Optimal lattices for interpolation of stationary random fields

Author(s):
Künsch, Hansruedi; Agrell, Erik; Hamprecht, Fred A.

Publication date:
2003

Permanent link:
https://doi.org/10.3929/ethz-a-004578812

Originally published in:
Research Report / Seminar für Statistik, Eidgenössische Technische Hochschule (ETH) 119

Optimal lattices for interpolation of stationary random fields

by

Hans R. Künsch [1] Erik Agrell [2] and Fred A. Hamprecht [3]

Research Report No. 119
August 2003

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

[1]Seminar for Statistics, ETH Zurich
[2]Chalmers University of Technology Göteborg
[3]University of Heidelberg

Optimal lattices for interpolation of stationary random fields

Hans R. Künsch [§] Erik Agrell [¶] and Fred A. Hamprecht [‖]

Seminar für Statistik
ETH Zentrum
CH-8092 Zürich, Switzerland

August 2003

Abstract

We consider interpolation of a stationary random field that has been observed on a lattice. Exact expressions for the mean square error of the best linear unbiased estimator are given in the frequency domain. Morevoer, we derive asymptotic expansions of the average mean square error when the sampling rate tends to zero and to infinity respectively. This allows us to determine the optimal lattices for interpolation. 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 surfaces, the optimal lattice is the one which solves the dual packing problem. In addition, we compare the best linear unbiased interpolation with cardinal interpolation.

AMS 2000 subject classifications. Primary 60G60; secondary 60G25, 62M40, 11H31.

Key words and phrases. Cardinal interpolation, kriging, lattice theory, multichannel signal processing, packing problems, sampling, spectral representation.

1 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 (Unser, 2000). 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

[§]Seminar for Statistics, ETH Zurich
[¶]Chalmers University of Technology Göteborg
[‖]University of Heidelberg

be discretized, stored, and reconstructed. Applications include computer vision and image processing (Jähne, 2000; Thévenaz, Blu and Unser, 2000), remote sensing (Jimenez and Landgrebe, 1998), medical imaging (Stark, 1993), and experimental design (Hamprecht, Thiel and van Gunsteren, 2002; Hamprecht and Agrell, 2003). 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.

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 Conway and Sloane (1999) 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 (Jähne, 2000), 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 (Hamprecht and Agrell, 2003).

We assume that the multidimensional signal is a realization of a stationary random field 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 (Stein, 1999, Sec. 6.6). The present article is complimentary to earlier efforts that have focused on finding an optimal sampling scheme on a finite domain (Sacks, Welch, Mitchell and Wynn, 1989; Johnson, Moore and Ylvisaker, 1990; Lim, Sacks, Studden and Welch, 2002). In Johnson et al. (1990), a different kind of asymptotics is considered, whereas Lim et al. (2002) introduces numerical procedures. The former paper (Johnson et al., 1990) also notes the "obvious connection" with lattice theory, without investigating it further.

In Section 2, we introduce notation, define basic concepts, and summarize Fourier analysis on lattices. In Section 3, the best linear unbiased estimator for interpolation of the signal is derived and its average error variance is calculated, as a function of the covariance function and lattice. It is concluded that the optimal lattice type depends on the sampling rate. In Sections 4–5, 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 6 we give some numerical examples to illustrate our results. The proofs of all Theorems are given in the Appendix.

2 Preliminaries

2.1 Stationary random fields

We consider a zero-mean stationary random field $(Z(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{R}^d)$ with finite second moments. We denote its covariance function by

$$R(\boldsymbol{x}) = \mathbb{E}[Z(\boldsymbol{y})Z(\boldsymbol{y}+\boldsymbol{x})]. \tag{1}$$

If Z is mean square continuous, then, by Bochner's theorem (see, e.g., Gihman and Skorohod, 1974, 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 f. This means that

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

A sufficient condition for this is

$$\int_{\mathbb{R}^d} |R(\boldsymbol{x})| d\boldsymbol{x} < \infty, \tag{3}$$

and then the spectral density can be obtained as

$$f(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} R(\boldsymbol{x}) \exp(-i\boldsymbol{\omega}^T\boldsymbol{x}) d\boldsymbol{x}. \tag{4}$$

For $D \subset \mathbb{R}^d$, we denote the subspace of $L_2(\Omega, \mathbb{P})$ spanned by the random variables $(Z(\boldsymbol{x}), \boldsymbol{x} \in D)$ by $H_Z(D)$. Instead of $H_Z(\mathbb{R}^d)$, we simply write H_Z. By standard results (see e.g., Gihman and Skorohod, 1974, Ch. 4, Sec. 5, esp. Theorem 3), there is an isometric correspondence between H_Z and the subspace of $L_2(\mathbb{R}^d, (2\pi)^{-d} f(\boldsymbol{\omega}) d\boldsymbol{\omega})$ containing all functions ψ satisfying $\psi(-\boldsymbol{\omega}) = \overline{\psi(\boldsymbol{\omega})}$. Under this correspondence, $Z(\boldsymbol{x}) \leftrightarrow \exp(i\boldsymbol{x}^T\boldsymbol{\omega})$, and $H_Z(D)$ corresponds to the subspace spanned by the functions $\exp(i\boldsymbol{x}^T\boldsymbol{\omega}), \boldsymbol{x} \in D$.

2.2 Lattices

The standard reference for lattices is Conway and Sloane (1999). A d-dimensional lattice $\Lambda(\boldsymbol{B})$ is a 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

$$\boldsymbol{B}_1 = \begin{bmatrix} 2 & 0 \\ 1 & \sqrt{3} \end{bmatrix}, \quad \boldsymbol{B}_2 = \begin{bmatrix} 1 & -\sqrt{3} \\ 1 & \sqrt{3} \end{bmatrix}. \tag{5}$$

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

$$\boldsymbol{B}_3 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{6}$$

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:

$$\Omega(\boldsymbol{B}, \boldsymbol{u}) \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{u}\| \leq \|\boldsymbol{x} - \boldsymbol{u}'\| \quad \forall \boldsymbol{u}' \in \Lambda(\boldsymbol{B}) \right\} \tag{7}$$

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 $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$:

$$\exp(i\boldsymbol{u}^T(\boldsymbol{\omega} + \boldsymbol{\lambda})) = \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \exp(i\boldsymbol{u}^T\boldsymbol{\lambda}) = \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \tag{8}$$

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.

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*

$$\frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) d\boldsymbol{\omega} = \begin{cases} 1 & (\boldsymbol{u} = \boldsymbol{0}) \\ 0 & (\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}). \end{cases} \tag{9}$$

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

$$\widetilde{\Omega}(\boldsymbol{A}) \stackrel{\text{def}}{=} \left\{ \boldsymbol{\omega} \in \mathbb{R}^d : \boldsymbol{\omega} = \boldsymbol{A}^T\boldsymbol{w} \text{ with } \boldsymbol{w} \in [0,1)^d \right\}. \tag{10}$$

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

$$\frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) d\boldsymbol{\omega} = \int_{[0,1)^d} \exp(i\boldsymbol{u}^T\boldsymbol{A}^T\boldsymbol{w}) d\boldsymbol{w}. \tag{11}$$

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. □

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 $(\exp(i\boldsymbol{u}^T\boldsymbol{\omega}); \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$:

$$g(\boldsymbol{\omega}) = \sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B})} c(\boldsymbol{u}) \exp(i\boldsymbol{u}^T\boldsymbol{\omega}). \tag{12}$$

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

$$c(\boldsymbol{u}) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} g(\boldsymbol{\omega}) \exp(-i\boldsymbol{u}^T\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{13}$$

3 The interpolation error

Let $\widehat{Z}(\boldsymbol{x})$ be the best linear unbiased estimator of $Z(\boldsymbol{x})$ based on observations $(Z(\boldsymbol{u}); \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$ on a lattice. Here, "best" is understood in the sense of minimizing the mean square error $\mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2]$. By Hilbert space geometry, $\widehat{Z}(\boldsymbol{x})$ is the orthogonal projection of $Z(\boldsymbol{x})$ on the subspace of $H_Z(\Lambda(\boldsymbol{B}))$. It is much easier to determine first the element $\psi_{\boldsymbol{x}}$ in $L_2(\mathbb{R}^d, (2\pi)^{-d} f(\boldsymbol{\omega}) d\boldsymbol{\omega})$ that corresponds to $\widehat{Z}(\boldsymbol{x})$ under the isometry introduced in Section 2. This is done in the following theorem, generalizing a result by Stein (1999, pp. 98–99) for cubic lattices.

