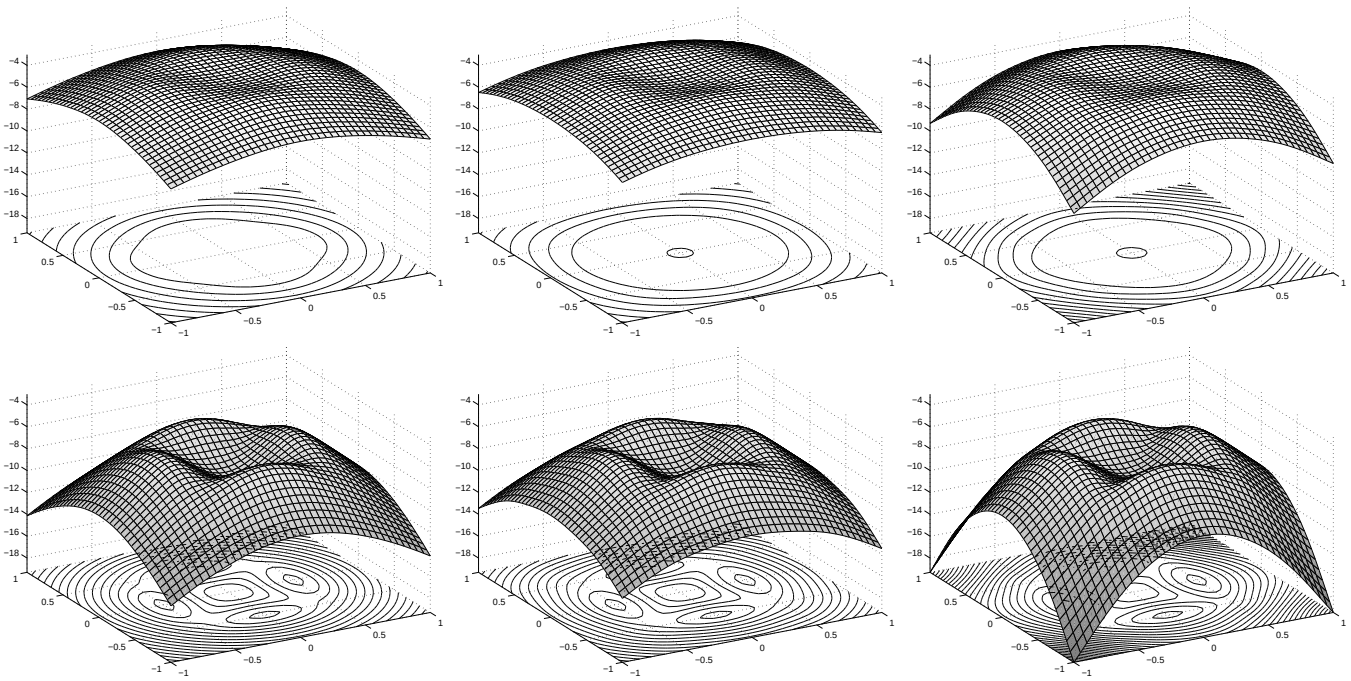

Fig. 4. The function $r(\boldsymbol{\omega})$ (eq. (31), left) and its two approximations (middle and right) in logarithmic scale as a function of the twodimensional vector $\boldsymbol{\omega}$, for the square lattice and the exponential spectral density. The scale parameter α is 5 (top) and 10 (bottom).

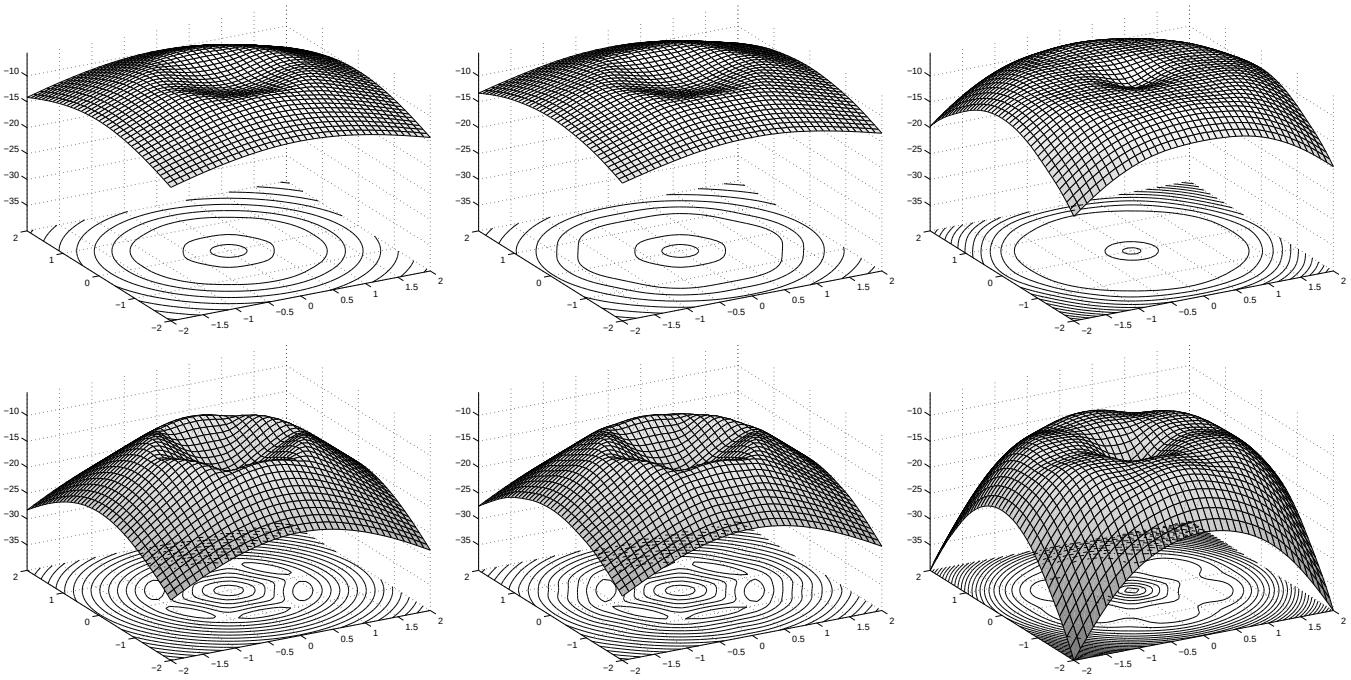

Fig. 5. Same as Fig. 4 for the hexagonal lattice.

