CHALMERS
UNIVERSITY OF TECHNOLOGY

Exploring a space of materials: Spatial sampling design and subset selection

Exploring a space of materials: spatial sampling design and subset selection

Fred A. Hamprecht
IWR
University of Heidelberg
D-69120 Heidelberg, Germany
fred.hamprecht@iwr.uni-heidelberg.de

Erik Agrell
Dept. of Signals and Systems
Chalmers University of Technology
SE-41296 Göteborg, Sweden
agrell@s2.chalmers.se

Abstract

This chapter is concerned with the following experimental situation: a fixed number of experiments are to be used to learn as much as possible about a space of potential materials. The researcher is required to first commit him-/herself to the specific materials that are to be investigated before these are all evaluated simultaneously.

We start from the assumption that the response of interest (such as activity, toxicity, durability, permeability, etc.) shows some spatial correlation over the experimental space to be explored. This means that the evaluation of a specific material tells the researcher something about the properties of similar materials, where similarity decreases with distance in material space.

The lack of knowledge of the response surfaces—which motivates experimentation in the first place—is taken into account by using stochastic processes. We give quantitative expressions for the uncertainty in the response that persists after experiments will have been performed and discuss strategies to find designs that minimize this uncertainty.

Furthermore, we introduce lattices that become optimal under certain limiting conditions and use these to illustrate the nature of the space-filling designs that are derived from our assumptions.

Finally, we discuss both the case in which the specific materials to be investigated can be chosen arbitrarily as well as the case in which the samples must be selected from a finite set, e.g., an existing collection or library.

Throughout the chapter, special emphasis is put on the problems arising in multidimensional spaces. All required tools (stochastic processes, the best linear unbiased estimator and lattices) are introduced to make the chapter self-

contained.

Contents

13.1 Introduction 1
13.2 Modeling the response surface 1
13.2.1 Polynomial approximation 2
13.2.2 Stochastic processes 2
13.3 Kriging 5
13.4 Optimality criteria 7
13.4.1 Integrated mean square error 7
13.4.2 Maximum mean square error 8
13.5 Infinite experimental regions 9
13.5.1 Uniformly spaced points: lattices 9
13.5.2 Packing, quantizing, and covering problems 10
13.5.3 The geometry of Voronoi cells 14
13.5.4 Discussion of optimality criteria 20
13.6 Finite experimental regions 21
13.7 Extensions 23
13.7.1 Application to subset selection 23
13.7.2 Adaptive sampling 24
13.7.3 Relation to other algorithms 24
13.7.4 Technical details 25
13.8 Conclusions and practical recommendations 26
Appendix A Generator matrices for lattices 27
A.1 Generating lattice points 27
A.2 Generator matrices 28

13.1 Introduction

We assume you are facing the following situation: you wish to find a material that features the best possible value for some property which you can measure quantitatively. This is your **response**. You can control the composition as well as the formulation and synthesis conditions. These variables form the basis of your **experimental space**. We define the **experimental region** as that part of experimental space which you decide to explore in your quest for a better material.

Previous chapters have dealt with optimization of the response. This works fine once you *have* found an area in your experimental region in which the response is greater than zero. What, however, if you have not? Also, you may already have explored an area of your experimental region until you have found the local response maximum, but you may be dissatisfied with what you have found: you now wish to screen the remainder of your experimental region for other areas which show a nonzero response.

In both cases, you are obliged to sample your experimental region systematically, at least until you have found a new active area. A similar approach may be useful even to locate the response maximum within a known active area, in cases when the experiments are too time-consuming to admit a sequential design, e.g., if long annealing is required. This entire chapter is about how to perform this sampling most effectively in the sense of gaining the clearest possible picture of your response surface.

13.2 Modeling the response surface

To prevent the waste of time and money, we wish to sample space systematically in an optimal fashion. Fine, but what is "systematic" and "optimal"? Intuitively, we understand that, in the absence of detailed prior knowledge about the response surface, "systematic" means the samples should be distributed uniformly throughout the experimental region. However, different systematic sampling schemes are conceivable, and a great variety of these have been proposed in the literature (see references in [1]). Many have been brought forward with eloquent arguments, but as long as these are purely verbal, it is difficult to weigh one against the other. We advocate the use of methods which are derived from clearly (and mathematically) stated assumptions concerning the **response surface**. The name of the latter derives from the case of a two-dimensional experimental region. If the response is plotted in the third dimension for each point in the experimental region, the impression of a surface results. In this analogy, a steepest ascent algorithm can be compared to a hiker that chooses the steepest path to the nearest peak, which corresponds to a local response maximum. We will always talk of "surfaces," even in higher dimensions.

13.2.1 Polynomial approximation

The **design** is the sampling scheme, or the set of all points at which experiments are to be performed. The theory of optimal experiments [2] assumes that the response can be approximated with an analytical model. Once a model is fixed, those points are selected for the design that have the greatest statistical leverage, that is, those points which minimize the uncertainty in the estimates of the model parameters. Such optimality criteria go under the name of "alphabetic" criteria; namely, A-optimality (minimizing the trace of the inverse of the information matrix), D-optimality (maximizing the determinant of the information matrix), and G-optimality (minimizing the maximum prediction variance) [3].

The by far most popular models are polynomials, that is the sum of a constant plus a plane plus a parabola and so on. The difficulty with most implementations is that they consider only low-order polynomials as models. This is problematic if the response surface can be expected to be irregular, for instance featuring multiple peaks (statisticians refer to these as "modes") and valleys. A second-order polynomial can provide a good model for a single peak, but not for a chain of mountains. One consequence is that, especially in higher dimensions, most of the design points placed optimally by the above criteria lie on the hull of the experimental region while few points are selected within. This lack of representation of the inner region has started a quest for more "space-filling" algorithms.

In the following sections, we will describe a different model of the response surface that provides a lot of flexibility. Optimal sampling strategies will then be derived under that model.

13.2.2 Stochastic processes

Not knowing what the actual response surface looks like, we can model it using a stochastic approach. For starters, we could draw numbers from a random number generator, one for each of the points in the experimental region. The resultant surface would be pure noise. If you have reason to believe that this is a good model for the particular response surface you are studying, you may skip this chapter: the best you can then do is to avoid sampling the same point in experimental space twice, but other than that, all conceivable designs will be equally good, on average.

In reality, however, many response surfaces exhibit spatial correlation. That is, one can expect two proximate points in the experimental region to feature similar responses. Exceptions do certainly occur, for instance in the case of sharp phase transitions, but even then the remainder of the experimental region will usually show some spatial correlation. This "similar property principle" applies to a wide range of systems and we assume it to hold in the following.

What we want, then, is a mathematical tool that can exhibit a range of spatial correlations while being random in character—because the true response surface is not known. In other words, we wish to represent in our model the expected **smoothness** of the unknown response surface (which may be guessed or inferred

from previous experiments on similar systems), but not the exact location of the response minima and maxima. Once such a model is found, we can derive designs that are optimal for it.

A suitable technique which satisfies our requirements is a **stochastic process**, also known as random field or random process. Such processes arise in the description of systems that evolve in space subject to probabilistic laws. We do not claim that a response surface is probabilistic in nature, but we do assume that a response surface looks similar to a suitably chosen stochastic process.

Before we illustrate these concepts graphically, we need a little more terminology: consider a random number generator from which random numbers can be drawn, one at a time. These individual numbers are called **realizations**. In the case of a particular stochastic process, an entire random surface (and it can generate infinitely many of these) is just *one* realization. These realizations will have their peaks and valleys at different positions, but they will also have one thing in common: the smoothness.

Figure 13.1 shows two realizations of each of four different stochastic processes. The surfaces are obviously all different, but pairs of them appear similar: they share the same smoothness. The smoothness of the surfaces is governed by the range[1] of the covariance functions as well as by the behavior of the covariance functions near the origin. A cusp as in the exponential covariance function leads to surfaces that are much rougher on a short scale.

It is now time to quantify this smoothness and the class of stochastic processes we use. First, consider the mean of the stochastic process $Z(x)$ at position x,

$$\bar{Z}(x) = E[Z(x)]$$

where E denotes expectation. One obtains the mean as one averages the response at a specific position x over all possible realizations of a stochastic process. In the following development, we will assume this mean to be a constant $\bar{Z}$ over the experimental region. Next, consider the **covariance** between two points x_1, x_2, given by

$$\mathrm{Cov}(x_1, x_2) = E[(Z(x_1) - \bar{Z})(Z(x_2) - \bar{Z})].$$

The covariance function describes how similar the response is at positions x_1, x_2, averaged over all possible realizations. To make things easier, we assume the covariance function to depend only on the length and orientation of the vector connecting x_1 and x_2. This assumption and the one of a constant mean basically express the belief that the response surface does not change in character throughout the experimental region. It can still exhibit wild cliffs, deep canyons and vertiginous peaks, but the character of the response surface landscape should be homogeneous throughout. Mathematically, these assumptions are denoted

[1]The range indicates the interval over which the covariance function differs "substantially" from zero. This definition is admittedly vague and conventions in the literature differ. In the calculations shown here, the exponential and Gaussian covariance functions with range ρ are given by $\mathrm{Cov}(x_1, x_2) \sim \exp(-||x_1 - x_2||/\rho)$ and $\mathrm{Cov}(x_1, x_2) \sim \exp(-||x_1 - x_2||^2/\rho^2)$, respectively.