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

$$\psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \frac{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} \exp(i\boldsymbol{x}^T(\boldsymbol{\omega} + \boldsymbol{\lambda})) f(\boldsymbol{\omega} + \boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})}. \tag{14}$$

The proof is given in the Appendix.

In most cases, 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, compare (12). Moreover, it is easily seen that the Fourier coefficients (13) of $\psi_{\boldsymbol{x}}$ are of the form $c(\boldsymbol{x} - \boldsymbol{u})$ where

$$c(\boldsymbol{x} - \boldsymbol{u}) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \frac{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} \exp(i(\boldsymbol{x} - \boldsymbol{u})^T(\boldsymbol{\omega} + \boldsymbol{\lambda})) f(\boldsymbol{\omega} + \boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})} d\boldsymbol{\omega}. \tag{15}$$

Introducing

$$f^*(\boldsymbol{\omega}) = \frac{f(\boldsymbol{\omega})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})} \tag{16}$$

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

$$c(\boldsymbol{x} - \boldsymbol{u}) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{A}))} \int_{\mathbb{R}^d} \exp(i(\boldsymbol{x} - \boldsymbol{u})^T\boldsymbol{\omega}) f^*(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{17}$$

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

$$\psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \sum_{u \in \Lambda(\boldsymbol{B})} c(\boldsymbol{x} - \boldsymbol{u}) \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) \tag{18}$$

that also

$$\widehat{Z}(\boldsymbol{x}) = \sum_{u \in \Lambda(\boldsymbol{B})} c(\boldsymbol{x} - \boldsymbol{u}) Z(\boldsymbol{u}). \tag{19}$$

However, (18) converges in $L_2(\Omega(\boldsymbol{A}), d\boldsymbol{\omega})$ and not necessarily in $L_2(\mathbb{R}^d, (2\pi)^{-d} f(\boldsymbol{\omega}) d\boldsymbol{\omega})$. A sufficient condition for this to hold is for instance that $\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})$ is bounded. The difference between the two L_2-spaces also shows up in cases where the set $\{\boldsymbol{\omega} \in \Omega(\boldsymbol{A}); \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda}) = 0\}$ is not empty. For Theorem 1, it is irrelevant how we define $\psi_{\boldsymbol{x}}$ on this set. However, for the Fourier coefficients in (17) this can make a difference: These coefficients and the representation (19) are then not unique.

Theorem 1 can be used to compute the mean square error for a fixed $\boldsymbol{x}$

$$\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) \stackrel{\mathrm{def}}{=} \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2] \tag{20}$$

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

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \stackrel{\text{def}}{=} \frac{1}{\text{vol}(\Omega(\boldsymbol{B}))} \int_{\Omega(\boldsymbol{B})} \sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) d\boldsymbol{x}. \tag{21}$$

Theorem 2 *The following expressions hold:*

$$\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f^*(\boldsymbol{\omega}) \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\boldsymbol{0}\}} (1 - \exp(i\boldsymbol{x}^T \boldsymbol{\lambda}) f(\boldsymbol{\omega} + \boldsymbol{\lambda}) d\boldsymbol{\omega}, \tag{22}$$

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(\boldsymbol{\omega})(1 - f^*(\boldsymbol{\omega})) d\boldsymbol{\omega} \tag{23}$$

$$= R(\boldsymbol{0}) - \frac{1}{(2\pi)^d} \int_{\Omega(\boldsymbol{A})} \frac{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f^2(\boldsymbol{\omega} + \boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})} d\boldsymbol{\omega}. \tag{24}$$

where f^* is defined in (16) and a value "$0/0$" should be interpreted as zero. In particular, we have

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \leq \sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) \leq 2\, \sigma^2(ave, \Lambda(\boldsymbol{B})). \tag{25}$$

Equations (17), (22) and (23) are already in Petersen and Middleton (1962, Sec. VI). The following bounds for the average mean square error which follow from Theorem 2 are new to our knowledge.

Theorem 3 *For any spectral density we have*

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \leq 2 \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus \Omega(\boldsymbol{A})} f(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{26}$$

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

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \geq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus \Omega(\boldsymbol{A})} f(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{27}$$

As a simple example for Theorem 2 we consider a spectral density f 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 (24) is equal to the nonzero value of f if $\boldsymbol{\omega} + \boldsymbol{\lambda} \in D$ for some $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})$ and zero otherwise. This implies that

$$\frac{\sigma^2(ave, \Lambda(\boldsymbol{B}))}{R(\boldsymbol{0})} = 1 - \frac{\text{vol}(\{\boldsymbol{\omega} \in \Omega(\boldsymbol{A}) | (\boldsymbol{\omega} + \boldsymbol{\lambda}) \in D \text{ for some } \boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})\})}{\text{vol}(D)}. \tag{28}$$

For a similar example, we look at the case where f is zero outside of $\Omega(\boldsymbol{A})$. Then by Theorem 3, $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ 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 (22)). 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 (Nyquist, 1928) due to Petersen and Middleton (1962). Moreover, we can compute the coefficients (17) explicitly. By Theorem 1, the function $\psi_{\boldsymbol{x}}$ corresponding to $\widehat{Z}(\boldsymbol{x})$ (or more precisely, one possible choice of this function) is

$$\psi_{\boldsymbol{x}}(\boldsymbol{\omega}) = \exp(i\boldsymbol{x}^T(\boldsymbol{\omega} \bmod \Lambda(\boldsymbol{A}))) \tag{29}$$

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 (17) in the representation (19) are

$$c(\boldsymbol{x} - \boldsymbol{u}) = \frac{1}{\text{vol}(\Omega(\boldsymbol{A}))} \int_{\Omega(\boldsymbol{A})} \exp(i(\boldsymbol{x} - \boldsymbol{u})^T \boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{30}$$

independently of f. 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 unbiased 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 unbiased 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 $\Omega(\boldsymbol{A})$)

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

In one dimension, these expressions have been derived in Brown (1978). Because $|1 - \exp(ix)|^2 = 2(1 - \cos(x))$, by Lemma 1 the average interpolation error with cardinal interpolation is

$$\sigma_c^2(ave, \Lambda(\boldsymbol{B})) = 2 \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus \Omega(\boldsymbol{A})} f(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{31}$$

which is the upper bound from Theorem 3. 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.

In the next two sections we determine the lattice that minimizes $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ among all lattices with equal volume $\mathrm{vol}(\Omega(\boldsymbol{B}))$. Note that $1/\mathrm{vol}(\Omega(\boldsymbol{B}))$ is the sampling rate, that is, the limit of the number of points in $\Lambda(\boldsymbol{B}) \cap D$ divided by $\mathrm{vol}(D)$ 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 (25) for any other lattice Λ with the same sampling rate

$$\sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda_0) \le 2 \sup_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}, \Lambda). \tag{32}$$

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.

4 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 $\sigma^2(ave, \Lambda(\beta\boldsymbol{B}))$ as β tends to infinity for a fixed lattice $\Lambda(\boldsymbol{B})$ and a fixed covariance $R_0(\boldsymbol{x})$. This implies that the dependence between any two observed values is small and thus the sampled realizations of Z look rough. Instead of rescaling the lattice, we can equivalently rescale the covariance function, that is, we will consider $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ for covariance functions of the form $R(\boldsymbol{x}) = R_0(\beta\boldsymbol{x})$ with β increasing to infinity.

Without loss of generality, we assume that $R(\mathbf{0}) = 1$. We first derive expressions for $\widehat{Z}(\boldsymbol{x})$ and for $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ in the space domain. More precisely, we determine the coefficients $c(\boldsymbol{x} - \boldsymbol{u})$ in the representation (19) directly. The orthogonality condition $\mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))Z(\boldsymbol{u})] = 0$ leads to the system of equations

$$R(\boldsymbol{x} - \boldsymbol{u}) = \sum_{\boldsymbol{u}' \in \Lambda(\boldsymbol{B})} R(\boldsymbol{u} - \boldsymbol{u}')c(\boldsymbol{x} - \boldsymbol{u}) \quad (\boldsymbol{u} \in \Lambda(\boldsymbol{B})). \tag{33}$$