By the definition of the dual lattice, $\boldsymbol{u}^T\boldsymbol{A}^T$ is an integer vector times 2π, and thus the claim follows from the basic properties of the complex exponential. □

Proof of Proposition 1: Because

$$\int_{\Omega(\boldsymbol{A})} \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})d\boldsymbol{\omega} = \int_{\mathbb{R}^d} S(\boldsymbol{\omega})d\boldsymbol{\omega} = (2\pi)^d R(\mathbf{0}), \tag{34}$$

$\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})$ is finite almost everywhere and thus $\psi_{\boldsymbol{x}}$ is well defined and bounded by one. We have to show two things: First, the function $\psi_{\boldsymbol{x}}$ from (6) corresponds to a random variable in $H_Z(\Lambda(\boldsymbol{B}))$, and second, $\exp(i\boldsymbol{\omega}^T\boldsymbol{x}) - \psi_{\boldsymbol{x}}(\boldsymbol{\omega})$ is orthogonal to $\exp(i\boldsymbol{\omega}^T\boldsymbol{u})$ for any $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$. The first claim holds because $\psi_{\boldsymbol{x}}$ is periodic with period $\boldsymbol{\lambda}$ for any $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$. For the second claim,

we split the integration over $\mathbb{R}^d$ into integrations over all translates of $\Omega(\boldsymbol{A})$ and use the periodicity of $\psi_{\boldsymbol{x}}(\boldsymbol{\omega})$ and $\exp(i\boldsymbol{\omega}^T\boldsymbol{u})$. Then we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} \psi_{\boldsymbol{x}}(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T\boldsymbol{u}) S(\boldsymbol{\omega}) d\boldsymbol{\omega} \\
&= \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \int_{\Omega(\boldsymbol{A})} \psi_{\boldsymbol{x}}(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T\boldsymbol{u}) S(\boldsymbol{\omega}+\boldsymbol{\lambda}) d\boldsymbol{\omega} \\
&= \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \int_{\Omega(\boldsymbol{A})} \exp(i\boldsymbol{x}^T(\boldsymbol{\omega}+\boldsymbol{\lambda})) S(\boldsymbol{\omega}+\boldsymbol{\lambda}) \exp(-i\boldsymbol{\omega}^T\boldsymbol{u}) d\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^d} \exp(i\boldsymbol{x}^T\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T\boldsymbol{u}) S(\boldsymbol{\omega}) d\boldsymbol{\omega}.
\end{aligned}$$

□

Proof of Theorem 1: Because $\widehat{Z}(\boldsymbol{x})$ is an orthogonal projection,

$$\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) = \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2] = \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))Z(\boldsymbol{x})].$$

Using Proposition 1, the mean square error is therefore equal to

$$\begin{aligned}
&\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (\exp(i\boldsymbol{x}^T\boldsymbol{\omega}) - \psi_{\boldsymbol{x}}(\boldsymbol{\omega})) \exp(-i\boldsymbol{x}^T\boldsymbol{\omega}) S(\boldsymbol{\omega}) d\boldsymbol{\omega} \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} (1 - \exp(i\boldsymbol{x}^T\boldsymbol{\lambda})) S(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})} S(\boldsymbol{\omega}) d\boldsymbol{\omega},
\end{aligned}$$

which is the first claim. For the second claim, we note that by Lemma 1, averaging $\exp(i\boldsymbol{x}^T\boldsymbol{\lambda})$ over $\boldsymbol{x} \in \Omega(\boldsymbol{B})$ gives zero for $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\boldsymbol{0}\}$. In a similar way, the second inequality in (15) is proved because $\sup_{\boldsymbol{x}} |1 - \exp(i\boldsymbol{x}^T\boldsymbol{\lambda})| = 2$. The first inequality in (15) is trivial.

The expression (13) follows by splitting the integration over $\mathbb{R}^d$ into integrations over all translates of $\Omega(\boldsymbol{A})$ and using the periodicity of $\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} S(\boldsymbol{\omega}+\boldsymbol{\lambda})$. Finally (14) follows by applying the Poisson summation formula to the numerator and the denominator of the integrand in (13). □

Proof of Proposition 2: We use the expression (13) in the equivalent form

$$\begin{aligned}
&\sigma^2(\mathrm{ave}, \Lambda(\boldsymbol{B})) \\
&\quad = \frac{1}{(2\pi)^d} \int_{\Omega(\boldsymbol{A})} \frac{(\sum_{\boldsymbol{\lambda}} S(\boldsymbol{\omega}+\boldsymbol{\lambda}))^2 - \sum_{\boldsymbol{\lambda}} S^2(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}} S(\boldsymbol{\omega}+\boldsymbol{\lambda})} d\boldsymbol{\omega},
\end{aligned}$$

where all summations are over $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$. The proof follows from some simple algebraic manipulations. In order to simplify the notation, let $(a_k;\, k = 0, 1, \ldots)$ be an arbitrary nonnegative and summable sequence. Then we have

$$\left(\sum_{k=0}^{\infty} a_k\right)^2 + \left(\sum_{k=1}^{\infty} a_k\right)^2 = a_0^2 + 2\sum_{k=0}^{\infty} a_k \sum_{k=1}^{\infty} a_k.$$