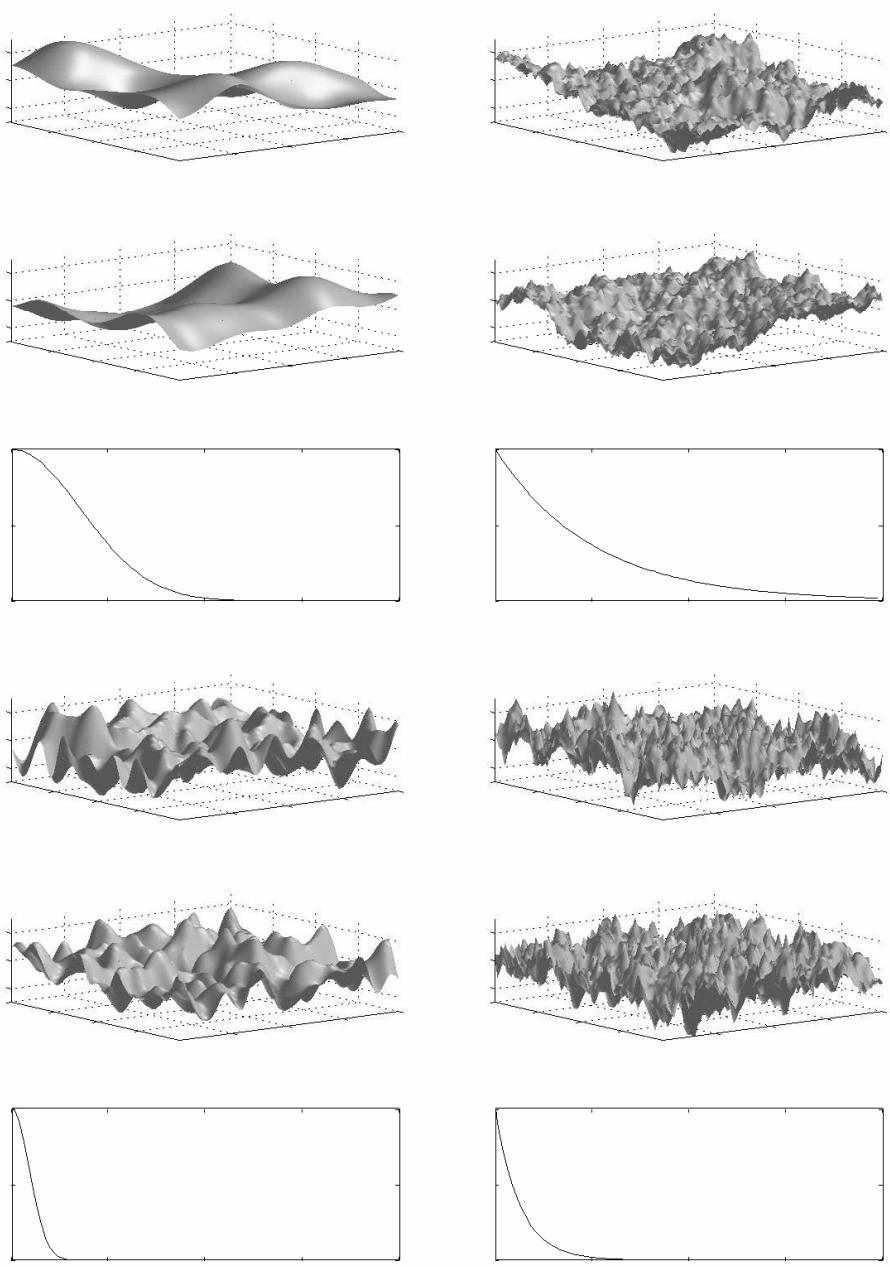


Figure 13.1: Stochastic models for response surfaces. Shown are two realizations each of four stochastic processes with their corresponding isotropic (circularly symmetric) covariance functions. Left: processes with a Gaussian covariance function. Right: processes with an exponential covariance function. Top: long range. Bottom: short range. The smoothness is governed by the covariance function: its degree of differentiability around the origin determines the smoothness at a microscopic scale, its decay with distance governs the appearance of mountains and valleys at a large scale. All simulations performed with gstat [4].

second-order (or weak or wide-sense) **stationarity.**[2] As a final simplification, we may assume that the covariance function depends only on the length, but not on the orientation, of the vector connecting x_1 and x_2. Such covariances are called **isotropic** and the phenomena they describe admit, on average, no discernible orientation.

Summarizing this section, we recommend to follow arguments for the selection of experimental designs that are directly derived from clearly stated models for the response surface. We have quoted two such models: the polynomial approximation, leading to the alphabetic criteria; and stochastic processes, for which we will derive optimality criteria in the following sections.

13.3 Kriging

The central assumption now is that the response surface can be modeled as the realization of a second-order stationary stochastic process. Bear with us while we introduce an interpolation method that will, in turn, lead to criteria for optimal designs.

Actually, an interpolator has some interest in its own right: given a number of measured responses at some points in the experimental region, it will give an estimate of the response in-between these points. Reflecting the importance of this task, a large number of methods have been proposed: polynomials, splines, B-splines, thin-plate splines, etc. The **kriging**[3] or **best linear unbiased estimator** stands out because it provides not only an estimate of the response surface throughout the experimental region, but also the uncertainty of that estimate, as a function of space: the uncertainty is low around the available data and grows with distance from these. Just how quickly the uncertainty grows with distance depends on the smoothness of the surface, see Fig. 13.2.

The simple kriging estimator $\hat{Z}(x)$ of the true response surface $Z(x)$ is given by

$$\hat{Z}(x) = \sum_{i=1}^{n} \lambda_i(x) Z(y_i),$$

where y_i are those points at which experiments have been performed, that is, the design; and $\lambda_i(x)$ are optimal weights, one for each of the n measurements $Z(y_i)$.[4]

[2]Fixing the first two moments, the mean and the covariance, is not very restrictive in the sense that there is an infinite number of different stochastic processes with the same first two moments. Generally, all finite-dimensional moments [5] are required to characterize a stochastic process exhaustively. Gaussian processes (which were used in Fig. 13.1) are an exception because they are completely determined by their first two moments.

[3]This interpolator has been developed independently in different disciplines, but has probably gained the greatest popularity in the mining community that is compelled to extract a maximum of information from each of their expensive samples. The name honors the mining engineer D. G. Krige [6].

[4]We will in the following assume that $\bar{Z} = 0$. To estimate a process $Y(x)$ with nonzero, but constant and known, mean, simply subtract the mean from every measurement, apply the regular kriging estimator to the obtained zero-mean process $Z(x) = Y(x) - \bar{Y}$, and add $\bar{Y}$ again to $\hat{Z}(x)$ to obtain an unbiased estimate $\hat{Y}(x)$.

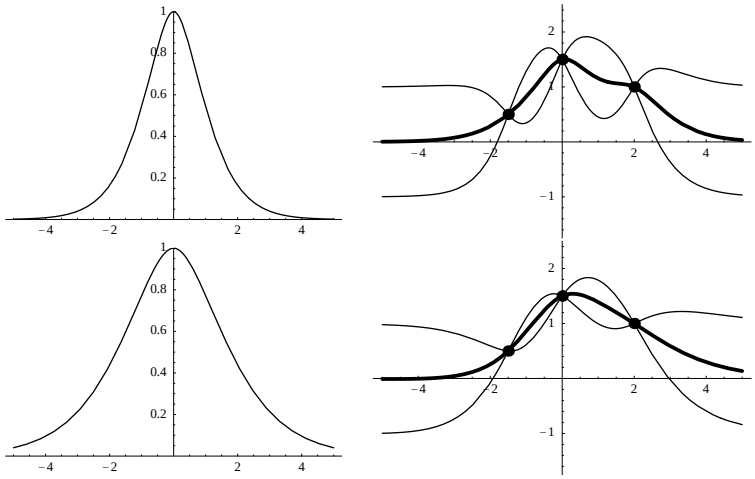


Figure 13.2: Simple kriging example. The left side shows two covariance functions that differ in their range. The right side shows an interpolation (bold line) of the three bold points. Far from the data, the interpolation goes to the assumed mean, in this case zero. The thin lines indicate the uncertainty of the interpolation in terms of $\pm\sqrt{MSE(x)}$, given by eq. 13.2. The estimates lie within the marked intervals with a likelihood of 68 %. The uncertainty grows faster for the covariance function with the shorter range.