This can be written formally with infinite matrices $R_{\Lambda(\boldsymbol{B})}(\boldsymbol{u}, \boldsymbol{u}') = R(\boldsymbol{u} - \boldsymbol{u}')$ and infinite vectors $\boldsymbol{r}_{\Lambda(\boldsymbol{B}),\boldsymbol{x}}(\boldsymbol{u}) = R(\boldsymbol{x} - \boldsymbol{u})$ and $\boldsymbol{c}_{\Lambda(\boldsymbol{B}),\boldsymbol{x}}(\boldsymbol{u}) = c(\boldsymbol{x} - \boldsymbol{u})$ as $R_{\Lambda(\boldsymbol{B})}\boldsymbol{c}_{\Lambda(\boldsymbol{B}),\boldsymbol{x}} = \boldsymbol{r}_{\Lambda(\boldsymbol{B}),\boldsymbol{x}}$. We are going to show that this equation has a well-defined solution which moreover gives the correct coefficients for $\widehat{Z}(\boldsymbol{x})$ if we assume that (3) holds and that

$$\epsilon \stackrel{\text{def}}{=} \sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}} |R(\boldsymbol{u})| < 1. \tag{34}$$

Since we assume $R(\mathbf{0}) = 1$, we can write $R_{\Lambda(\boldsymbol{B})} = I_{\Lambda(\boldsymbol{B})} - \Delta_{\Lambda(\boldsymbol{B})}$ where $I_{\Lambda(\boldsymbol{B})}$ is an infinite identity matrix and $\Delta_{\Lambda(\boldsymbol{B})}(\boldsymbol{u}, \boldsymbol{u}') = \Delta(\boldsymbol{u} - \boldsymbol{u}')$ where

$$\Delta(\boldsymbol{u}) \stackrel{\text{def}}{=} \begin{cases} 0 & (\boldsymbol{u} = \mathbf{0}) \\ -R(\boldsymbol{u}) & (\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}). \end{cases} \tag{35}$$

Then formally

$$R^{-1}_{\Lambda(\boldsymbol{B})} = (I_{\Lambda(\boldsymbol{B})} - \Delta_{\Lambda(\boldsymbol{B})})^{-1} = \sum_{k=0}^{\infty} \Delta^k_{\Lambda(\boldsymbol{B})}. \tag{36}$$

The elements of the matrix powers $\Delta^k_{\Lambda(\boldsymbol{B})}$ are of the form $\Delta^k(\boldsymbol{u}_i - \boldsymbol{u}_j)$ where we denote by $\Delta^k(\boldsymbol{u})$ the k-fold discrete convolution of Δ

$$\Delta^k(\boldsymbol{u}) \stackrel{\text{def}}{=} \begin{cases} \sum_{\boldsymbol{u}' \in \Lambda(\boldsymbol{B})} \Delta^{k-1}(\boldsymbol{u} - \boldsymbol{u}')\Delta(\boldsymbol{u}') & (k \geq 1) \\ 1 & (k = 0 \text{ and } \boldsymbol{u} = \mathbf{0}) \\ 0 & (k = 0 \text{ and } \boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}). \end{cases} \tag{37}$$

By assumption (34) and an induction argument, we see that

$$\sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B})} |\Delta^k(\boldsymbol{u})| \leq \epsilon^k. \tag{38}$$

Thus the sum $\sum_{k=0}^{\infty} \Delta^k_{\Lambda(\boldsymbol{B})}$ converges, and we will show in the next theorem rigorously that we in this way indeed obtain the best interpolation $\widehat{Z}(\boldsymbol{x})$ and its mean square error.

In order to formulate the result, it is convenient to introduce a short notation for the continuous convolution of the covariance function

$$R^{*2}(\boldsymbol{x}) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} R(\boldsymbol{y})R(\boldsymbol{x} - \boldsymbol{y})d\boldsymbol{y} = \int_{\mathbb{R}^d} R(\boldsymbol{y})R(\boldsymbol{x} + \boldsymbol{y})d\boldsymbol{y}. \tag{39}$$

The equality above follows from the symmetry of the covariance function $R(-\boldsymbol{x}) = R(\boldsymbol{x})$.

Theorem 4 *If $R(\mathbf{0}) = 1$ and assumptions (3) and (34) hold, then the best linear unbiased estimator and its mean square error are given by*

$$\widehat{Z}(\boldsymbol{x}) = \sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B})} \left(\sum_{\boldsymbol{u}' \in \Lambda(\boldsymbol{B})} \sum_{k=0}^{\infty} \Delta^k(\boldsymbol{u} - \boldsymbol{u}')R(\boldsymbol{x} - \boldsymbol{u}') \right) Z(\boldsymbol{u}) \tag{40}$$

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) = 1 - \frac{1}{\text{vol}(\Omega(\boldsymbol{B}))} \sum_{k=0}^{\infty} \sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B})} \Delta^k(\boldsymbol{u})R^{*2}(\boldsymbol{u}). \tag{41}$$

The proof is given in the Appendix.

Heuristically, the contribution of the terms in the expression (41) becomes smaller as k increases, cf. (38). Moreover, because in the sparse sampling case $R^{*2}(\mathbf{0})$ is much larger than $R^{*2}(\boldsymbol{u})$ for $\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}$, for a fixed k the term with $\boldsymbol{u} = \mathbf{0}$ has to be considered separately. Retaining thus the terms with $k = 0$ and $k = 1$ as well as the term with $k = 2$ and $\boldsymbol{u} = \mathbf{0}$, we obtain the approximation

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\mathbf{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \left(1 + \sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}} R^2(\boldsymbol{u}) - \sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}} R(\boldsymbol{u}) \frac{R^{*2}(\boldsymbol{u})}{R^{*2}(\mathbf{0})} \right). \quad (42)$$

By (38) and the Cauchy-Schwarz inequality for $R^{*2}(\boldsymbol{u})$, the error in this approximation can be bounded by

$$\frac{R^{*2}(\mathbf{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \left(\epsilon^2 \max_{\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \{\mathbf{0}\}} \frac{R^{*2}(\boldsymbol{u})}{R^{*2}(\mathbf{0})} + \frac{\epsilon^3}{1 - \epsilon}) \right). \quad (43)$$

The first two terms in (42) do not depend on the lattice. The third and fourth term do, but they are still too complicated for optimization. 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, i.e., the *kissing number,* and by $\boldsymbol{e}$ an arbitrary unit vector.

Theorem 5 *Consider a sequence of isotropic covariance functions $R(\boldsymbol{x}) = R_0(\beta||\boldsymbol{x}||)$ depending on a parameter $\beta \geq 1$ and assume that R_0 is monotonically decreasing and satisfies*

$$C_1 \exp(-r^p) \leq R_0(r) \leq C_2 \exp(-r^p) \quad (44)$$

for some constants $0 < C_1 < C_2 < \infty$ and $p > 0$. Then

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\mathbf{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \left(1 + \tau R^2(2\rho\boldsymbol{e}) - \tau R(2\rho\boldsymbol{e}) \frac{R^{*2}(2\rho\boldsymbol{e})}{R^{*2}(\mathbf{0})} \right) \quad (45)$$

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

Note that in (45), 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., de Bruijn (1958, 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.

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 (44). Although Laplace approximations are well known, we could not find a result in the literature that applies directly to our problem. Therefore we give the proofs of the following two theorems in the Appendix.

Theorem 6 *If the assumptions of Theorem 5 hold for $p \geq 1$, then*

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\mathbf{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} + \frac{\tau}{\mathrm{vol}(\Omega(\boldsymbol{B}))} R(2\rho\boldsymbol{e}) R^{*2}(2\rho\boldsymbol{e}) \tag{46}$$

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

For $p < 1$, the two last terms in (45) are in general of the same order as will be shown in the next Theorem. Hence we need a more precise analysis of the convolution, which in turn requires a slightly more precise assumption on the function R_0 for large arguments.

Theorem 7 *If the assumptions of Theorem 5 hold for $0 < p < 1$ and if (44) is strengthened to*

$$\frac{R_0(r)}{C \exp(-r^p)} \to 1 \quad (r \to \infty) \tag{47}$$

for a constant $0 < C < \infty$, then

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\mathbf{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} + \frac{\tau}{\mathrm{vol}(\Omega(\boldsymbol{B}))} R^2(2\rho\boldsymbol{e}) \left(2 \int_{\mathbb{R}^d} R(\boldsymbol{x}) d\boldsymbol{x} - R^{*2}(\mathbf{0}) \right) \tag{48}$$

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

As a corollary, the lattice $\Lambda(\boldsymbol{B})$ minimizing $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ for given $\mathrm{vol}(\Omega(\boldsymbol{B}))$ is in all cases considered the one maximizing the packing radius $\rho(\boldsymbol{B})$. In situations covered by Theorem 6 this follows from the monotonicity of R^{*2} and in situations covered by Theorem 7 it follows because

$$2 \int_{\mathbb{R}^d} R(\boldsymbol{x}) d\boldsymbol{x} - R^{*2}(\mathbf{0}) = \beta^{-d} \int_{\mathbb{R}^d} (2R_0(||\boldsymbol{x}||) - R_0^2(||\boldsymbol{x}||)) d\boldsymbol{x} \tag{49}$$

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 $\tau(\boldsymbol{B})$.