Therefore

$$\begin{aligned}
\left(\sum_{k=0}^{\infty} a_k\right)^2 - \sum_{k=0}^{\infty} a_k^2 &= 2\sum_{k=0}^{\infty} a_k \sum_{k=1}^{\infty} a_k - \left(\sum_{k=1}^{\infty} a_k\right)^2 - \sum_{k=1}^{\infty} a_k^2 \\
&\le 2\sum_{k=0}^{\infty} a_k \sum_{k=1}^{\infty} a_k.
\end{aligned}$$

Substituting $S(\boldsymbol{\omega}+\boldsymbol{\lambda}_k)$ for a_k, where $\{\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \ldots\}$ is an enumeration of $\Lambda(\boldsymbol{A})$ with $\boldsymbol{\lambda}_0 = \boldsymbol{0}$, the first claim follows.

For the second claim, we observe that

$$a_0 \sum_{k=1}^{\infty} a_k - \sum_{k=1}^{\infty} a_k^2 = \sum_{k=1}^{\infty} (a_0 - a_k) a_k \ge 0$$

if $a_0 \ge a_k$ for all k. By the definition of the Voronoi cell, $\|\boldsymbol{\omega}\|^2 \le \|\boldsymbol{\omega} - \boldsymbol{\lambda}\|^2$ for all $\boldsymbol{\omega} \in \Omega(\boldsymbol{A})$ and all $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$, and thus the second claim follows. □

Proof of Proposition 3: By a change of summation, it follows from the definition of the k-fold convolution (20) that for all $k \ge 1$

$$\begin{aligned}
&\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \Delta^k(\boldsymbol{u}) \\
&= \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \Delta(\boldsymbol{u}) \cdot \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \Delta^{k-1}(\boldsymbol{u}) \qquad (35) \\
&= \left(\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \Delta(\boldsymbol{u})\right)^k \qquad (36)
\end{aligned}$$

where (36) follows recursively by repeated application of (35). Moreover, by a similar substition

$$\begin{aligned}
&\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) R^{*2}(\boldsymbol{u}) \cdot \sum_{\boldsymbol{v}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{v}^T\boldsymbol{\omega}) \Delta^k(\boldsymbol{v}) = \\
&\qquad \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \sum_{\boldsymbol{v}\in\Lambda(\boldsymbol{B})} R^{*2}(\boldsymbol{u}-\boldsymbol{v}) \Delta^k(\boldsymbol{v}). \qquad (37)
\end{aligned}$$

The claim (22) now follows from (14) by applying in turn (19), (36), (37), and Lemma 1 (note that $\mathrm{vol}(\Omega(\boldsymbol{A})) = (2\pi)^d/\mathrm{vol}(\Omega(\boldsymbol{B}))$). □

Proof of Theorem 2: First we show that the quantity ϵ in (18) is asymptotically equal to $\tau R(2\rho\boldsymbol{e})$. We have

$$\epsilon = R(2\rho\boldsymbol{e}) \left(\tau + \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B});\, \|\boldsymbol{u}\|>2\rho} \frac{R(\boldsymbol{u})}{R(2\rho\boldsymbol{e})}\right).$$

By the assumption (26), $R(\boldsymbol{u})/R(2\rho\boldsymbol{e})$ converges to zero for any fixed $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$ with $\|\boldsymbol{u}\| > 2\rho$ and it is for all $\beta \ge 1$ upper-bounded by a constant times $\exp(-\|\boldsymbol{u}\|^p)$. Hence by Lebesgue's dominated convergence theorem, the sum over all $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$ with $\|\boldsymbol{u}\| > 2\rho$ converges also to zero provided that $\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(-\|\boldsymbol{u}\|^p)$ is finite. This follows by adapting an argument of Gunning [8, p. 71]. By the definition of a lattice, any $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$ has the form $\boldsymbol{B}^T\boldsymbol{w}$ with $\boldsymbol{w} \in \mathbb{Z}^d$. Denoting the smallest eigenvalue of the matrix $\boldsymbol{B}\boldsymbol{B}^T$ by $\lambda_{\min}$, we have

$$\|\boldsymbol{u}\|^2 = \boldsymbol{w}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{w} \ge \lambda_{\min}\|\boldsymbol{w}\|^2.$$

Moreover, for $p \ge 2$ we have by Jensen's inequality $\|\boldsymbol{w}\|^p \ge d^{(p-2)/2}(|w_1|^p + \cdots + |w_d|^p)$, whereas for $p < 2$ we obtain $\|\boldsymbol{w}\|^p \ge (|w_1|^p + \cdots + |w_d|^p)$ by summing the inequalities

$$\frac{w_i^p}{(w_1^2 + \cdots + w_d^2)^{p/2}} \ge \frac{w_i^2}{w_1^2 + \cdots + w_d^2}.$$

Hence, we conclude that, with $c \stackrel{\text{def}}{=} \lambda_{\min}^{p/2} \min(1, d^{(p-2)/2})$,

$$\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(-\|\boldsymbol{u}\|^p) \le \left(\sum_{w=-\infty}^{\infty} \exp(-c|w|^p) \right)^d < \infty.$$

The same argument shows that $\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} R^2(\boldsymbol{u})$ is equal to $\tau R^2(2\rho\boldsymbol{e})$ times a factor that converges to one. Finally, as we will show below, R^{*2} is also monotonically decreasing and thus the same argument can be used once again to show that $\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} R(\boldsymbol{u})R^{*2}(\boldsymbol{u})$ is equal to $\tau R(2\rho\boldsymbol{e})R^{*2}(2\rho\boldsymbol{e})$ times a factor that converges to one. Hence, (24) is asymptotically equivalent to (27), and the error term (25) is of lower order than (27).