Note that these weights vary over space (otherwise the estimate $\hat{Z}(x)$ would be a constant). They are found as the solution of a set of n linear equations

$$\sum_{i=1}^{n} \lambda_i(x)\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(x, y_j) \qquad j = 1, \ldots, n \tag{13.1}$$

depending on the covariance function of the stochastic process and are optimal in the sense of minimizing the mean square error [5, 1] between the true and the estimated response surfaces,

$$MSE(x) = E\left[(\hat{Z}(x) - Z(x))^2\right].$$

Again, the mean square error of an interpolator should be minimal on *average*, that is, considering all possible realizations of the stochastic process.

We have stated previously that the growth of the uncertainty of the interpolation with the distance from design points depends on the smoothness of the surface. This smoothness enters through the covariance function in equation 13.1 for the optimal weights $\lambda_i(x)$.

The mean square error of the kriging estimator, or the uncertainty of the interpolation as a function of space, is given [5, 1] by

$$MSE(x) = \mathrm{Cov}(x, x) - \sum_{i=1}^{n} \lambda_i(x)\mathrm{Cov}(y_i, x). \tag{13.2}$$

What does this equation depend on? On the covariance function, the design y_i, and the optimal weights which, in turn, depend only on the covariance function

and y_i. If you ponder this, the all-important conclusion is that the uncertainty in the interpolation depends on the measurement sites (the design) and the covariance function, but not on the measured values $Z(y_i)$.

As a consequence, the design can now be optimized so as to minimize the mean-square error, or the uncertainty in the interpolation, *before* performing any experiments. The catch is in the assumptions: the estimated uncertainty is only correct when the "true" covariance function is used.

13.4 Optimality criteria

Equation 13.2 gives the expected error for location x. The design can then be modified to control the uncertainty at that specific location. We wish, however, to find a design that is optimal for the *entire* experimental region $\mathcal{R}$. We will introduce two natural measures of global performance that have been proposed [7], then illustrate in which sense they differ fundamentally (sections 13.5.2, 13.5.3), and finally argue which of the two should be used (section 13.5.4).

13.4.1 Integrated mean square error

This criterion aims to minimize the total mean square error, integrated over the entire experimental region:

$$\text{Min} \int_{\mathcal{R}} MSE(x)\, dx \tag{13.3}$$

Given a particular experimental region (see section 13.1) and covariance function, this criterion or objective function can be optimized with any deterministic (e.g., steepest descent) or stochastic (e.g., simulated annealing) optimization method[5]. Algorithmic details and several examples of such an optimization are given in [1].

Fig. 13.3 shows a square experimental region and a design of 121 points that has been optimized under eq. 13.3 using an exponential covariance function with a range of one-fifth of the square's side-length. The design obviously reflects the particular shape of the experimental region under study, but its inner points seem to have a will of their own: they are arranged in a hexagonal pattern which does not match the shape of the boundary! This might lead us to hypothesize that in two dimensions, the "intrinsic" structure of a design that minimizes the integrated mean square error may be hexagonal.

To find out whether this hypothesis is of any value, we should make the experimental region larger and larger and add more and more points to it to keep the density of the design constant. In this way, the points at the interior can be expected to be influenced less and less by the particular shape of the experimental region.

[5] If an opinion exists as to which parts of the experimental region should be sampled most intensely, this prior belief can be allowed for through a spatial weighting of the mean square error [1].

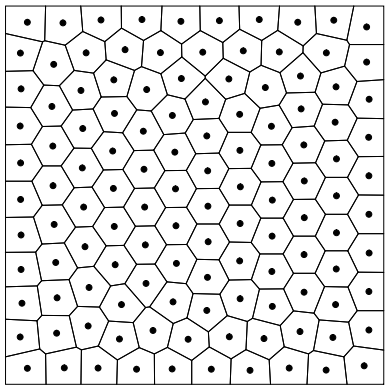


Figure 13.3: Design of 121 points for a square region, optimized under the integrated mean square error.

Ultimately, and to answer the question in full generality, the experimental region and the number of points should become infinitely large. This assumption greatly facilitates a mathematical analysis. We state, without proof, the following results:

- if the response surface becomes infinitely smooth or infinitely rough, it does not matter which sampling design is used
- if the response surface is very rough, the best design is given [8] by the densest sphere packing (see section 13.5.1)
- if the response surface is very smooth, the best design is given by the Fourier transform of the densest sphere packing [9].

These notions will be introduced in section 13.5.1, but to satisfy your curiosity, we will disclose the result that in two dimensions, both the best sphere packing and its Fourier transform are indeed given by a hexagonal arrangement and our inspection of Fig. 13.3 has thus allowed us to anticipate some deep results about the inherent characteristics of optimal designs in general.

13.4.2 Maximum mean square error

Whereas the integrated mean square error criterion discussed above strives to minimize the *average* ignorance, the maximum mean square error criterion aims to minimize the *maximum* interpolation error in the experimental region. A brute-force recipe to find a design which minimizes the maximum mean square error consists in identifying that location in space which has the greatest error, then perturbing the design, finding the location which now has the greatest error, etc.

If the stochastic process is assumed Gaussian and the response surface is very rough, this criterion becomes almost equivalent [10] to the problem of thinnest sphere covering which is introduced in the next section.

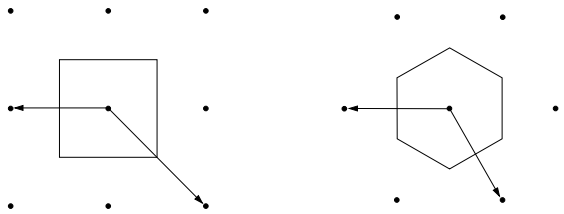


Figure 13.4: The cubic lattice Z_2 and the hexagonal lattice A_2 along with their Voronoi cells and one possible choice of basis vectors.

13.5 Infinite experimental regions

In a thought experiment, we have previously considered an infinitely large experimental region, to be explored with an infinite number of points. This is a trick to make boundary effects negligible and learn something about the inherent character of the different design criteria. We wish to exploit this trick in the present section to accentuate the differences between using integrated and maximum mean square error.

In section 13.6 we will turn our attention back to the practically more relevant case of a finite experimental region.

13.5.1 Uniformly spaced points: lattices

Intuitively, we can expect an optimal design on a perfectly homogeneous infinite space to be very homogeneous itself. The most regular arrangement of points conceivable is a **lattice**. Mathematically, a lattice consists of all linear combinations of some basis vectors, with all coefficients assuming integer values only (see Fig. 13.4). This yields a highly regular structure with the property that, with an oft-cited observation, "if you sit on one lattice point and view the surrounding set of lattice points, you will see the identical environment regardless of which point you are sitting on" [11]. In this sense, all lattices are perfectly uniform arrangements: there is not one part of space where the points lie closer together than in any other part.

The first dimension offers not much variety when it comes to lattices: there is only one way of arranging points at equal intervals, and the resultant lattice is called Z_1. (We follow approximately the terminology of [12], where the first letter indicates the family and the number gives the dimension.) In contrast, an infinite number of lattices can be conceived for any dimension higher than one. In two dimensions, two fundamental lattices have emerged as particularly interesting, the **cubic** lattice Z_2 and the **hexagonal** lattice A_2, see Fig. 13.4.

Fig. 13.4 also displays a set of basis vectors for each lattice. A lattice is fully specified through its basis vectors, but the converse is not true: the same lattice can be generated from different sets of basis vectors. If the basis vectors are collected into a matrix, we obtain a **generator matrix** for the lattice. Examples of generator matrices for Z_2 and A_2 are

$$\begin{bmatrix} -2 & 0 \\ 2 & -2 \end{bmatrix} \text{ and } \begin{bmatrix} -2 & 0 \\ 1 & -\sqrt{3} \end{bmatrix}$$

where the rows represent the same basis vectors as those shown in Fig. 13.4.

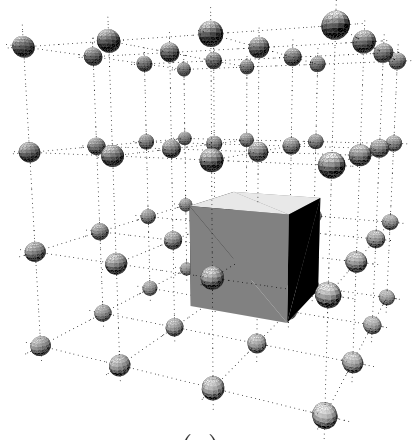
Also shown in the figures are the so-called **Voronoi cells** of the lattices [13, 14]. The Voronoi cell of a point in a point set comprises the part of space that is closer to this point than to any other. Contrary to the case of general point sets, there is only one type of Voronoi cells in a lattice, i.e., each Voronoi cell can be superimposed onto any other by translation only, without rotation or reflection.

When we move on to three-dimensional lattices, three basic lattices deserve particular attention, all of them well known from crystallography. They are illustrated in Fig. 13.5. The **cubic** lattice Z_3 simply consists of all three-dimensional points with integer coordinates. If we remove every second point of the cubic lattice as in Fig. 13.5 (b), the **face-centered cubic**, or fcc, lattice is obtained. (Alternatively, it can be constructed by centering a point on each of the six faces of a cube and stacking such cubes, hence the name.) And if we instead remove three fourths of the points in a cubic lattice, leaving only those with either all even or all odd coordinates, we obtain the **body-centered cubic**, or bcc, lattice, depicted in Fig. 13.5 (c). Even though the point sets in themselves may not look very different, their geometrical properties differ substantially, as illustrated by the shapes of their Voronoi cells. In standard lattice notation, the face-centered cubic lattice is denoted A_3 and the body-centered cubic lattice A_3^*. Their generator matrices, as well as their generalizations and other lattices in higher dimensions, are deferred to the Appendix.