5 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 $\sigma^2(ave, \Lambda(\beta\boldsymbol{B}))$ 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. Equivalently, in the frequency domain the spectral density function takes the form $f(\boldsymbol{\omega}) = \beta^{-d} f_0(||\boldsymbol{\omega}||/\beta)$ and thus the mass accumulates at the origin. In order to simplify notation, we will use the parameter $\alpha = 1/\beta$ which tends to infinity.

We rewrite (23) in the equivalent form

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) d\boldsymbol{\omega} \tag{50}$$

where

$$r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) = \frac{f(\boldsymbol{\omega}) \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\mathbf{0}\}} f(\boldsymbol{\omega} + \boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})}. \tag{51}$$

Like in the previous section, it turns out that for β increasing to infinity, the peaks of the function $r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega})$ become sharper and thus we can use Laplace approximations once

again. We first explain the result heuristically. Since for any two positive real numbers a and b, $ab = \max(a,b)\cdot\min(a,b)$ and $\max(a,b) \le a+b \le 2\max(a,b)$, the following bounds on $r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega})$ hold

$$\frac{1}{2}\min(f(\boldsymbol{\omega}), \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} f(\boldsymbol{\omega}+\boldsymbol{\lambda})) \le r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega}) \le \min(f(\boldsymbol{\omega}), \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} f(\boldsymbol{\omega}+\boldsymbol{\lambda})). \quad (52)$$

If f is isotropic, monotone and decays quickly, the infinite sum $\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}} f(\boldsymbol{\omega}+\boldsymbol{\lambda})$ is approximately equal to the largest summand which is the one where $||\boldsymbol{\omega}+\boldsymbol{\lambda}||$ is minimal, cf. the proof of Theorem 5. Together with the bounds (52) 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

$$r(\Lambda(\boldsymbol{A}),\boldsymbol{\omega}) \approx q(\hat{\boldsymbol{\lambda}},\boldsymbol{\omega}) \stackrel{\text{def}}{=} \frac{f(\boldsymbol{\omega})f(\hat{\boldsymbol{\lambda}}-\boldsymbol{\omega})}{f(\boldsymbol{\omega})+f(\hat{\boldsymbol{\lambda}}-\boldsymbol{\omega})}, \quad (53)$$

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

$$\sigma^2(ave,\Lambda(\boldsymbol{B})) \approx \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \sum_{\hat{\boldsymbol{\lambda}}\in\Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}},\boldsymbol{\omega})d\boldsymbol{\omega} = \frac{\tau}{(2\pi)^d}\int_{\mathbb{R}^d} q(2\rho\boldsymbol{e},\boldsymbol{\omega})d\boldsymbol{\omega} \quad (54)$$

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 (54) depends essentially only on the values of $f(\boldsymbol{\omega})$ for $\boldsymbol{\omega}$ near $\rho\boldsymbol{e}$. Since we assume f to be isotropic and monotone, $\sigma^2(ave,\Lambda(\boldsymbol{B}))$ 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 8 *Consider a sequence of isotropic spectral density functions $f(\boldsymbol{\omega}) = f_0(\alpha||\boldsymbol{\omega}||)$ depending on a parameter $\alpha \ge 1$ and assume that for some $p > 0$ and some $0 < C < \infty$*

$$\frac{f_0(r)}{C\exp(-r^p)} \to 1 \quad (r\to\infty). \quad (55)$$

Then for $\alpha\to\infty$, the error in the approximation (54) is of lower order and

$$\frac{\sigma^2(ave,\Lambda(\boldsymbol{B}))}{\exp(-(\alpha\rho)^p)(\alpha\rho)^{-p(d+1)/2}} \to C\frac{\tau}{4}\rho^d(2\pi/p)^{(d+1)/2}. \quad (56)$$

It is interesting to compare cardinal interpolation 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 (1999) point out that the high frequency behavior of f is crucial for the interpolation error in the high rate sampling case, and the conjecture is actually false. To show this, we first approximate (31) with

$$\sigma_c^2(ave,\Lambda(\boldsymbol{B})) \approx 2\tau(\boldsymbol{A})\frac{1}{(2\pi)^d}\int_{\omega_1>\rho} f(\boldsymbol{\omega})d\boldsymbol{\omega}, \quad (57)$$

which is the union-bound approximation commonly used in digital communications for high signal-to-noise ratios (Conway and Sloane, 1999, pp. 69–70). By analogous arguments as used in the proof of Theorem 8 we can show the following theorem.

Theorem 9 *Under the assumptions of Theorem 8,*

$$\frac{\sigma_c^2(ave, \Lambda(\boldsymbol{B}))}{\exp(-(\alpha\rho)^p)(\alpha\rho)^{-p(d+1)/2}} \to C\tau\rho^d(2\pi)^{(d-1)/2}p^{-(d+1)/2} \tag{58}$$

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

$$\frac{\sigma^2(ave, \Lambda(\boldsymbol{B}))}{\sigma_c^2(ave, \Lambda(\boldsymbol{B}))} \to \frac{\pi}{4} < 1. \tag{59}$$

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.

6 Some numerical results

If we have two lattices $\Lambda(\boldsymbol{B}_1)$ and $\Lambda(\boldsymbol{B}_2)$ such that $\mathrm{vol}(\Omega(\boldsymbol{B}_1)) = \mathrm{vol}(\Omega(\boldsymbol{B}_2)) = 1$ 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

$$\sigma^2(ave, \beta'\Lambda(\boldsymbol{B}_2)) = \sigma^2(ave, \beta\Lambda(\boldsymbol{B}_1)), \tag{60}$$

assuming 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 efficency 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 efficency is defined as the limit of the same expression as $\beta \to \infty$. It is easily seen that under the condition (55) of Theorem 8, the high-rate efficiency is equal to $(\rho(\boldsymbol{A}_1)/\rho(\boldsymbol{A}_2))^d$ and under the condition (44) of Theorem 5, 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, see Conway and Sloane (1999), has a number of optimality properties, both the high- and low-rate asymptotic relative efficiency of E_8 over the cubic lattice is as high as 16.

Next, we give a few examples to illustrate the quality of our approximations for the average mean square error. We restrict ourselves to the case $d = 2$ and we use the cubic and the hexagonal lattice, scaled to have sampling rate equal to one. In the low-rate sampling case, we compared a numerical approximation for $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ with the approximations (45) and the even simpler approximation

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \approx 1 - \frac{R^{*2}(\boldsymbol{0})}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \tag{61}$$

which depends on the lattice only via the sampling rate. As covariance function we took the so-called Matérn class

$$R_\nu(\boldsymbol{x}) = \frac{K_\nu(\beta||\boldsymbol{x}||)\,(\beta||\boldsymbol{x}||)^\nu}{\Gamma(\nu)2^{\nu-1}} \tag{62}$$

value	lattice	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$
exact	cubic	.2137	.4074	.5670	.6880	.7743	.8338
exact	hexagonal	.2123	.4052	.5649	.6864	.7732	.8331
(61)	any	-5.283	-.5708	.3019	.6073	.7487	.8255
(45)	cubic	-.1423	.4567	.5720	.6864	.7732	.8334
(45)	hexagonal	2.460	.8719	.6549	.7035	.7764	.8338

Table 1: Average mean square interpolation errors and their approximations for the exponential 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
(61)	any	-2.142	.2146	.6509	.803651	.874336	.9127335
(45)	cubic	-.829	.3135	.6524	.803655	.874336	.9127335
(45)	hexagonal	-.154	.3156	.6517	.803652	.874336	.9127335

Table 2: Average mean square interpolation errors and their approximations for the Gaussian covariance.

where K_ν is a modified Bessel funtion, see e.g. Stein (1999). For us, the main advantage of this class is that the convolution belongs to the same class:

$$R_\nu^{*2}(\boldsymbol{x}) = \frac{4\pi\nu^2}{(1+2\nu)\beta^2} R_{1+2\nu}(\boldsymbol{x}). \tag{63}$$

The shape parameter ν regulates the behavior of R at the origin. For $\nu = 0.5$, the Matérn covariance is simply the exponential covariance $R(\boldsymbol{x}) = \exp(-\beta||\boldsymbol{x}||)$. For $\nu \to \infty$ with $\beta = \sqrt{2\nu}$, the Matérn covariance converges to the Gaussian covariance $R(\boldsymbol{x}) = \exp(-(\beta||\boldsymbol{x}||)^2/2)$. We chose values of β in a range where the approximations change from being meaningless to being perfect which depends on the value of the shape parameter ν.