It remains to show the monotonicity of the convolution. For this, we write R as a superposition of indicator functions ϕ_r of the spheres with center $\boldsymbol{0}$ and radius r:

$$R(\boldsymbol{x}) = \int_0^\infty \phi_r(\boldsymbol{x}) dG(r).$$

Here $\phi_r(\boldsymbol{x})$ is equal to one if $\|\boldsymbol{x}\| \le r$ and zero otherwise, and $G(r) = R(\boldsymbol{0}) - R(r\boldsymbol{e})$. Then

$$R^{*2}(\boldsymbol{x}) = \int_0^\infty \int_0^\infty \phi_r * \phi_s(\boldsymbol{x}) dG(r) dG(s).$$

The convolution $\phi_r * \phi_s(\boldsymbol{x})$ is nothing else than the volume of the intersection of two spheres with radii r and s whose centers have distance $\|\boldsymbol{x}\|$. Thus for any r and s, $\phi_r * \phi_s$ is a non-increasing isotropic function, and the same thing holds for R^{*2}. Alternatively, one can use the result [7, eq. (36)].

Note that in (27), the two last terms go in opposite directions because both R and R^{*2} are positive. In order to find out which term dominates, we need to analyze the behavior of R^{*2}. As β increases, the maxima of $R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u})$ become more and more pronounced. Thus we obtain the leading term of the convolution by a Laplace approximation argument (see, e.g., [5, Ch. 4]). This technique restricts the integration for the convolution to a neighborhood where the integrand is maximal and replaces the integrand there by a simpler function. Although Laplace approximations are well known, we could not find a result in the literature that applies directly to our problem. Therefore we give here the proofs of the approximations (28) and (29).

It turns out that the location of the maxima of $R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u})$ and also the asymptotic behavior of the convolution depends on the value of p in the assumption (26). We begin with the case $p \ge 1$. In this case, the function $\|\boldsymbol{x}\|^p + \|\boldsymbol{u}+\boldsymbol{x}\|^p$ is minimal for $\boldsymbol{x} = -\boldsymbol{u}/2$ if $p > 1$ whereas for $p = 1$ it is minimal on the segment from $\boldsymbol{0}$ to $-\boldsymbol{u}$. We first consider the case $p > 1$. We let B denote the ball with center $-\boldsymbol{u}/2$ and radius $\eta\|\boldsymbol{u}\|$ where η will be chosen later: $B \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathbb{R}^d;\ \|\boldsymbol{x} + \boldsymbol{u}/2\| \le \eta\|\boldsymbol{u}\|\}$. By the triangle inequality, we have for $\boldsymbol{x} \in B$ $\|\boldsymbol{x}\| \le \|\boldsymbol{u}\|(\frac{1}{2} + \eta)$ and also $\|\boldsymbol{x}+\boldsymbol{u}\| \le \|\boldsymbol{u}\|(\frac{1}{2}+\eta)$. Assumption (26) implies that there is a constant $C_1 > 0$ such that $R_0(r) \ge C_1 \exp(-|r|^p)$ for all r. Thus on B

$$R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u}) \ge C_1^2 \exp(-2^{1-p}(1+2\eta)^p(\beta\|\boldsymbol{u}\|)^p)$$

and therefore by restricting the integration to B

$$R^{*2}(\boldsymbol{u}) \ge \text{const.} \|\boldsymbol{u}\|^d \exp(-2^{1-p}(1+2\eta)^p(\beta\|\boldsymbol{u}\|)^p).$$

By a simple change of variables, $R^{*2}(\boldsymbol{0})$ is equal to a constant times β^{-d} and thus the claim follows if we choose η such that $2^{1-p}(1+2\eta)^p < 1$, which is possible for $p > 1$.

Next, we consider the case $p = 1$. We may assume that $\boldsymbol{u} = (\|\boldsymbol{u}\|, 0, \ldots, 0)^T$ and we write $\boldsymbol{x} \in \mathbb{R}^d$ as $(x, \boldsymbol{y}^T)^T$ with $\boldsymbol{y} \in \mathbb{R}^{d-1}$. We will restrict the integration to the strip $B = \{\boldsymbol{x} \in \mathbb{R}^d;\ -\|\boldsymbol{u}\| \le x \le 0\}$. By the triangle inequality, $\|\boldsymbol{x}\| \le |x| + \|\boldsymbol{y}\|$ and $\|\boldsymbol{x}+\boldsymbol{u}\| \le \|\boldsymbol{u}\| - |x| + \|\boldsymbol{y}\|$ on B. This implies for $\boldsymbol{x} \in B$

$$R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u}) \ge C_1^2 \exp(-\beta\|\boldsymbol{u}\| - 2\beta\|\boldsymbol{y}\|)$$

and therefore by restricting the integration to B

$$R^{*2}(\boldsymbol{u}) \ge C_1^2 \exp(-\beta\|\boldsymbol{u}\|)\|\boldsymbol{u}\|\beta^{1-d} \int_{\mathbb{R}^{d-1}} \exp(-2\|\boldsymbol{y}\|) d\boldsymbol{y}.$$

The expression on the right is bounded below by a constant times $\beta\|\boldsymbol{u}\|R(\boldsymbol{u})R^{*2}(\boldsymbol{0})$ and (28) follows.

