Optimal lattices for interpolation of stationary random fields

Colloque en l'honneur de Xavier Guyon

Hans R. Künsch, Seminar für Statistik, ETH Zürich

Erik Agrell, Chalmers University, Göteborg

Fred A. Hamprecht, University of Heidelberg

Paris, May 30, 2003

1. Introduction

The problem:

For a given sampling rate, where should we observe a stationary random field $Z(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^d$, in order to make estimation at unobserved points as accurate as possible ?

The sampling rate is defined as the limit of number of observation points in a window divided by the volume of the window as the window increases to $\mathbb{R}^d$.

Applications:

Discretizing multidimensional signals ($\boldsymbol{x}$ represents for instance space and time and frequency of a tomografic image).

Design of experiments (computer experiments, discovery of active compounds from a chemical library).

Simplifying assumptions:

- Linear interpolators only.
- Mean and covariance function R known, covariance function isotropic.
- Observation points form a lattice.

If the covariance function is known, “observation points should be as uniformly as possible”. For estimating an isotropic covariance function, there should be observation points at arbitrary distances. These two goals conflict.

To obtain uniform observation points, lattices are a natural choice. However, in higher dimensions the “most uniform” lattice depends on how we define uniformity. The standard cubic lattice is very non-uniform with respect to all criteria.

2. Lattices

A **lattice** Λ is a subset of $\mathbb{R}^d$ consisting of integer linear combinations of d linearly independent generating vectors. The generator $\boldsymbol{B}$ is the matrix with the generating vectors as rows:

$$\Lambda = \Lambda(\boldsymbol{B}) = \{\boldsymbol{u} = \boldsymbol{B}^T \boldsymbol{w} : \boldsymbol{w} \in \mathbb{Z}^n\}.$$

Example Two possible generators for the hexagonal lattice in $\mathbb{R}^2$ are

$$\begin{bmatrix} 2 & 0 \\ 1 & \sqrt{3} \end{bmatrix}, \quad \begin{bmatrix} 1 & -\sqrt{3} \\ 1 & \sqrt{3} \end{bmatrix}.$$

The basic **Voronoi cell** $\Omega(\boldsymbol{B})$ of a lattice is the set of all vectors in $\mathbb{R}^d$ that are at least as close to the origin $\mathbf{0}$ as to any other lattice point. All other Voronoi regions are translates of $\Omega(\boldsymbol{B})$. The sampling rate of a lattice is $1/\mathrm{vol}(\Omega(\boldsymbol{B}))$.

The **packing radius** $\rho(\boldsymbol{B})$ is half the minimum distance between two points of the lattice, i.e. the inradius of $\Omega(\boldsymbol{B})$. The **kissing number** $\tau(\boldsymbol{B})$ is the number of lattice points at distance 2ρ.

Uniform lattices have a large packing-radius. Other measures of uniformity are the covering radius, i.e. the circumradius of $\Omega(\boldsymbol{B})$, and the vector quantization distortion, i.e. the average square distance of a point in $\Omega(\boldsymbol{B})$ from $\mathbf{0}$.

In the frequency domain, an important role is played by the **dual lattice** of $\Lambda(\boldsymbol{B})$, scaled by 2π, which we denote by $\Lambda(\boldsymbol{A})$. 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 for the dual lattice is $\boldsymbol{A} = 2\pi(\boldsymbol{B}^{-T})$.

The functions $(\boldsymbol{\omega} \to \exp(i\boldsymbol{u}^T\boldsymbol{\omega}) : \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$ form an orthonormal base of the space of periodic functions with periods in $\Lambda(\boldsymbol{A})$. Hence under the usual isometry $Z(\boldsymbol{x}) \leftrightarrow \exp(i\boldsymbol{\omega}^T\boldsymbol{x})$, the subspace generated by $(Z(\boldsymbol{u}); \boldsymbol{u} \in \Lambda(\boldsymbol{B}))$ corresponds to the space of periodic functions with periods in $\Lambda(\boldsymbol{A})$.

3. The main result

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. The average mean square error of $\widehat{Z}(\boldsymbol{x})$ is

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) = \frac{1}{\mathrm{vol}(\Omega(\boldsymbol{B}))} \int_{\Omega(\boldsymbol{B})} \mathbb{E}[(Z(\boldsymbol{x}) - \widehat{Z}(\boldsymbol{x}))^2] d\boldsymbol{x}.$$

Now consider lattices $\Lambda(\boldsymbol{B})$ with $\mathrm{vol}(\Omega(\boldsymbol{B})) = 1$ and scale each lattice by a parameter β, i.e. take $\Lambda(\beta\boldsymbol{B})$. Then for each β we can ask which lattice minimizes $\sigma^2(ave, \Lambda(\beta\boldsymbol{B}))$. Instead of considering a fixed covariance function and scaling the lattice, we can also take a fixed lattice and scale the covariance function: $R(\boldsymbol{x}) = R_0(\beta||\boldsymbol{x}||)$. For the spectral density f, this means $f(\boldsymbol{\omega}) = f_0(||\boldsymbol{\omega}||/\beta)$.

For a subclass of covariance functions or spectral densities to be defined later, we have the following results

β	$\sigma^2(ave, \Lambda(\beta \boldsymbol{B}))$	Optimality criterion
0	0	none
small	see later	dual packing radius
≈ 1	??	??
large	see later	packing radius
very large	$R(\mathbf{0})(1 - \int \rho(\boldsymbol{x})^2 d\boldsymbol{x}\, \beta^{-d})$	none
∞	$R(\mathbf{0})$	none

Here, ρ is the correlation function: $\rho(\boldsymbol{x}) = R(\boldsymbol{x})/R(\mathbf{0})$.