In order to compute $\sigma^2(ave, \Lambda(\boldsymbol{B}))$, we need to discretize the averaging integral and we have to consider the interpolation error based on observing Z at a finite number of lattice points. By trial and error, we chose approximations in such a way that the resulting error should be smaller than the precision in the tables below.

The results for the exponential and the Gaussian covariance are given in Tables 1 and 2. For $\nu = 1$, the results were similar. We see that the approximations are excellent and cover also a range of sampling rates where interpolation is still reasonable. The second approximation (45) is always better than (61), and the hexagonal lattice is always better than the cubic lattice. We conjecture that the hexagonal lattice is optimal for all sampling rates because it is self-dual. Note however that in the exponential case, the approximation (45) is larger for the hexagonal lattice than for the cubic lattice. The reason is that the kissing number $\tau(\boldsymbol{B})$ which appears as a factor in the difference between (61) and (45) is larger for the hexagonal lattice. Of course, for larger β's, the larger packing radius of the hexagonal lattice will dominate as predicted by our theory.

In the high-rate sampling case, we compared the function $r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega})$ (see (51)) whose integral is equal to $\sigma^2(ave, \Lambda(\boldsymbol{B}))$ with its approximations $\sum_{\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega})$ (see (54)) and the function we obtain when we replace each term $q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega})$ by its Laplace approximation, compare (110). We considered the spectral densities $f(\boldsymbol{\omega}) = \exp(-(\alpha||\boldsymbol{\omega}||)^p)$ for various values of p. Since the general behavior is similar, we show only the results for $p =$

1 in Figures 1 and 2. The value of α where peaks at the points $\boldsymbol{\omega} = \hat{\boldsymbol{\lambda}}/2$ with $\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})$ start to appear and the approximations become accurate depends strongly on p. For $p = 2$, this occurs around $\alpha = 2$, whereas for $p = 1/2$ this occurs only around $\alpha = 50$. Note that if $f(\boldsymbol{\omega}) = \exp(-(\alpha||\boldsymbol{\omega}||)^2)$, then the covariance $R(\boldsymbol{x})$ is proportional to $\exp(-(||\boldsymbol{x}||/(2\alpha))^2)$. Hence for the Gaussian covariance $R(\boldsymbol{x}) = \exp(-||\boldsymbol{x}||^2/2)$, our approximations cover sampling rates less or equal to $1/4$ (corresponding to $\beta = 2$ in Table 2) and rates greater or equal than 8 (corresponding to the value $\alpha = 2$ mentioned above).

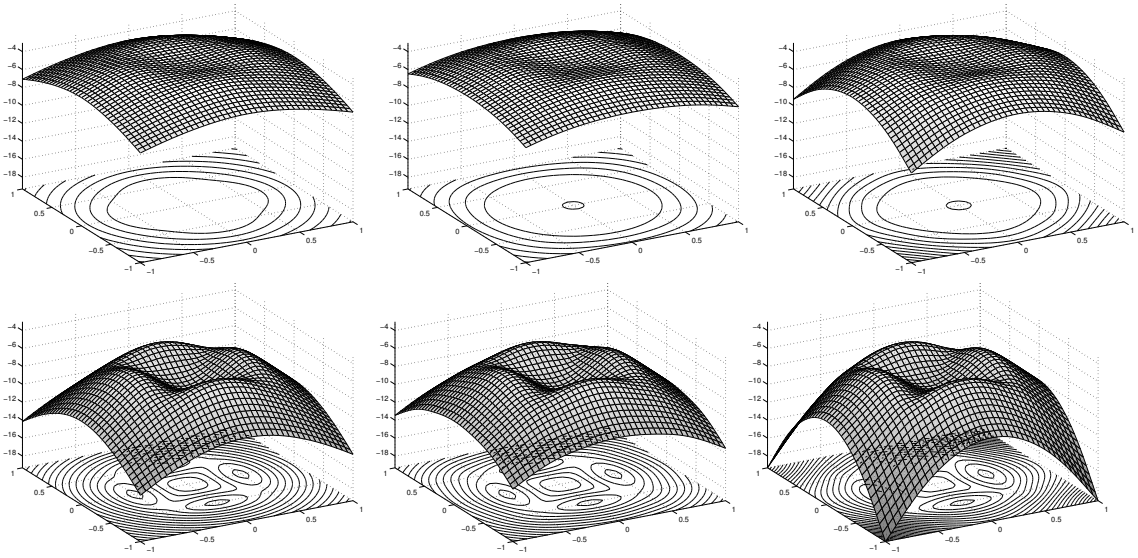

Figure 1: The function r (eq. 50, left) and its two approximations (middle and right) in logarithmic scale for the cubic lattice and the exponential spectral density. The scale parameter α is equal to 5 (upper row) and 10 (lower row).

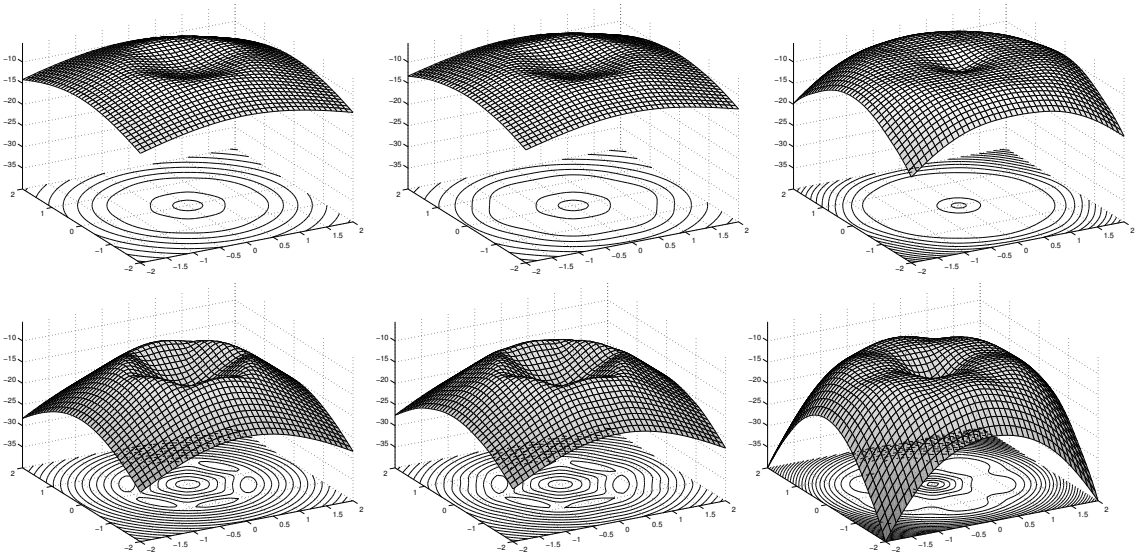

Figure 2: Same as Fig. 1 for the hexagonal lattice.

sampling rate	$\sigma^2(ave, \Lambda(\beta \boldsymbol{B}))$	Optimality criterion
∞	0	none
large	Theorem 8	dual packing radius
≈ 1	Theorem 2	numerical
small	Theorem 5	packing radius
very small	equation (61)	none
0	$R(\mathbf{0})$	none

Table 3: Summary of main results.

7 Summary and conclusions

If we consider lattices with sampling rate 1, i.e. $\mathrm{vol}(\Omega(\boldsymbol{B})) = 1$, and a fixed covariance R, then we can ask for any sampling rate β^{-d}, how large is the average interpolation error $\sigma^2(ave, \Lambda(\beta \boldsymbol{B}))$ and which lattice minimizes this average interpolation error. The results obtained in this paper can be summarized in the Table 7.

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.