For $p < 1$, the two last terms in (27) are in general of the same order as will become apparent in the proof. Hence we need a more precise analysis of the convolution. We use again a Laplace approximation argument. For $p < 1$, the minima of $\|\boldsymbol{x}\|^p + \|\boldsymbol{x}+\boldsymbol{u}\|^p$ are at $\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{x} = -\boldsymbol{u}$. We will assume that $\boldsymbol{u} = (\|\boldsymbol{u}\|, 0, \ldots, 0)^T$ and we introduce the half-space $H = \{\boldsymbol{x} \in \mathbb{R}^d;\ x_1 \ge -\|\boldsymbol{u}\|/2\}$. Using the symmetry of R, we obtain after a change of variables

$$\frac{R^{*2}(\boldsymbol{u})}{R(\boldsymbol{u})} = 2 \int_H \frac{R(\boldsymbol{u}+\boldsymbol{x})R(\boldsymbol{x})}{R(\boldsymbol{u})} d\boldsymbol{x}.$$

Then we have a single maximum near $\boldsymbol{x} = \boldsymbol{0}$, and we will replace the integral over H by the integral over B where B denotes the ball with center $\boldsymbol{0}$ and radius $\eta\beta^{-p}$ where η will be chosen later: $B \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathbb{R}^d;\ \|\boldsymbol{x}\| \le \eta\beta^{-p}\}$. On B, $R(\boldsymbol{u}+\boldsymbol{x})/R(\boldsymbol{u})$ is practically equal to one for any $\boldsymbol{u}$, leading to the approximations

$$\begin{aligned} \frac{R^{*2}(\boldsymbol{u})}{R(\boldsymbol{u})} &\approx 2 \int_B \frac{R(\boldsymbol{u}+\boldsymbol{x})R(\boldsymbol{x})}{R(\boldsymbol{u})} d\boldsymbol{x} \\ &\approx 2 \int_B R(\boldsymbol{x}) d\boldsymbol{x} \approx 2 \int_{\mathbb{R}^d} R(\boldsymbol{x}) d\boldsymbol{x}. \end{aligned} \tag{38}$$

The approximation (29) follows by combining (38) with (27).

The rest of the proof consists of controlling the errors due to the three approximations in (38). For the first approximation, the key argument is to show that for some $\delta > 0$, constants $C_1 < C_2$, and all $\boldsymbol{x} \in H$

$$\begin{aligned} \frac{R(\boldsymbol{u}+\boldsymbol{x})R(\boldsymbol{x})}{R(\boldsymbol{u})} &\le \frac{C_2^2}{C_1} \exp(\beta^p(\|\boldsymbol{u}\|^p - \|\boldsymbol{u}+\boldsymbol{x}\|^p - \|\boldsymbol{x}\|^p)) \\ &\le \frac{C_2^2}{C_1} \exp(-\delta\beta^p\|\boldsymbol{x}\|^p). \end{aligned} \tag{39}$$

The first inequality holds because (26) implies that if we choose C_1 small enough and C_2 big enough, then

$C_1 \exp(-(\beta\|\boldsymbol{x}\|)^p) \le R(\boldsymbol{x}) \le C_2 \exp(-(\beta\|\boldsymbol{x}\|)^p)$ for all $\boldsymbol{x}$. For the second inequality, we distinguish between the two cases $x_1 \ge 0$ and $-\|\boldsymbol{u}\|/2 \le x_1 \le 0$. If $x_1 \ge 0$, then $\|\boldsymbol{u}+\boldsymbol{x}\| \ge \|\boldsymbol{u}\| + x_1 \ge \|\boldsymbol{u}\|$ and thus (39) holds for all $\delta \le 1$. If $-\|\boldsymbol{u}\|/2 \le x_1 \le 0$, then $\|\boldsymbol{u}+\boldsymbol{x}\| \ge \|\boldsymbol{u}\| - |x_1|$. Hence it is sufficient to show that for all $0 \le r \le 1/2$

$$(1-\delta)r^p + (1-r)^p \ge 1 \tag{40}$$

(simply put $r = |x_1|/\|\boldsymbol{u}\|$). But by concavity of $t \to t^p$, $(1-t)^p \ge (1-t)$ for all $0 \le t \le 1$ and $t^p \ge t2^{1-p}$ for $0 \le t \le 1/2$. Hence (40) and thus also (39) hold for $1-\delta = 2^{p-1}$. Finally, (39) is sufficient to justify the first approximation in (38) since by a simple change of variables and the definition of B

$$\begin{aligned}\int_{H\setminus B} \exp(-\delta\beta^p\|\boldsymbol{x}\|^p)d\boldsymbol{x} &\le \beta^{-d}\int_{\|\boldsymbol{x}\|>\eta\beta^{1-p}} \exp(-\delta\|\boldsymbol{x}\|^p)d\boldsymbol{x}\\ &= o(\beta^{-d}) = o\left(\int_{\mathbb{R}^d} R(\boldsymbol{x})d\boldsymbol{x}\right).\end{aligned}$$

This last argument can be repeated in order to justify the third approximation in (38).

Thus there remains the justification of the middle approximation. Because $\|\boldsymbol{u}\| - \|\boldsymbol{x}\| \le \|\boldsymbol{u}+\boldsymbol{x}\| \le \|\boldsymbol{u}\| + \|\boldsymbol{x}\|$, we have for $\|\boldsymbol{x}\| \to 0$ by the definition of the derivative

$$\begin{aligned}|\,\|\boldsymbol{u}\|^p - \|\boldsymbol{u}+\boldsymbol{x}\|^p| &\le \|\boldsymbol{u}\|^p(1-(1-\|\boldsymbol{x}\|/\|\boldsymbol{u}\|)^p)\\ &= p\|\boldsymbol{u}\|^{p-1}\|\boldsymbol{x}\| + o(\|\boldsymbol{x}\|).\end{aligned}$$