Example: d=3. Here the packing radius is maximized for the so-called face-centered cubic lattice which is obtained by adding the centers of the faces to the cubic lattice. It has $\tau = 12$ and $\rho = 2^{-5/6} = 0.5612\ldots$.

The dual lattice of the face-centered cubic lattice is the body-centered cubic lattice (the centers of the cubes are added). It has $\tau = 8$ and $\rho = 2^{-5/3}3^{1/2}$ $= 0.5456\ldots$.

Hence the optimal lattice depends on the sampling rate.

Generator matrices are (up to constants)

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0 \\ 0.5 & 0 & -0.5 \end{bmatrix}.$$

4. High rate sampling

First we have a formula for the average MSE in the frequency domain.

Theorem 1 *If Z has a spectral density f, then*

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) = \frac{1}{(2\pi)^d} \int \frac{f(\boldsymbol{\omega}) \sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\mathbf{0}\}} f(\boldsymbol{\omega} + \boldsymbol{\lambda})}{\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A})} f(\boldsymbol{\omega} + \boldsymbol{\lambda})} d\boldsymbol{\omega}.$$

For β going to zero, the peaks of the integrand on the right hand side become dominant which allows us to approximate the integral.

Theorem 2 *If $f(\boldsymbol{\omega}) = f_0(||\boldsymbol{\omega}||/\beta)$ with $f_0(r) \sim C\exp(-r^p) \quad (r \to \infty)$ then*

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \quad \sim \quad \left(\frac{2\pi}{p}\right)^{(d+1)/2} \beta^{-p(d+1)/2} \frac{\tau(\boldsymbol{A})}{4} \cdot f_0(\rho(\boldsymbol{A})/\beta)\rho(\boldsymbol{A})^{d+p(d+1)/2}.$$

For β small enough, the right hand side decreases as $\rho(\boldsymbol{A})$ increases. Hence, the optimal lattice maximizes $\rho(\boldsymbol{A})$.

Steps of the proof:

- $\sum_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{A}) \setminus \{\mathbf{0}\}} f(\boldsymbol{\omega} + \boldsymbol{\lambda})$ is dominated by the largest summand,
- the integrand is maximal for $\boldsymbol{\omega} = \hat{\boldsymbol{\lambda}}/2$ where $\hat{\boldsymbol{\lambda}}$ is one of the shortest non-zero vectors in $\Lambda(\boldsymbol{A})$ (that is $||\hat{\boldsymbol{\lambda}}|| = \rho(\boldsymbol{A})$).
- near such a point, the integrand is approximately equal to

$$\frac{f(\boldsymbol{\omega}) f(\hat{\boldsymbol{\lambda}} - \boldsymbol{\omega})}{f(\boldsymbol{\omega}) + f(\hat{\boldsymbol{\lambda}} - \boldsymbol{\omega})},$$

- Finally, use Laplace approximations.

5. Low rate sampling

Theorem 3 *If* $R(\mathbf{0}) = 1$ *and*

$$\sum_{\boldsymbol{u} \in \Lambda(\boldsymbol{B}) \setminus \mathbf{0}} |R(\boldsymbol{u})| < 1,$$

then

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

where

$$R^{*2}(\boldsymbol{x}) = \int_{\mathbb{R}^d} R(\boldsymbol{y}) R(\boldsymbol{x} - \boldsymbol{y}) d\boldsymbol{y},$$

$$\Delta^k(\boldsymbol{u}) = \begin{cases} \sum_{\boldsymbol{u}' \in \Lambda(\boldsymbol{B}) \setminus \mathbf{0}} \Delta^{k-1}(\boldsymbol{u} - \boldsymbol{u}') R(\boldsymbol{u}') & (k \geq 1) \\ 1_{\{\boldsymbol{u} = \mathbf{0}\}} & (k = 0) \end{cases}$$

This is an analogue of the usual kriging formulae for countably many observations, using the "infinite matrix inversion formula"

$$(I - \Delta)^{-1} = \sum_{k=0}^{\infty} \Delta^k.$$

(Under the conditions above, this is well-defined).

In the infinite sum, the leading term is the one for $k = 0$, i.e.

$$\sigma^2(ave, \Lambda(\boldsymbol{B})) \approx 1 - R^{*2}(\mathbf{0}),$$

which is independent of the lattice.

The next terms are those for $k = 1$ and the one with $k = 2, \boldsymbol{u} = \mathbf{0}$. If R decays exponentially, sums over $\boldsymbol{u} \in \Lambda(\boldsymbol{B})$ are asymptotically equivalent to the largest summand.

Theorem 4 *If $R(\boldsymbol{x}) = R_0(\beta||\boldsymbol{x}||)$ with $R_0(r) \sim C\exp(-r^p) \quad (r \to \infty)$, and if $R(\mathbf{0}) = 1$, then up to terms of lower order*

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

where $\boldsymbol{e}$ is a unit vector. The asymptotic behavior of the right hand side depends on p, but it is always a decreasing function of $\rho(\boldsymbol{B})$. Therefore, the optimal lattice maximizes $\rho(\boldsymbol{B})$.

6. Cardinal interpolation

If the spectral density f is zero outside of $\Omega(\boldsymbol{A})$, then by Theorem 1 the interpolation MSE is zero and

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

where

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

does not depend on f. This is called **cardinal interpolation**.

One can use cardinal interpolation even if the field is not band-limited. How much does one loose over optimal interpolation ?

By analyzing the formula in Theorem 1, for isotropic and decreasing spectral densities, the efficiency of cardinal interpolation is at least 0.5.

For high-rate sampling, under the conditions of Theorem 2, the limiting relative efficiency of cardinal interpolation is $\frac{\pi}{4} = 0.785...$, independently of the lattice and the spectral density f.