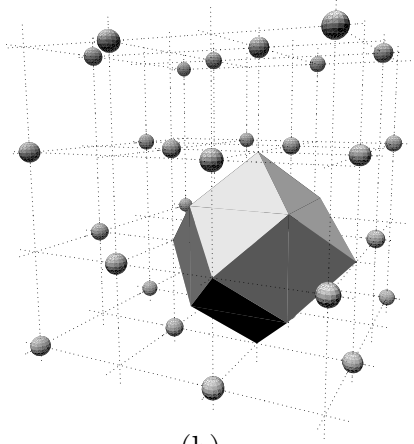
13.5.2 Packing, quantizing, and covering problems

Packing

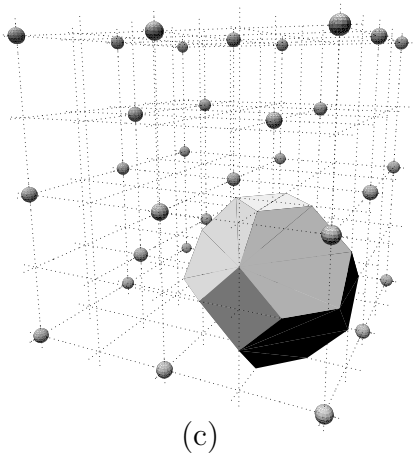
In section 13.4.1 we have claimed that, in the case of an infinite design region and a rough response surface, the integrated mean square error is minimized by a design that corresponds to a best sphere packing. The **sphere packing** problem asks for the densest possible way of arranging non-overlapping spheres of equal size. The density becomes maximal when the volume of the interstitial regions is minimized. In lattices, the multiple translations of a single Voronoi cell tile all of space. As a consequence, it is sufficient to study a single Voronoi cell to understand the properties of the entire lattice. To reduce interstitial space, it is advantageous to have a Voronoi cell that is as similar as possible to a sphere. Since spheres do not tile space in dimensions greater than one, the best approximation is sought. If we scale all lattices such that their Voronoi



(a)



(b)



(c)

Figure 13.5: The three basic three-dimensional lattices and their Voronoi cells. (a) The cubic lattice, whose Voronoi cell is simply a cube. (b) The fcc (face-centered cubic) lattice. Its Voronoi cell is a rhombic dodecahedron and has 12 faces. (c) The bcc (body-centered cubic) lattice. Its Voronoi cell is a truncated octahedron and has 14 faces.

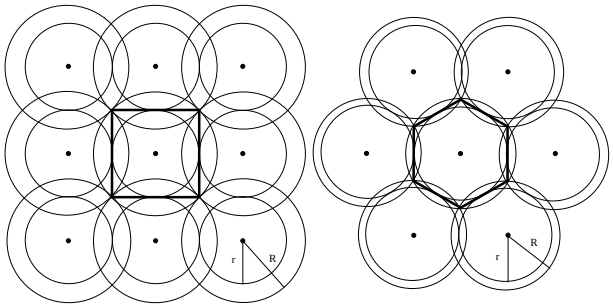


Figure 13.6: The cubic and the hexagonal lattice with the largest inscribed and smallest circumscribed spheres of their Voronoi cells. If both lattices are scaled such that their Voronoi cells have unit volume, the hexagonal lattice has the greater packing radius r and the smaller covering radius R, cf. section 13.5.2 and Fig. 13.7.

Dimension	packing	best quantizing lattice	covering
1	Z_1	Z_1	Z_1
2	A_2	A_2	A_2
3	A_3 (fcc)	A_3^* (bcc)	A_3^*
4	D_4	D_4	A_4^*
5	D_5	D_5^*	A_5^*
6	E_6	E_6^*	A_6^*
7	E_7	E_7^*	A_7^*
8	E_8	E_8	A_8^*

Table 13.1: The best known lattices according to three criteria. A lattice L^* is the dual, or Fourier transform, of the lattice L. Z_d is the d-dimensional cubic lattice and A_d is the generalized hexagonal lattice. These and the other lattice families are defined in the Appendix.

cells have unit volume, the best packing lattice will be the one with the largest inscribed sphere[6]. Its radius, the **packing radius** r, becomes the figure of merit for the packing problem: the greater r, the denser can a lattice pack spheres, see Fig. 13.6

As we go to higher dimensions, we encounter an ever increasing number of potentially useful lattices, while at the same time the evaluation of their characteristics becomes more complex. Fortunately, a great amount of research has been devoted to the search for good lattices and the results are tabulated for several quality measures and for all dimensions that are relevant for our purposes [12]. The best packing, quantizing, and covering (see below) lattices known for dimensions up to eight are summarized in Table 13.1. Definitions of all lattices mentioned in the table, in terms of their generator matrices, are found in the

[6]Lattices are not the only possible method of stacking spheres densely, but the best lattices are generally either the best packing method known or not much worse than the best known.

Appendix.

Quantizing

Work on the optimal sampling problem in signal processing [9] has led to the conclusion that for very smooth response surfaces in an infinite experimental region, the integrated mean square error is minimized by a design that is the dual[7] of a best sphere packer.

A glance at Table 13.1 shows that the duals of the best packing lattices can be found[8] in a column entitled "quantizing." The **quantizing problem** (or vector quantization), which owes its name to an application in digital communications, seeks an arrangement of points such that their Voronoi cells become as compact as possible, where compactness is measured by their average second moment. Indeed, the intriguing fact that the best known lattices for packing and quantizing are duals of each other has been observed outside the range of the table, too, and it was earlier conjectured to be a generally valid relation between optimal lattices in any dimension [12]. This conjecture does not appear to be true for arbitrary dimensions [15], but the general tendency is still valid: the dual of a good packing lattice is good for quantization and vice versa [16]. We will have reason to recall this property in section 13.6.

As before, if we restrict ourselves to lattices, it is sufficient to study a single Voronoi cell; and also as before, the most compact body of a given volume according to this measure is—the sphere. Yet again, tiling space with spherical cells is not possible and thus one seeks a lattice whose Voronoi cells are as nearly spherical as possible, by the above measure, while still tiling space. The figure of merit now is $\bar{r}$ which gives the root of the mean squared Euclidean distance of all points in a Voronoi cell to its center.

Covering

Finally, we noted (section 13.4.2) that, at least for rough Gaussian[9] stochastic processes, for an infinite experimental region, the maximum mean square error is minimized by a design that offers a **thinnest covering**. In this problem, the aim is to seek an arrangement of spheres of equal size such that each point in space is covered by at least one sphere, while the average number of spheres covering a point in space should be as low as possible. Arguing in terms of lattices, the radius of the smallest circumscribed sphere around the Voronoi cell, the **covering radius** R, should become minimal, see Fig. 13.6.

We have seen that all criteria, packing, quantizing, and covering, seek to find lattices with as spherical a Voronoi cell as possible; however, different definitions of sphericity are used and the following section will show that these lead to wildly differing lattices.

[7] The dual of a lattice is essentially its Fourier transform, if the point set is regarded as a multidimensional sum of Dirac impulse functions.

[8] The lattices Z_1, A_2, D_4, and E_8 are self- or isodual.

[9] Gaussian here refers to the finite-dimensional distribution, not to the covariance function.