This implies that for any $c > p\|\boldsymbol{u}\|^{p-1}$ and β sufficiently large, we have for all $\boldsymbol{x} \in B$

$$\exp(-c\eta) \le \exp(\beta^p(\|\boldsymbol{u}\|^p - \|\boldsymbol{u}+\boldsymbol{x}\|^p)) \le \exp(c\eta).$$

By the assumption (26), we therefore have also

$$\exp(-c\eta) \le \frac{R(\boldsymbol{x}+\boldsymbol{u})}{R(\boldsymbol{u})} \le \exp(c\eta).$$

Because η can be chosen arbitrarily small, this justifies the middle approximation in (38). □

Proof of Theorem 3: First we determine the points where $r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega})$ is maximal. Since for any two positive real numbers a and b, $ab = \max(a,b)\cdot\min(a,b)$ and $\max(a,b) \le a+b \le 2\max(a,b)$, $r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega})$ can be bounded as

$$\begin{aligned}\frac{1}{2}\min(S(\boldsymbol{\omega}), \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} S(\boldsymbol{\omega}+\boldsymbol{\lambda})) &\le r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega})\\ &\le \min(S(\boldsymbol{\omega}), \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} S(\boldsymbol{\omega}+\boldsymbol{\lambda})).\end{aligned} \tag{41}$$

Our assumptions on S allow us to replace the infinite sum $\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} S(\boldsymbol{\omega}+\boldsymbol{\lambda})$ by the largest summand, cf. the proof of Theorem 2. Together with the bounds (41), this implies that the maxima of $r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega})$ are asymptotically at $\boldsymbol{\omega} = \hat{\boldsymbol{\lambda}}$ with $\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})$. Moreover, as α tends to infinity, the maxima become more and more pronounced so that we can use Laplace approximations once again.

The proof consists of two steps. First, we are going to show that there is a $\delta > 0$ and an integrable function $h(\boldsymbol{\omega})$ such that for all $\alpha \ge 1$ and all $\boldsymbol{\omega} \in \mathbb{R}^d$

$$\left| r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega}) - \sum_{\hat{\boldsymbol{\lambda}}\in\Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}},\boldsymbol{\omega})\right| \le \alpha^d \exp(-\alpha^p(\rho+\delta)^p)h(\boldsymbol{\omega}). \tag{42}$$

In a second step we are going to show that

$$\frac{\int_{\mathbb{R}^d} q(2\rho\boldsymbol{e},\boldsymbol{\omega})d\boldsymbol{\omega}}{\exp(-(\alpha\rho)^p)(\alpha\rho)^{d-p(d+1)/2}} \to \frac{C}{4}\left(\frac{2\pi}{p}\right)^{(d+1)/2}. \tag{43}$$

From this, the theorem follows, cf. (30) and (32).

For (42) we choose δ such that for any $\boldsymbol{\omega} \in \mathbb{R}^d$, there exist at most two lattice points $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$ such that $\|\boldsymbol{\omega}-\boldsymbol{\lambda}\| \le \rho+\delta$, and if two such points exist, they must have distance 2ρ in addition. For $\|\boldsymbol{\omega}\| > \rho+\delta$, we will use the bound

$$\left| r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega}) - \sum_{\hat{\boldsymbol{\lambda}}\in\Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}},\boldsymbol{\omega})\right| \le (\tau+1)S(\boldsymbol{\omega}), \tag{44}$$

which follows from (41) and the analogous bound

$$\frac{1}{2}\min(S(\boldsymbol{\omega}), S(\boldsymbol{\omega}-\hat{\boldsymbol{\lambda}})) \le q(\hat{\boldsymbol{\lambda}},\boldsymbol{\omega}) \le \min(S(\boldsymbol{\omega}), S(\boldsymbol{\omega}-\hat{\boldsymbol{\lambda}})). \tag{45}$$

By (44) and our assumption (33) on S, there is a constant $C_1 \ge C$ such that

$$\begin{aligned}\alpha^{-d}S(\boldsymbol{\omega}) &\le C_1\exp(-(\alpha\|\boldsymbol{\omega}\|)^p)\\ &= C_1\exp(-\|\boldsymbol{\omega}\|^p)\exp(-(\alpha^p-1)\|\boldsymbol{\omega}\|^p)\\ &\le C_1\exp(-\|\boldsymbol{\omega}\|^p)\exp((\rho+\delta)^p)\exp(-\alpha^p(\rho+\delta)^p).\end{aligned}$$

Hence (42) holds on $\|\boldsymbol{\omega}\| > \rho+\delta$ if we set

$$h(\boldsymbol{\omega}) = C_1(\tau+1)\exp((\rho+\delta)^p)\exp(-\|\boldsymbol{\omega}\|^p).$$

For $\|\boldsymbol{\omega}\| \le \rho+\delta$, we use the bound

$$\left| r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega}) - \sum_{\hat{\boldsymbol{\lambda}}\in\Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}},\boldsymbol{\omega})\right| \le 2\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0},\hat{\boldsymbol{\xi}}\}} S(\boldsymbol{\omega}-\boldsymbol{\lambda}), \tag{46}$$

where $\hat{\boldsymbol{\xi}}$ is the point in $\Psi(\boldsymbol{A})$ closest to $\boldsymbol{\omega}$. In order to see why this bound holds, denote $S(\boldsymbol{\omega})$ by a, $S(\boldsymbol{\omega}-\hat{\boldsymbol{\xi}})$ by b and the sum of $S(\boldsymbol{\omega}-\boldsymbol{\lambda})$ over all $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\boldsymbol{0},\hat{\boldsymbol{\xi}}\}$ by c. By simple algebraic manipulations we find

$$\left|\frac{a(b+c)}{a+b+c} - \frac{ab}{a+b}\right| = \frac{a^2c}{(a+b)(a+b+c)} \le c.$$