A Proof of Theorems

Proof of Theorem 1: Because

$$\int_{\Omega(\boldsymbol{A})} \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda}) d\boldsymbol{\omega} = \int_{\mathbb{R}^d} f(\boldsymbol{\omega}) d\boldsymbol{\omega} = (2\pi)^d R(\mathbf{0}), \tag{64}$$

$\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})$ 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 (14) lies 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

$$\begin{aligned}
& \int_{\mathbb{R}^d} \psi_{\boldsymbol{x}}(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T \boldsymbol{u}) f(\boldsymbol{\omega}) d\boldsymbol{\omega} && (65)\\
= & \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} \int_{\Omega(\boldsymbol{A})} \psi_{\boldsymbol{x}}(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T \boldsymbol{u}) f(\boldsymbol{\omega} + \boldsymbol{\lambda}) d\boldsymbol{\omega} && (66)\\
= & \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} \int_{\Omega(\boldsymbol{A})} \exp(i\boldsymbol{x}^T (\boldsymbol{\omega} + \boldsymbol{\lambda})) f(\boldsymbol{\omega} + \boldsymbol{\lambda}) \exp(-i\boldsymbol{\omega}^T \boldsymbol{u}) d\boldsymbol{\omega} && (67)\\
= & \int_{\mathbb{R}^d} \exp(i\boldsymbol{x}^T \boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T \boldsymbol{u}) f(\boldsymbol{\omega}) d\boldsymbol{\omega}. && (68)
\end{aligned}$$

□

Proof of Theorem 2: Because $\widehat{Z}(\boldsymbol{x})$ is an orthogonal projection,

$$\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) = \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2] = \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))Z(\boldsymbol{x})]. \tag{69}$$

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

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}(\exp(i\boldsymbol{x}^T\boldsymbol{\omega})-\psi_{\boldsymbol{x}}(\boldsymbol{\omega}))\exp(-i\boldsymbol{x}^T\boldsymbol{\omega})f(\boldsymbol{\omega})d\boldsymbol{\omega} \quad = \tag{70}$$

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\frac{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}(1-\exp(i\boldsymbol{x}^T\boldsymbol{\lambda}))f(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}f(\boldsymbol{\omega}+\boldsymbol{\lambda})}f(\boldsymbol{\omega})d\boldsymbol{\omega} \tag{71}$$

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 $\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})\setminus\{\boldsymbol{0}\}$. This also proves the second inequality in (25) because $\sup_{\boldsymbol{x}}|1-\exp(i\boldsymbol{x}^T\boldsymbol{\lambda})|=2$. The first inequality in (25) is trivial.

Finally, (24) follows by splitting the integration over $\mathbb{R}^d$ into integrations over all translates of $\Omega(\boldsymbol{A})$ and using the periodicity of $\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}f(\boldsymbol{\omega}+\boldsymbol{\lambda})$. □

Proof of Theorem 3: We use the expression (24) in the equivalent form

$$\sigma^2(ave,\Lambda(\boldsymbol{B}))=\frac{1}{(2\pi)^d}\int_{\Omega(\boldsymbol{A})}\frac{(\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}f(\boldsymbol{\omega}+\boldsymbol{\lambda}))^2-\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}f^2(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})}f(\boldsymbol{\omega}+\boldsymbol{\lambda})}d\boldsymbol{\omega},$$

The proof follows from some simple algebraic manipulations. In order to simplify the notation, let $(a_k;k=0,1,2,\ldots)$ be an arbitrary nonnegative and summable sequence. Then we have

$$\left(\sum_{k=0}^{\infty}a_k\right)^2=a_0^2+a_0\sum_{k=1}^{\infty}a_k+\sum_{k=0}^{\infty}a_k\sum_{k=1}^{\infty}a_k. \tag{72}$$

Therefore

$$\left(\sum_{k=0}^{\infty}a_k\right)^2-\sum_{k=0}^{\infty}a_k^2=2\sum_{k=0}^{\infty}a_k\sum_{k=1}^{\infty}a_k-\left(\sum_{k=1}^{\infty}a_k\right)^2-\sum_{k=1}^{\infty}a_k^2\le 2\sum_{k=0}^{\infty}a_k\sum_{k=1}^{\infty}a_k. \tag{73}$$

Substituting $f(\boldsymbol{\omega}+\boldsymbol{\lambda}_k)$ 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

$$a_0\sum_{k=1}^{\infty}a_k-\sum_{k=1}^{\infty}a_k^2=\sum_{k=1}^{\infty}(a_0-a_k)a_k\ge 0 \tag{74}$$

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 Theorem 4: First we note that (3) implies that

$$\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})}|R(\boldsymbol{x}+\boldsymbol{u})|<\infty \tag{75}$$

for almost all $x\in\Omega(\boldsymbol{B})$, cf. the argument following (64). Hence by (38) for almost all $\boldsymbol{x}\in\Omega(\boldsymbol{B})$

$$\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})}\left|\sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})}\sum_{k=0}^{\infty}\Delta^k(\boldsymbol{u}-\boldsymbol{u}')R(\boldsymbol{x}-\boldsymbol{u}')\right|\le\frac{1}{1-\epsilon}\sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})}|R(\boldsymbol{x}-\boldsymbol{u}')|<\infty \tag{76}$$

(the exceptional set does not matter since we are interested in the average mean square error). Therefore, the right hand side of (40) converges in H_Z (and also almost surely), see

e.g. Brockwell and Davis (1987, Prop. 3.1.1). To prove (40), we verify $\mathbb{E}[(Z(\boldsymbol{x})\widehat{Z}(\boldsymbol{x}))Z(\boldsymbol{v})] = 0$, or equivalently (compare (33))

$$R(\boldsymbol{x}-\boldsymbol{v}) = \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \left(\sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})} \sum_{k=0}^{\infty} \Delta^k(\boldsymbol{u}-\boldsymbol{u}')R(\boldsymbol{x}-\boldsymbol{u}') \right) R(\boldsymbol{u}-\boldsymbol{v}) \tag{77}$$

for all $\boldsymbol{v} \in \Lambda(\boldsymbol{B})$. This is true because

$$\begin{aligned} \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \Delta^k(\boldsymbol{u}-\boldsymbol{u}')R(\boldsymbol{u}-\boldsymbol{v}) &= \Delta^k(\boldsymbol{v}-\boldsymbol{u}') - \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \Delta^k(\boldsymbol{u}-\boldsymbol{u}')\Delta(\boldsymbol{v}-\boldsymbol{u}) \\ &= \Delta^k(\boldsymbol{v}-\boldsymbol{u}') - \Delta^{k+1}(\boldsymbol{v}-\boldsymbol{u}'). \end{aligned}$$

Hence all terms cancel except the one with $k = 0$.

Moreover, we can also compute $\sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) = R(\boldsymbol{0}) - \mathbb{E}[\widehat{Z}(\boldsymbol{x})Z(\boldsymbol{x})]$. We obtain

$$\begin{aligned} \sigma^2(\boldsymbol{x}, \Lambda(\boldsymbol{B})) &= 1 - \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})} \sum_{k=0}^{\infty} \Delta^k(\boldsymbol{u}-\boldsymbol{u}')R(\boldsymbol{x}-\boldsymbol{u}')R(\boldsymbol{x}-\boldsymbol{u}) \\ &= 1 - \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})} \sum_{k=0}^{\infty} \Delta^k(\boldsymbol{u})R(\boldsymbol{x}-\boldsymbol{u}')R(\boldsymbol{x}-\boldsymbol{u}'+\boldsymbol{u}). \end{aligned}$$

The theorem follows now by averaging this expression over $\boldsymbol{x} \in \Omega(\boldsymbol{B})$ and by noting that

$$\sum_{\boldsymbol{u}'\in\Lambda(\boldsymbol{B})} \int_{\Omega(\boldsymbol{B})} R(\boldsymbol{x}-\boldsymbol{u}')R(\boldsymbol{x}-\boldsymbol{u}'+\boldsymbol{u})d\boldsymbol{x} = \int_{\mathbb{R}^d} R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u})d\boldsymbol{x}. \tag{78}$$

□

Proof of Theorem 5: First we show that the quantity ϵ in (34) is asymptotically equal to $\tau R(2\rho\boldsymbol{e})$. We have

$$\epsilon = R(2\rho\boldsymbol{e}) \left(\tau + \sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B});||\boldsymbol{u}||>2\rho} \frac{R(\boldsymbol{u})}{R(2\rho\boldsymbol{e})} \right). \tag{79}$$

By the assumption (44),

$$\frac{R(\boldsymbol{u})}{R(2\rho\boldsymbol{e})} \leq \frac{C_2}{C_1} \exp(-\beta^p(||\boldsymbol{u}||^p - (2\rho)^p)) \tag{80}$$

converges to zero for any fixed $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$ with $||\boldsymbol{u}|| > 2\rho$ and it is for all $\beta \geq 1$ 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 the argument on p. 71 in Gunning (1962). 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

$$||\boldsymbol{u}||^2 = \boldsymbol{w}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{w} \geq \lambda_{\min}||\boldsymbol{w}||^2 \tag{81}$$

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

$$\frac{w_i^p}{(w_1^2 + \cdots + w_d^2)^{p/2}} \geq \frac{w_i^2}{w_1^2 + \cdots + w_d^2}. \tag{82}$$

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

$$\sum_{\boldsymbol{u}\in\Lambda(\boldsymbol{B})} \exp(-||\boldsymbol{u}||^p) \le \left(\sum_{w=-\infty}^{\infty} \exp(-c|w|^p) \right)^d < \infty. \tag{83}$$

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, (42) is asymptotically equivalent to (45), and the error term (43) is of lower order than (45).