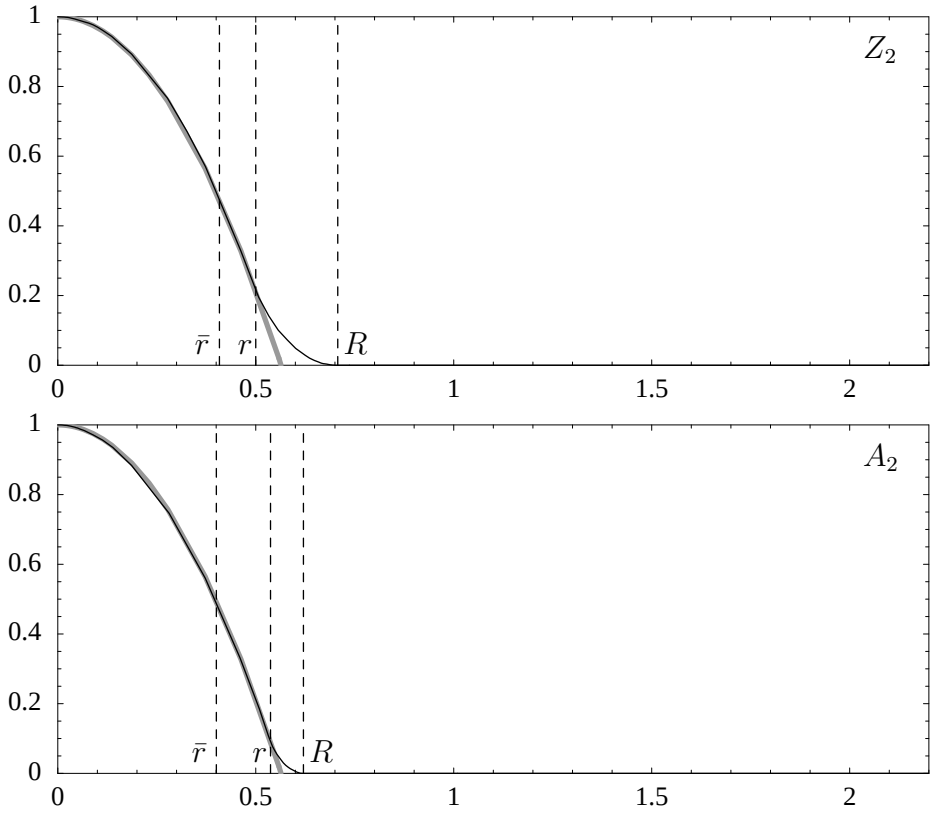


Figure 13.7: The solid curves show the radial distribution, i.e.,the percentage of the Voronoi cells of the two-dimensional lattices Z_2 and A_2 (see Fig. 13.4) that lies outside a given radius. Dashed lines indicate the figures of merit for the packing (packing radius r), quantizing (root mean square distance $\bar{r}$) and covering (covering radius R) problems. The gray curve illustrates the radial distribution of a circle with the same area as the considered Voronoi cells.

13.5.3 The geometry of Voronoi cells

The by far most popular multidimensional lattice in experimental design, as well as in a wide range of other applications, is the cubic lattice, Z_d. It is what one obtains by simply allowing the same discrete set of equidistant values for all variables, independently of each other. While this property makes it extremely easy to generate and utilize this lattice, the lattice is by no means a good lattice in the sense of the previous section. Note that it appears nowhere in Table 13.1 for dimensions higher than one. Figures 13.4—13.6 suggests why: its Voronoi cell is a square, in general a hypercube, which is not a very good approximation to a sphere.

The most important characteristic of a point in a Voronoi cell is its "radius," i.e., its distance to the center of the cell—the direction is irrelevant for the performance, as long as a rotation-invariant figure of merit is used. Hence, the performance of a lattice is fully determined by the distribution of the radii in the Voronoi cell. We define the **radial distribution** of a body as the percentage of

the volume that lies outside a circle of a certain radius. For lattice Voronoi cells, this is equivalent to the likelihood that the distance between a random point in multidimensional space and its closest lattice point is greater than a certain value. A uniform distribution is assumed over a large enough region to make boundary effects negligible.

To prepare ourselves for a journey into higher dimensions, we begin with the radial distributions of the Voronoi cells of the two-dimensional lattices of Figure 13.4. The corresponding functions are shown in Figure 13.7. The density of both lattices have been normalized to one lattice point per unit volume. The three quality measures r, $\bar{r}$, and R (see section 13.5.2) are marked in the diagrams for both lattices.

If we look for a d-dimensional body of unit volume with greatest inscribed sphere (largest r), smallest moment of inertia (smallest $\bar{r}$) or smallest circumscribed sphere (smallest R) without requiring that it allow a tiling of space, we find that the sphere is optimal by all criteria. Hence we include in the diagrams the corresponding curve for the distance between points in a sphere and its center. This curve serves as a lower bound in the diagrams. A good lattice, in the sense that it has as much as possible of the Voronoi cell located close to its center, would in these diagrams be identified by its proximity to the spherical lower bound. As we can see in Figure 13.7, the curves for the two lattices differ only in their "tails." They follow the same shape down to the packing radius r of Z_2. From there on, A_2 has a steeper slope, reflecting the rounder shape of its Voronoi cell.

Figures 13.8–13.10 show the same type of curves for higher dimensions. For dimensions 4, 8, and 16, the best lattices for packing, quantizing, and covering, respectively, are illustrated, along with the cubic lattice. Again, the best lattice would have a curve that in some sense lies as close as possible to that of the sphere. The curves coincide with that of the sphere for distances below their packing radii r and above their covering radii R, respectively. They deviate between these points, where the cubic lattice always shows the most prominent "tail." The ratio R/r is equal to $\sqrt{d}$ for Z_d and thus tends to infinity as the dimension increases. "Good" lattices, on the other hand, typically have a ratio less than two for any dimension. In particular, this is always true for the optimal packing lattice.[10]

In all cases, in particular for the cubic lattice, the curves approach zero rapidly long before the theoretical zero, which occurs at the covering radius R. This may be somewhat surprising in view of the large number of vertices that a typical Voronoi cell possesses.[11] Apparently, the volume close to such a vertex is so small that even the sum of them accounts for a negligible volume only.

One might think of a high-dimensional Voronoi cell as being similar to a sea urchin: a spherical shape with a large number of thin needles on its surface. Such

[10]Because for any lattice with $R \geq 2r$, a new lattice with the same r but higher point density can be created by inserting points in the void between points in the original lattice.

[11]The Voronoi cell of any d-dimensional lattice has between 2^d and $(d+1)!$ vertices, inclusively, where the lowest value is attained by Z_d and the highest by A_d^* [17].

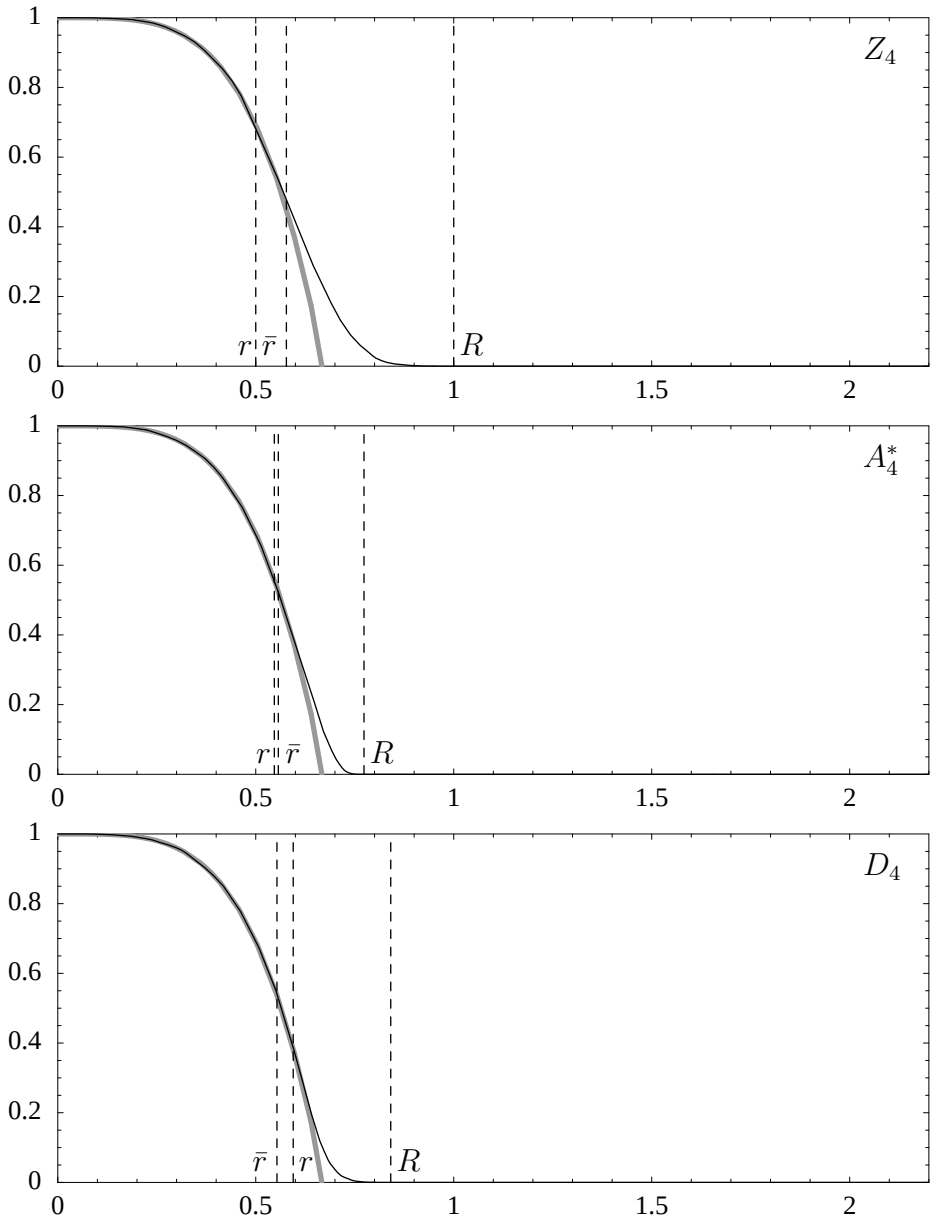


Figure 13.8: The radial distribution of Z_4, A_4^* (the best known lattice for covering), and D_4 (the best known for packing and quantizing). The gray curve represents a four-dimensional sphere.