From this and (45), (46) follows. If $\|\boldsymbol{\omega}\| \le \rho+\delta$, then by the definition of $\hat{\boldsymbol{\xi}}$ and our choice of δ, $\|\boldsymbol{\omega}-\boldsymbol{\lambda}\| > \rho+\delta$ for all $\Lambda(\boldsymbol{A}) \setminus \{\boldsymbol{0},\hat{\boldsymbol{\xi}}\}$. Hence, arguing as above, we obtain in this case

$$\begin{aligned}\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0},\hat{\boldsymbol{\xi}}\}} S(\boldsymbol{\omega}-\boldsymbol{\lambda}) &\le C_1\alpha^d\exp((\rho+\delta)^p)\\ &\quad\cdot\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}\exp(-\|\boldsymbol{\omega}+\boldsymbol{\lambda}\|^p)\exp(-\alpha^p(\rho+\delta)^p)\end{aligned}$$

and thus (42) holds on $\|\boldsymbol{\omega}\| \le \rho + \delta$ if we set

$$h(\boldsymbol{\omega}) = 2C_1 \exp((\rho+\delta)^p) \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(-\|\boldsymbol{\omega}+\boldsymbol{\lambda}\|^p).$$

It is easy to see that h is integrable because

$$\begin{aligned}\int_{\|\boldsymbol{\omega}\|\le\rho+\delta} &\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(-\|\boldsymbol{\omega}+\boldsymbol{\lambda}\|^p)d\boldsymbol{\omega} \\ &\le (\tau+1)\int_{\Omega(\boldsymbol{A})} \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(-\|\boldsymbol{\omega}+\boldsymbol{\lambda}\|^p)d\boldsymbol{\omega} \\ &= (\tau+1)\int_{\mathbb{R}^d} \exp(-\|\boldsymbol{\omega}\|^p)d\boldsymbol{\omega}.\end{aligned}$$

(By our choice of δ, there are $\tau+1$ Voronoi cells that intersect $\{\boldsymbol{\omega}\in\mathbb{R}^d;\ \|\boldsymbol{\omega}\|\le\rho+\delta\}$.)

Finally, we turn to the proof of (43). Replacing $\boldsymbol{\omega}$ by $\boldsymbol{\omega}+\rho\boldsymbol{e}$, we have

$$\int_{\mathbb{R}^d} q(2\rho\boldsymbol{e},\boldsymbol{\omega})d\boldsymbol{\omega} = \int_{\mathbb{R}^d} \frac{S(\rho\boldsymbol{e}+\boldsymbol{\omega})S(\rho\boldsymbol{e}-\boldsymbol{\omega})}{S(\rho\boldsymbol{e}+\boldsymbol{\omega})+S(\rho\boldsymbol{e}-\boldsymbol{\omega})}d\boldsymbol{\omega},$$

and we will work with this symmetric form. Without loss of generality, we assume that $\boldsymbol{e} = (1,0,\ldots,0)^T$ and we write $\boldsymbol{\omega}\in\mathbb{R}^d$ as $(\omega_1,\boldsymbol{\xi}^T)^T$ with $\boldsymbol{\xi}\in\mathbb{R}^{d-1}$. By a Taylor expansion of the function $x\to x^{p/2}$ at the point $x=\rho^2$ we conclude that for any $\varepsilon>0$ there is a $\delta>0$ such that for $|\omega_1|\le\delta$ and $\|\boldsymbol{\xi}\|\le\delta$

$$\left|\|\boldsymbol{\omega}+\rho\boldsymbol{e}\|^p - \rho^p - a_1\omega_1 - a_2\|\boldsymbol{\xi}\|^2\right| \le \varepsilon(|\omega_1|+\|\boldsymbol{\xi}\|^2),$$

where $a_1 = p\rho^{p-1}$ and $a_2 = p\rho^{p-2}/2$. By our assumption (33) on S, we thus have the approximation

$$\begin{aligned}&\frac{S(\rho\boldsymbol{e}+\boldsymbol{\omega})S(\rho\boldsymbol{e}-\boldsymbol{\omega})}{S(\rho\boldsymbol{e}+\boldsymbol{\omega})+S(\rho\boldsymbol{e}-\boldsymbol{\omega})} \\ &\qquad\approx C\alpha^d\exp(-(\alpha\rho)^p)\frac{\exp(-a_2\alpha^p\|\boldsymbol{\xi}\|^2)}{2\cosh(\alpha^p a_1\omega_1)}. \qquad (47)\end{aligned}$$

Moreover, in a neighborhood of $\mathbf{0}$, we obtain upper and lower bounds if we multiply the right hand side (47) by $\exp(\pm 3\varepsilon\alpha^p(\|\xi\|^2+|\omega_1|))$. By similar arguments as above, the integrals outside this neighborhood are asymptotically negligible and thus we can integrate the upper and lower bounds over $\mathbb{R}^d$. Then the integral is the product of two integrals, one with respect to ω_1 and one with respect to $\boldsymbol{\xi}$. By well known properties of the multivariate Gaussian density, the one with respect to $\boldsymbol{\xi}$ is equal to

$$\left(\frac{2\pi}{2\alpha^p(a_2-3\varepsilon)}\right)^{(d-1)/2}.$$

After a change of variables $u=\exp(\alpha^p a_1\omega_1)$, the integral with respect to ω_1 is equal to

$$\frac{2}{\alpha^p a_1}\int_1^\infty \frac{u^{3\varepsilon/a_1}}{1+u^2}du.$$

Using Lebesgue's dominated convergence theorem, it is easy to see that this converges for $\varepsilon\to 0$ to

$$\frac{2}{\alpha^p a_1}\int_1^\infty \frac{1}{1+u^2}du = \frac{\pi}{2\alpha^p a_1}.$$

The integration of the lower bound is similar. Thus by taking all the arguments together and by letting ε go to zero, we obtain

$$\begin{aligned}\int_{\mathbb{R}^d} q(2\rho\boldsymbol{e},\boldsymbol{\omega})d\boldsymbol{\omega} &\sim C\exp(-(\alpha\rho)^p)\alpha^{d-p(d+1)/2} \\ &\quad\cdot\frac{1}{4}(2\pi)^{(d+1)/2}a_1^{-1}(2a_2)^{-(d-1)/2}.\end{aligned}$$

The claim now follows by substituting the values of a_1 and a_2. □

References

[1] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques: Part I—Interpolators and projectors," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2783–2795, Oct. 1999.