It remains to show the monotonicity of the convolution. We can multiply R by a constant such that it becomes an isotropic probability density on $\mathbb{R}^d$. Then also R^{*2} has the same property. The Fourier transform of R, that is, the spectral density f, is also isotropic and thus has the form $f(\boldsymbol{\omega}) = \phi(||\boldsymbol{\omega}||)$ with $\phi : [0,\infty) \to [0,\infty)$. By the identity (36) in Gneiting (1998), monotonicity of R implies that $\phi(||\boldsymbol{\omega}||)$ with $\boldsymbol{\omega} \in \mathbb{R}^{d+2}$ is a positive definite function. Because the product of two positive definite functions is positive definite, $\phi^2(||\boldsymbol{\omega}||)$ is also positive definite on $\mathbb{R}^{d+2}$. But $\phi^2(||\boldsymbol{\omega}||)$ with $\boldsymbol{\omega} \in \mathbb{R}^d$ is the Fourier transform of R^{*2} and so by applying the identity (36) in Gneiting (1998) in the opposite direction, it follows that R^{*2} is monotonically decreasing. □

Proof of Theorem 6:

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 $||\boldsymbol{x}+\boldsymbol{u}|| \le ||\boldsymbol{u}||(\frac{1}{2}+\eta)$. Thus on B

$$R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u}) \ge C_1^2 \exp(-2^{1-p}(1+2\eta)^p(\beta||\boldsymbol{u}||)^p) \tag{84}$$

and therefore by restricting the integration to B

$$R^{*2}(\boldsymbol{u}) \ge const.||\boldsymbol{u}||^d \exp(-2^{1-p}(1+2\eta)^p(\beta||\boldsymbol{u}||)^p). \tag{85}$$

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 $||\boldsymbol{x}+\boldsymbol{u}|| \le ||\boldsymbol{u}|| - |x| + ||\boldsymbol{y}||$ hold on B. This implies for $\boldsymbol{x} \in B$

$$R(\boldsymbol{x})R(\boldsymbol{x}+\boldsymbol{u}) \ge C_1^2 \exp(-\beta||\boldsymbol{u}|| - 2\beta||\boldsymbol{y}||) \tag{86}$$

and therefore by restricting the integration to B

$$R^{*2}(\boldsymbol{u}) \ge C_1^2 \exp(-\beta||\boldsymbol{u}||)||\boldsymbol{u}||\beta^{1-d} \int_{\mathbb{R}^{d-1}} \exp(-2||\boldsymbol{y}||)d\boldsymbol{y}. \tag{87}$$

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

Proof of Theorem 7: 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 $H = \{\boldsymbol{x} \in \mathbb{R}^d; x_1 \geq -||\boldsymbol{u}||/2\}$. Using the symmetry of R, we obtain after a change of variables

$$\frac{R^{*2}(\boldsymbol{u})}{R(\boldsymbol{u})} = 2\int_H \frac{R(\boldsymbol{u}+\boldsymbol{x})R(\boldsymbol{x})}{R(\boldsymbol{u})} d\boldsymbol{x}. \tag{88}$$

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: $B \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathbb{R}^d; ||\boldsymbol{x}|| \leq \eta\beta^{-p}\}$. On B, $R(\boldsymbol{u}+\boldsymbol{x})/R(\boldsymbol{u})$ is practically equal to one for any $\boldsymbol{u}$, leading to the approximations

$$\frac{R^{*2}(\boldsymbol{u})}{R(\boldsymbol{u})} \approx 2\int_B \frac{R(\boldsymbol{u}+\boldsymbol{x})R(\boldsymbol{x})}{R(\boldsymbol{u})} d\boldsymbol{x} \approx 2\int_B R(\boldsymbol{x})d\boldsymbol{x} \approx 2\int_{\mathbb{R}^d} R(\boldsymbol{x})d\boldsymbol{x}. \tag{89}$$

The theorem follows by combining (89) with (45).

The rest of the proof consists of controlling the errors due to the three approximations in (89). For the first approximation, the key argument is to show that for some $\delta > 0$, constants $C_1 < C_2$ and all $\boldsymbol{x} \in H$

$$\frac{R(\boldsymbol{u}+\boldsymbol{x})R(\boldsymbol{x})}{R(\boldsymbol{u})} \leq \frac{C_2^2}{C_1}\exp(\beta^p(||\boldsymbol{u}||^p - ||\boldsymbol{u}+\boldsymbol{x}||^p - ||\boldsymbol{x}||^p)) \leq \frac{C_2^2}{C_1}\exp(-\delta\beta^p||\boldsymbol{x}||^p). \tag{90}$$

The first inequality holds because (47) implies (44). For the second inequality, note that if $x_1 \geq 0$, then $||\boldsymbol{u}+\boldsymbol{x}|| \geq ||\boldsymbol{u}|| + x_1 \geq ||\boldsymbol{u}||$ and thus (90) holds for all $\delta \leq 1$. If $-||\boldsymbol{u}||/2 \leq x_1 \leq 0$, then $||\boldsymbol{u}+\boldsymbol{x}|| \geq ||\boldsymbol{u}|| - |x_1|$. Hence it is sufficient to show that for all $0 \leq r \leq 1/2$

$$(1-\delta)r^p + (1-r)^p \geq 1 \tag{91}$$

(simply put $r = |x_1|/||\boldsymbol{u}||$). But by concavity of $t \to t^p$, $(1-t)^p \geq (1-t)$ for all $0 \leq t \leq 1$ and $t^p \geq t2^{1-p}$ for $0 \leq t \leq 1/2$. Hence (91) and thus also (90) hold for $1 - \delta = 2^{p-1}$. Finally, (90) is sufficient to justify the first approximation in (89) since by a simple change of variables and the definition of B

$$\begin{aligned}\int_{H\setminus B} \exp(-\delta\beta^p||\boldsymbol{x}||^p)d\boldsymbol{x} &\leq \beta^{-d}\int_{||\boldsymbol{x}||>\eta\beta^{1-p}} \exp(-\delta||\boldsymbol{x}||^p)d\boldsymbol{x} \\ &= o(\beta^{-d}) = o\left(\int_{\mathbb{R}^d} R(\boldsymbol{x})d\boldsymbol{x}\right).\end{aligned} \tag{92}$$

This last argument can be repeated in order to justify the third approximation in (89).

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

$$|||\boldsymbol{u}||^p - ||\boldsymbol{u}+\boldsymbol{x}||^p| \leq ||\boldsymbol{u}||^p(1-(1-||\boldsymbol{x}||/||\boldsymbol{u}||)^p) = p||\boldsymbol{u}||^{p-1}||\boldsymbol{x}|| + o(||\boldsymbol{x}||). \tag{93}$$

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

$$\exp(-c\eta) \leq \exp(\beta^p(||\boldsymbol{u}||^p - ||\boldsymbol{u}+\boldsymbol{x}||^p)) \leq \exp(c\eta). \tag{94}$$

By the assumption (47), we therefore have also

$$\exp(-c\eta) \leq \frac{R(\boldsymbol{x}+\boldsymbol{u})}{R(\boldsymbol{u})} \leq \exp(c\eta). \tag{95}$$

Because η can be chosen arbitrarily small, this justifies the middle approximation in (89). □

Proof of Theorem 8: 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 $\alpha \geq 1$ and all $\boldsymbol{\omega} \in \mathbb{R}^d$

$$\left| r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) - \sum_{\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega}) \right| \leq \exp(-\alpha^p (\rho + \delta)^p) h(\boldsymbol{\omega}). \tag{96}$$

In a second step we are going to show that

$$\frac{\int_{\mathbb{R}^d} q(2\rho \boldsymbol{e}, \boldsymbol{\omega}) d\boldsymbol{\omega}}{\exp(-(\alpha\rho)^p)(\alpha\rho)^{-p(d+1)/2}} \to \frac{C\rho^d}{4} \left(\frac{2\pi}{p}\right)^{(d+1)/2}. \tag{97}$$

From this, the theorem follows, compare (50) and (54).