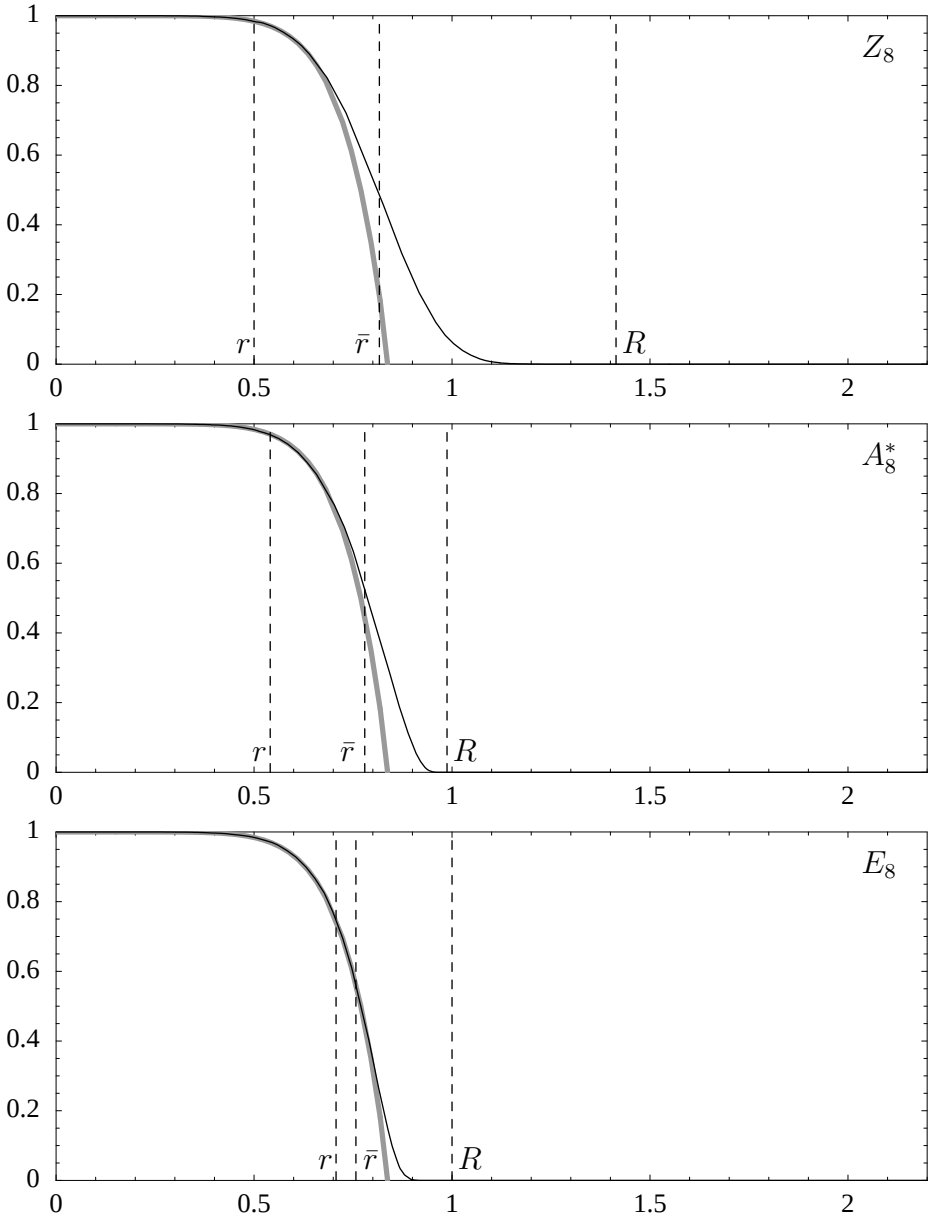


Figure 13.9: The radial distribution of Z_8, A_8^* (the best known lattice for covering), and E_8 (the best known for packing and quantizing). The gray curve represents a sphere.

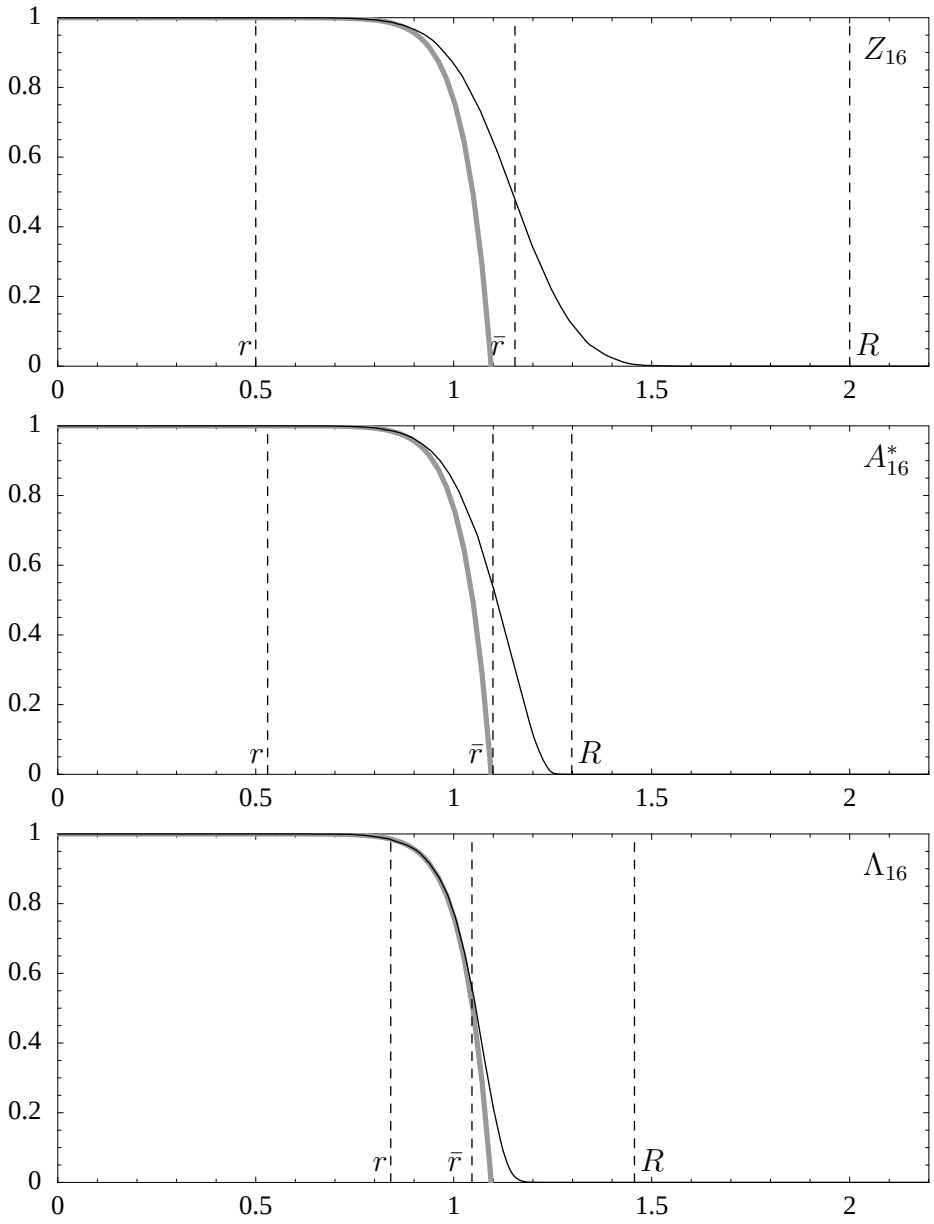


Figure 13.10: The radial distribution of Z_{16}, A_{16}^* (the best known lattice for covering), and Λ_{16} (the best known for packing and quantizing). The gray curve represents a sphere.