[2] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Springer Series in Statistics, New York: Springer, 1987.

[3] J. L. Brown, Jr., "On mean-square aliasing error in the cardinal series expansion of random processes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 254–256, Mar. 1978.

[4] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, vol. 290 of *Grundlehren der Mathematischen Wissenschaften*. New York: Springer, 3rd ed., 1999.

[5] N. G. de Bruijn, *Asymptotic Methods in Analysis*, vol. IV of *Bibliotheca Mathematica*. Amsterdam: North Holland, 1958.

[6] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes I*, vol. 210 of *Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen*. Berlin: Springer, 1974.

[7] T. Gneiting, "On α-symmetric multivariate characteristic functions," *Journal of Multivariate Analysis*, vol. 64, pp. 131–147, Feb. 1998.

[8] R. C. Gunning, *Lectures on Modular Forms*. Princeton Univ. Press, 1962.

[9] F. A. Hamprecht and E. Agrell, "Exploring a space of materials: Spatial sampling design and subset selection," in *Experimental Design for Combinatorial and High Throughput Materials Development* (J. N. Cawse, ed.), ch. 13, New York: Wiley, 2003.

[10] F. A. Hamprecht, W. Thiel, and W. F. van Gunsteren, "Chemical library subset selection algorithms: a unified derivation using spatial statistics," *Journal of Chemical Information and Computer Science*, vol. 42, pp. 414–428, 2002.

[11] B. Jähne, "Representation of multidimensional signals," in *Computer Vision and Applications*, ch. 8, San Diego: Academic Press, 2000.

[12] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Syst. Man Cybern. C*, vol. 28, pp. 39–54, Feb. 1998.

[13] M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, vol. 26, pp. 131–148, 1990.

[14] Y. B. Lim, J. Sacks, W. J. Studden, and W. J. Welch, "Design and analysis of computer experiments when the output is highly correlated over the input space," *The Canadian Journal of Statistics*, vol. 30, pp. 109–126, 2002.

[15] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. Amer. Inst. Elect. Eng.*, vol. 47, pp. 617–644, 1928.

[16] D. P. Petersen and D. Middleton, "Sampling and reconstruction of wave-number-limited functions in N-dimensional Euclidean spaces," *Inform. Contr.*, vol. 5, pp. 279–323, Dec. 1962.

[17] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments (with discussion)," *Statistical Science*, vol. 4, pp. 409–435, 1989.

[18] H. Stark, "Polar, spiral, and generalized sampling and interpolation," in *Advanced Topics in Shannon Sampling and Interpolation Theory* (R. J. Marks II, ed.), ch. 6, New York: Springer, 1993.

[19] M. L. Stein, *Interpolation of Data: Some Theory for Kriging*. Springer Series in Statistics, New York: Springer, 1999.

[20] P. Thévenaz, T. Blu, and M. Unser, "Interpolation revisited," *IEEE Trans. Med. Imag.*, vol. 19, pp. 739–758, July 2000.

[21] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.

Hans R. Künsch received the the M.S. degree in mathematics in 1975 and the Ph.D. degree in mathematics in 1980, both from ETH Zürich, Switzerland.

From 1982-1983, he was research fellow at the University of Tokyo. In 1983 he joined the faculty of the mathematics department at ETH Zürich where he is a full professor since 1992. His research interests include robust statistics, time series analysis and spatial statistics, in particular Gaussian and Markov random field models, resampling methods for dependent data and particle filters.

Dr. Künsch served as co-editor of the Annals of Statistics from 1998 to 2000.

Erik Agrell (M'99–SM'02) received the M.S. degree in electrical engineering in 1989 and the Ph.D. degree in information theory in 1997, both from Chalmers University of Technology, Sweden.

From 1988 to 1990, he was with Volvo Technical Development as a Systems Analyst, and from 1990 to 1997, with the Department of Information Theory, Chalmers University of Technology, as a Research Assistant. In 1997–1999, he was a Postdoctoral Researcher with the University of Illinois at Urbana-Champaign and the University of California, San Diego. In 1999, he joined the faculty of Chalmers University of Technology as an Associate Professor, holding a position first at Chalmers Lindholmen University College and since 2001 in the Department of Signals and Systems. His current research interests include geometrical aspects of coding theory, in particular lattices and block codes, bit-to-symbol mappings in digital communication systems, and coding and modulation for fiber-optic channels.

Dr. Agrell served as Publications Editor for IEEE Transactions on Information Theory from 1999 to 2002.

Fred A. Hamprecht studied chemistry at the Swiss Federal Institute of Technology (ETH) in Zurich and received his Ph.D. from the same school. After a brief period at the Seminar for Statistics, ETH Zurich, he became a Robert Bosch-endowed Associate Professor for Multidimensional Image Processing at the University of Heidelberg in 2001.

He is interested in all aspects of data analysis, ranging from signal and image processing to supervised and unsupervised learning. With his group, he currently develops automated diagnostic systems with applications both in industrial quality control and medicine, where particular emphasis is put on a principled treatment of spatial context.