For (96) 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 $||\boldsymbol{\omega} - \boldsymbol{\lambda}|| \leq \rho + \delta$, and if two such points exist, they must have distance 2ρ in addition. For $||\boldsymbol{\omega}|| > \rho + \delta$, we will use the bound

$$\left| r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) - \sum_{\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega}) \right| \leq (\tau + 1) f(\boldsymbol{\omega}) \tag{98}$$

which follows from (52) and the analogous bound

$$\frac{1}{2} \min(f(\boldsymbol{\omega}), f(\boldsymbol{\omega} - \hat{\boldsymbol{\lambda}})) \leq q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega}) \leq \min(f(\boldsymbol{\omega}), f(\boldsymbol{\omega} - \hat{\boldsymbol{\lambda}})). \tag{99}$$

By (98) and our assumption on f, there is a constant $C_1 \geq C$ such that

$$\begin{aligned} f(\boldsymbol{\omega}) &\leq C_1 \exp(-(\alpha||\boldsymbol{\omega}||)^p) = C_1 \exp(-||\boldsymbol{\omega}||^p) \exp(-(\alpha^p - 1)||\boldsymbol{\omega}||^p) && (100) \\ &\leq C_1 \exp(-||\boldsymbol{\omega}||^p) \exp((\rho + \delta)^p) \exp(-\alpha^p (\rho + \delta)^p). && (101) \end{aligned}$$

Hence (96) holds if we set

$$h(\boldsymbol{\omega}) = C_1(\tau + 1) \exp((\rho + \delta)^p) \exp(-||\boldsymbol{\omega}||^p) \quad (||\boldsymbol{\omega}|| > \rho + \delta). \tag{102}$$

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

$$\left| r(\Lambda(\boldsymbol{A}), \boldsymbol{\omega}) - \sum_{\hat{\boldsymbol{\lambda}} \in \Psi(\boldsymbol{A})} q(\hat{\boldsymbol{\lambda}}, \boldsymbol{\omega}) \right| \leq 2 \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\boldsymbol{0}, \hat{\boldsymbol{\xi}}\}} f(\boldsymbol{\omega} - \boldsymbol{\lambda}), \tag{103}$$

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

$$\left| \frac{a(b + c)}{a + b + c} - \frac{ab}{a + b} \right| = \frac{a^2 c}{(a + b)(a + b + c)} \leq c. \tag{104}$$

From this and (99), (103) follows. If $||\boldsymbol{\omega}|| \leq \rho + \delta$, 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

$$\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\boldsymbol{0}, \hat{\boldsymbol{\xi}}\}} f(\boldsymbol{\omega} - \boldsymbol{\lambda}) \leq C_1 \exp((\rho + \delta)^p) \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} \exp(-||\boldsymbol{\omega} + \boldsymbol{\lambda}||^p) \exp(-\alpha^p (\rho + \delta)^p) \tag{105}$$

and thus (96) holds if we set

$$h(\boldsymbol{\omega}) = 2C_1 \exp((\rho+\delta)^p) \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(-||\boldsymbol{\omega}+\boldsymbol{\lambda}||^p) \quad (||\boldsymbol{\omega}|| \le \rho+\delta). \tag{106}$$

It is easy to see that h is integrable because

$$\begin{aligned}\int_{||\boldsymbol{\omega}||\le\rho+\delta} \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(-||\boldsymbol{\omega}+\boldsymbol{\lambda}||^p)d\boldsymbol{\omega} &\le (\tau+1)\int_{\Omega(\boldsymbol{A})} \sum_{\boldsymbol{\lambda}\in\Lambda(\boldsymbol{A})} \exp(-||\boldsymbol{\omega}+\boldsymbol{\lambda}||^p)d\boldsymbol{\omega} \\ &= (\tau+1)\int_{\mathbb{R}^d} \exp(-||\boldsymbol{\omega}||^p)d\boldsymbol{\omega}, \end{aligned} \tag{107}$$

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

Finally, we turn to the proof of (97). Replacing $\boldsymbol{\omega}$ by $\boldsymbol{\omega}+\rho\boldsymbol{e}$, we have

$$\int_{\mathbb{R}^d} q(2\rho\boldsymbol{e},\boldsymbol{\omega})d\boldsymbol{\omega} = \int_{\mathbb{R}^d} \frac{f(\rho\boldsymbol{e}+\boldsymbol{\omega})f(\rho\boldsymbol{e}-\boldsymbol{\omega})}{f(\rho\boldsymbol{e}+\boldsymbol{\omega})+f(\rho\boldsymbol{e}-\boldsymbol{\omega})}d\boldsymbol{\omega}, \tag{108}$$

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 $||\boldsymbol{\xi}|| \le \delta$

$$\left| ||\boldsymbol{\omega}+\rho\boldsymbol{e}||^p - \rho^p - a_1\omega_1 - a_2||\boldsymbol{\xi}||^2 \right| \le \varepsilon(|\omega_1| + ||\boldsymbol{\xi}||^2) \tag{109}$$

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

$$\frac{f(\rho\boldsymbol{e}+\boldsymbol{\omega})f(\rho\boldsymbol{e}-\boldsymbol{\omega})}{f(\rho\boldsymbol{e}+\boldsymbol{\omega})+f(\rho\boldsymbol{e}-\boldsymbol{\omega})} \approx \exp(-(\alpha\rho)^p)\frac{\exp(-a_2\alpha^p||\boldsymbol{\xi}||^2)}{2\cosh(\alpha^p a_1\omega_1)}. \tag{110}$$

Moreover, in a neighborhood of $\mathbf{0}$, we obtain upper and lower bounds if we multiply the right hand side (110) 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

$$\left(\frac{2\pi}{2\alpha^p(a_2-3\varepsilon)}\right)^{(d-1)/2}. \tag{111}$$

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

$$\frac{2}{\alpha^p a_1}\int_1^\infty \frac{u^{3\varepsilon/a_1}}{1+u^2}du. \tag{112}$$

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

$$\frac{2}{\alpha^p a_1}\int_1^\infty \frac{1}{1+u^2}du = \frac{\pi}{2\alpha^p a_1}. \tag{113}$$

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

$$\int_{\mathbb{R}^d} q(2\rho\boldsymbol{e},\boldsymbol{\omega})d\boldsymbol{\omega} \sim \exp(-(\alpha\rho)^p)\alpha^{-p(d+1)/2}\frac{1}{4}(2\pi)^{(d+1)/2}a_1^{-1}(2a_2)^{-(d-1)/2}. \tag{114}$$

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

Acknowledgments.

We thank Tilman Gneiting for showing us how to prove that the convolution of an isotropic decreasing covariance is decreasing.

References

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

Brown, J. L. (1978). On mean-square aliasing error in the cardinal series expansion of random processes, *IEEE Transactions on Information Theory* **IT-24**(2): 254–256.

Conway, J. H. and Sloane, N. J. A. (1999). *Sphere Packings, Lattices and Groups*, Vol.290 of *Grundlehren der Mathematischen Wissenschaften*, 3rd edn, Springer, New York.

de Bruijn, N. G. (1958). *Asymptotic Methods in Analysis*, Vol. IV of *Bibliotheca Mathematica*, North Holland, Amsterdam.

Gihman, I. I. and Skorohod, A. V. (1974). *The Theory of Stochastic Processes I*, Vol.210 of *Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen*, Springer, Berlin.

Gneiting, T. (1998). On α-symmetric multivariate characteristic functions, *J. Multivar. Anal.* **64**(2): 131–147.

Gunning, R. C. (1962). *Lectures on Modular Forms*, Princeton Univ. Press.

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

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

Jähne, B. (2000). Representation of multidimensional signals, *Computer Vision and Applications*, Academic Press, San Diego, chapter 8.

Jimenez, L. O. and Landgrebe, D. A. (1998). Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data, *IEEE Trans. Syst. Man Cybern. C* **28**(1): 39–54.

Johnson, M. E., Moore, L. M. and Ylvisaker, D. (1990). Minimax and maximin distance designs, *J. Stat. Plann. Inf.* **26**: 131–148.

Lim, Y. B., Sacks, J., Studden, W. J. and Welch, W. J. (2002). Design and analysis of computer experiments when the output is highly correlated over the input space, *The Canadian Journal of Statistics* **30**: 109–126.

Nyquist, H. (1928). Certain topics in telegraph transmission theory, *Trans. Amer. Inst. Elect. Eng.* **47**: 617–644.

Petersen, D. P. and Middleton, D. (1962). Sampling and reconstruction of wave-number-limited functions in N-dimensional Euclidean spaces, *Inform. Contr.* **5**(4): 279–323.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments (with discussion), *Stat. Sci.* **4**: 409–435.

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

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

Thévenaz, P., Blu, T. and Unser, M. (2000). Interpolation revisited, *IEEE Trans. Med. Imag.* **19**(7): 739–758.

Unser, M. (2000). Sampling—50 years after Shannon, *Proc. IEEE* **88**(4): 569–587.