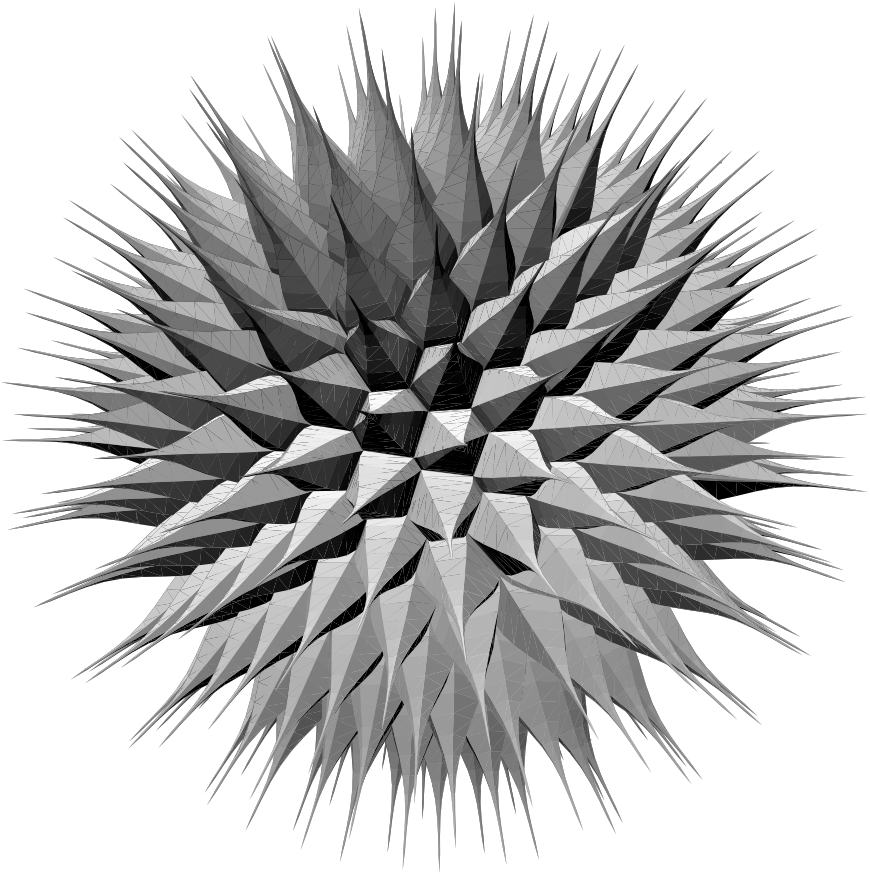


Figure 13.11: A three-dimensional representation of an eight-dimensional hypercube. The body has the same radial distribution and the same number of vertices as the hypercube. A very small fraction of the mass lies near a vertex. Also, most of the interior is void.

a shape is illustrated in Figure 13.11, which is a nonlinear but radius-preserving mapping from an eight-dimensional hypercube to three-dimensional space. We observe that the spines are extremely thin, indicating that the space close to a vertex accounts for a negligible fraction of the volume of a hypercube. On the other hand, the shape is also almost empty in the center. This behavior is typical for high-dimensional polytopes, not only hypercubes. Most of the mass is concentrated on a belt with relatively small radial variation.[12]

13.5.4 Discussion of optimality criteria

We have characterized the integrated and maximum mean square error criteria by means of the lattices that they lead to when the experimental region becomes infinitely large and the response surface is very smooth or very rough. Figures 13.8–13.10 show that almost all of the volume of a Voronoi cell is confined to a radius that is much smaller than the distance of the vertices from the center. Repeating our metaphor, illustrated in Figure 13.11, the corners of a Voronoi cell in a higher dimension take the shape of needles and their volume in experimental space is low. If the probability of finding an interesting area of high response is uniform throughout the experimental region, it is unlikely that it will come to lie in one of the needles. For covering (and thus the maximum mean square error criterion), however, all that matters is the distance of the tips of these needles from the center, irrespective of what little volume they occupy. The integrated mean square error criterion leads to Voronoi cells with longer needles. This criterion pays no heed to the large interpolation uncertainty near the tip of these needles, simply because their statistical weight is so low.

Also, in the sense of the integrated mean square error criterion, *any* additional design point will reduce the uncertainty of the experimenter (as long as it does not coincide with a previous point). By the maximum mean square error criterion, however, only *optimally* placed points reduce the uncertainty if the range of the covariance function is short.[13]

For these reasons, we encourage the use of the integrated mean square error criterion, assuming that the top priority in most applications is the best overall performance throughout the experimental region. High-dimensional lattices are however rarely useful in experimental design unless an extremely large number of experiments can be performed, as explained in the next section.

[12]For d-dimensional hypercubes, the radius of the belt approaches $R/\sqrt{3}$ as d increases.

[13]The reason being that for short ranges, the maximum mean square error criterion becomes similar to the sphere covering problem. The maximum mean square error then depends on the point (or points) being most remote from its closest design point. Only additional design points that reduce this greatest minimal distance lead to a reduction in maximum mean square error.

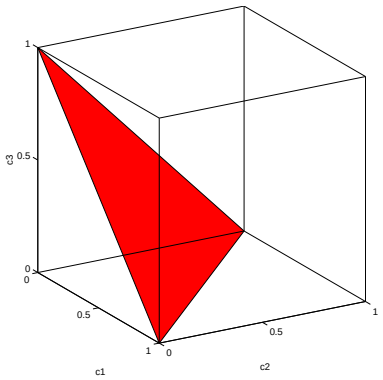


Figure 13.12: A cubic experimental region $0 \leq c_1, c_2, c_3 \leq 1$ and the regular triangle containing all ternary mixtures with $c_1 + c_2 + c_3 = 1$.

13.6 Finite experimental regions

Having learned more about the quintessential differences between the integrated and maximum mean square error criteria in the admittedly somewhat academic setting of infinite experimental regions, we are prepared to study the practically relevant case of finite experimental regions. Owing to the "curse of dimensionality," these have their peculiarities in higher dimensions.

If you can modify all of your experimental variables such as composition, formulation, and synthesis conditions independently, and if you set upper and lower limits individually for each of these variables, your experimental region is a rectangular body. For instance, if you can choose between 0 and 1 unit (milligram, mol, etc.) of any of the three components c_1, c_2, and c_3, the associated experimental region is a cube in the first octant of experimental space, see Fig. 13.12. If dependencies between the variables exist, the shape of the experimental space changes. The most common case is the so-called **mixture problem** in which all the components must add up to one unit. Geometrically, the condition $c_1 + c_2 + c_3 = 1$ is a plane that passes through the points $(0, 0, 1)$, $(0, 1, 0)$, and $(1, 0, 0)$. If one intersects this plane with the cube $0 \leq c_1, c_2, c_3 \leq 1$, the resultant experimental region is an equilateral triangle (see Fig. 13.12), often used for summarizing data on ternary mixtures. If one has a mixture problem with four components, the experimental region becomes a regular tetrahedron, and so on. The geometrical entities line segment (in dimension $d = 1$), equilateral triangle ($d = 2$), regular tetrahedron ($d = 3$), etc., are summarily denoted as d-dimensional **simplices.**

How do these polyhedra generalize to higher dimensions? We will look at cubic and simplicial experimental regions in higher dimensions.

For starters, imagine that you can set d different reaction parameters (temperature, pressure, etc.) to k different values each. You can then choose from k^d distinct sets of reaction conditions, see Fig. 13.13. For example, $d = 8$ parameters which are set to only $k = 3$ values each result in 6561 distinct combinations. How many of these lie on the surface of your experimental region? The answer is 6560 because only one point, namely, the one where all eight parameters assume

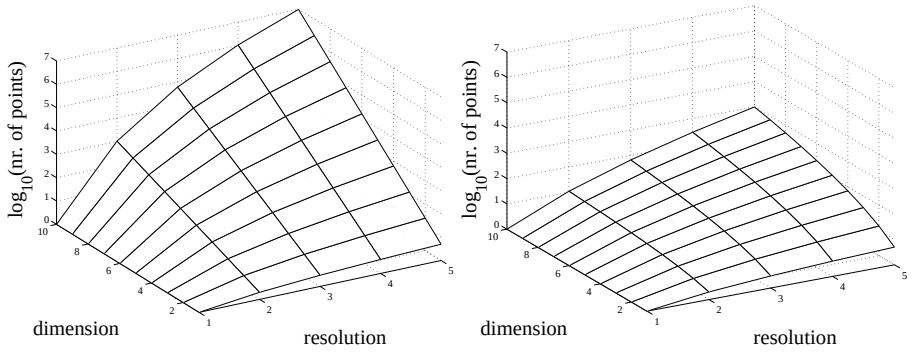


Figure 13.13: The logarithm of the number of points in cubes (left) and simplices (right) of various dimensions with the number of points on one edge given by "resolution." Points are arranged on the cubic lattice Z_d in the cubes and the generalized hexagonal lattice A_d in the simplices.

their central value, lies in the interior.

Similar properties hold for the simplices: if d components can be varied under the condition that they must sum to some constant, all possible combinations lie in a $(d-1)$-dimensional simplex. Assuming that the individual components can be varied in k steps, the total number of discrete combinations is $(k+d-1)!/((k-1)!d!)$. Again, most points will lie on the surface: in an 8-dimensional simplex, the resolution k must be 10 (or higher) to ensure that at least one point comes to lie in the interior of the simplex. At this stage (8-dimensional simplex with 10 points along each edge) the simplex comprises already 24310 points. In this example, the points have been packed in the generalized hexagonal lattice A_d, which has the right symmetry for a simplex [12].

Two conclusions from this section are that the volume of high-dimensional spaces is vast and most of the volume is concentrated on the surface. Unless special measures are taken, most of the design points will equally lie on (or near) the hull of the experimental region, (almost) equaling one or more of the constraints that define the experimental region.

Another implication is that if one can perform only few experiments compared to the volume of the experimental region, the shape of the experimental region will make a heavy impact on the optimal design. To explain it in terms of Fig. 13.3, if you put just as many points in a higher-dimensional experimental region, most points would make contact with one of the surfaces and the structure of the optimal design would have to deviate strongly from the nice lattice which it would really "like" to be.

As a consequence, if you have a large number of points (say, more than 10^d in dimension d), an extract from a lattice will be a good approximation to an optimal design. If you have less points, you may want to think about optimizing your design numerically.

If you can make a reasonable guess at the covariance structure of your response surface, use eq. 13.3 as a criterion in conjunction with your favorite optimizer.

If not, you should make a guess as to whether your response surface is more likely very rough or very smooth on the scale of your sampling density. Judging by the toy examples from this section, in case of doubt you are probably in the rough regime and should then use some sphere packing algorithm, e.g. [18]. If you have reason to believe that your response surface is very smooth, you would like to use the dual of a sphere packing. Since that is not defined for finite experimental regions, and not for nonlattice point sets either, you might want to put some trust in the conjecture mentioned in section 13.5.2 and decide to design a point set that is good for vector quantization. This problem has been studied extensively in connection with the encoding of analog signals for transmission over a digital network. (For an excellent overview, see [19].) Design algorithms for this purpose are readily applicable also to experimental design, for sufficiently smooth surfaces, with or without a weight function. Available algorithms fall into one of two main classes: "block-iterative" [20] or "sample-iterative" [21].

13.7 Extensions

The following subsections will mention further applications based on the kriging formalism (sections 13.7.1 and 13.7.2), compare the kriging approach to popular algorithms for the generation of space-filling designs (13.7.3), and discuss the repercussions of some of the assumptions made and how they can be dealt with (13.7.4).

13.7.1 Application to subset selection

In the foregoing, we have assumed that each point in the experimental region is eligible as a design point. This need not be true in all applications.

For instance, the finest resolution in the amount of deposited component that a pipetting robot can achieve may be an entire droplet. This leads to a discretization of the experimental region.

In other applications, for instance in pharmacological screening, a small number of compounds should be selected from a larger library. As in the continuous case, this selection should be optimal in the sense of allowing us to learn a maximum about the response in the experimental region. This is the **subset selection** problem. We assume that the compounds have been embedded in a property space (see chapter 1, table 1) such that the response varies as smoothly as possible from one compound to the next. Each compound is represented by one point in property space and as in the continuous case, points that are close in the experimental space should evoke a similar response.

In both examples mentioned, the experimental region now consists of a cloud of points in space. In cases such as the screening example, the density of the cloud will not be uniform: there may be a clutter of points in some areas of the experimental region while others may be nearly empty.

Now, since the continuous experimental region has been replaced by a cloud of points, and assuming the integrated mean square error criterion has been chosen,

the integral in equation 13.3 can be replaced by a summation of the mean square error over all points. However, the non-uniformity of the point density needs to be taken into account, otherwise the resultant design will have most of its points in areas that are highly populated by the library [1]. Such a design is called a **representative** subset and is not compatible with the assumption of uniformly distributed probability of high response, which in turn asks for **diverse** subsets.

If, instead, the maximum mean square error has been chosen as design criterion, the optimization process is simplified in a similar way. It then suffices to calculate, for a given trial design, the mean square error at each point in the cloud, and retain the largest value found. The trial design can then be perturbed and the new maximum mean square error found, etc.

In summary, modeling of the response surface as a stochastic procedure and design optimization based on the best linear unbiased estimator can be applied to the subset selection problem, with the computational simplification of considering only the given library members as candidate points for the design rather than the entire experimental region.

13.7.2 Adaptive sampling

All previous arguments were based on the assumption that first, a design is sought and afterwards, all experiments determining the responses are performed. Such experiments are denoted **simultaneous**. What if you have been lucky enough to locate an area of promising response, and if you now wish to switch from exploration to optimization? You could define a narrower experimental region and proceed as before, first optimizing a design and then performing measurements. If you do not wish to waste experiments, you can construct a design *around* the sampling points which you already have in that narrower experimental region (for details, see [1]). If you do have a fair idea of the smoothness of the response surface, however, you can do even better than that. You can construct a design that directly uses the information from the previous measurements. The basic idea in such an **adaptive** (also called sequential) experiment is to put your next design point where the response (as predicted by the kriging interpolator) *plus* the estimated uncertainty of your prediction are maximal. For a highly readable account of this approach, see [22].

13.7.3 Relation to other algorithms

The algorithm you employ should mirror your assumptions concerning

(i) the response surface and

(ii) your experimental strategy.

Concerning (i), modeling the response surface as realization of a stationary stochastic process allows to lay open the implicit assumptions that many published algorithms rely on.

For instance, the "maximin" algorithm (which seeks to maximize the minimum distance between any two design points) implicitly tries to find a good sphere packing, and as such relies on the assumption of a rough response surface in combination with the aim of minimizing the average uncertainty. Be aware that the popular maximin algorithm pushes design points into the hull of the experimental region. This can be avoided by using a definition of sphere packing which relies on the sphere packing density in the experimental region rather than on the interpoint distance [1].

If, instead, an algorithm tries to minimize the maximum distance between any point in the experimental region and its closest design point ("minimax"), it seeks to solve the covering problem, which in the current framework can be interpreted as trying to minimize the maximum uncertainty on a rough response surface, as in section 13.4.2. Some more correspondences are discussed in [1].

Concerning (*ii*), an algorithm that proposes design points in a sequential fashion is not compatible with a simultaneous experimental setup. If all experiments are performed without taking into account previous measurements, then an optimal design should also be found by varying all design points simultaneously, rather than by augmenting an initial small design, one by one.

13.7.4 Technical details

The expected uncertainty in a kriging prediction depends on knowledge of the true covariance function. If a guessed or fitted covariance function is used instead, the uncertainty is underestimated. An uncorrelated random error component can be admitted to account for random measurement error. The scaling of the coordinate axes is vital; an isotropic covariance function assumes that all spatial directions are of equal importance. If one dimension of the experimental region represents the quantity of a particular additive and its units are changed from, say, millimol to mol while keeping all other things equal, the experimental region is compressed thousandfold in that direction and much fewer points will be available to illuminate the effect of the additive on the response. In other words, prior knowledge must be used to scale the axes of the experimental region or, equivalently, parameterize a non-isotropic covariance function.[14] Also, less stringent assumptions concerning the nature of the response surface can be made [5, 1], in particular the response can be modeled as the sum of a trend (given, e.g., by a polynomial) and a stochastic process. Depending on the relative importance of the two, the resultant design will vary between one that is optimal according to the "alphabetic" criteria (section 13.2.1) and one that is optimal for the pure stochastic model.

[14] Rescaling of the experimental space amounts to replacing a covariance matrix that is constant along its diagonal and zero elsewhere, with a diagonal covariance matrix with differing diagonal elements. This is equivalent to making a shift from Euclidean to weighted Euclidean distance.

13.8 Conclusions and practical recommendations

Before anything else, you are required to define your experimental region (defined in the introduction). If your experimental space has no reasonable natural limits, you need to fix constraints arbitrarily. Upper and lower bounds on individual dimensions are the most convenient constraints to formulate and implement, but you should generally try to define your experimental region as tight as possible: when working in multiple dimensions you are almost always data starved and if you can afford only so many design points, better not waste them in an overly generous margin.

In exploring an unknown space of materials, we think that modeling the response surface as a realization of a stochastic process offers much flexibility in the incorporation of prior knowledge through weight and covariance functions. If nothing is known about the covariance structure of the response surface, at the very least a guess is required as to whether it is smooth or rough on the scale of the interpoint distance. In the light of section 13.5.4, we will only discuss the integrated mean square error criterion in the following.

Remembering that the volume of a multidimensional space is vast, the few thousand design points you can afford may easily look forlorn. Their isolation will usually lead to a situation in which your response surface can vary significantly between design points. In this case, if you have relatively few[15] design points, you should, for the reasons given in sections 13.4.1 and 13.6, consider numerical optimization of your design with a sphere packing algorithm, for instance as in [18]. If you have many design points available relative to the dimensionality of your problem, you may ignore boundary effects and excise part of a packing lattice for your design.

If you can assume that your response surface is smooth, you are in a lucky position: you can then hope to gain a fairly accurate picture of the true response surface from your experiments. Also, if you have too few design points to simply use the dual of a packing lattice (see sections 13.4.1 and 13.5.2), you can rely on efficient and robust algorithms for vector quantization (section 13.6).

If you are in a position to make an educated guess at the covariance structure of the response surface, you can optimize your design under eq. 13.3 using any local or global optimizer. Also, if for some reason you want more design points in a particular area, you can introduce this bias through a weight function [1].

In essence, you can get away with very little knowledge about your response surface, but if you do know more you have ways of putting this knowledge to work in a consistent framework. The above is condensed in the flowchart in Fig. 13.14.

In summary, if you can afford many experiments and wish to invest little into optimization of the experimental design, you can sample your experimental region on a lattice—though preferably not on the cubic. Switching to a more suitable lattice will improve the performance significantly for the same number

[15]It is difficult to give numbers that are useful for a wide range of situations, but if hard pressed for one, we might say that "few" means $\ll 10^d$ in dimension d.

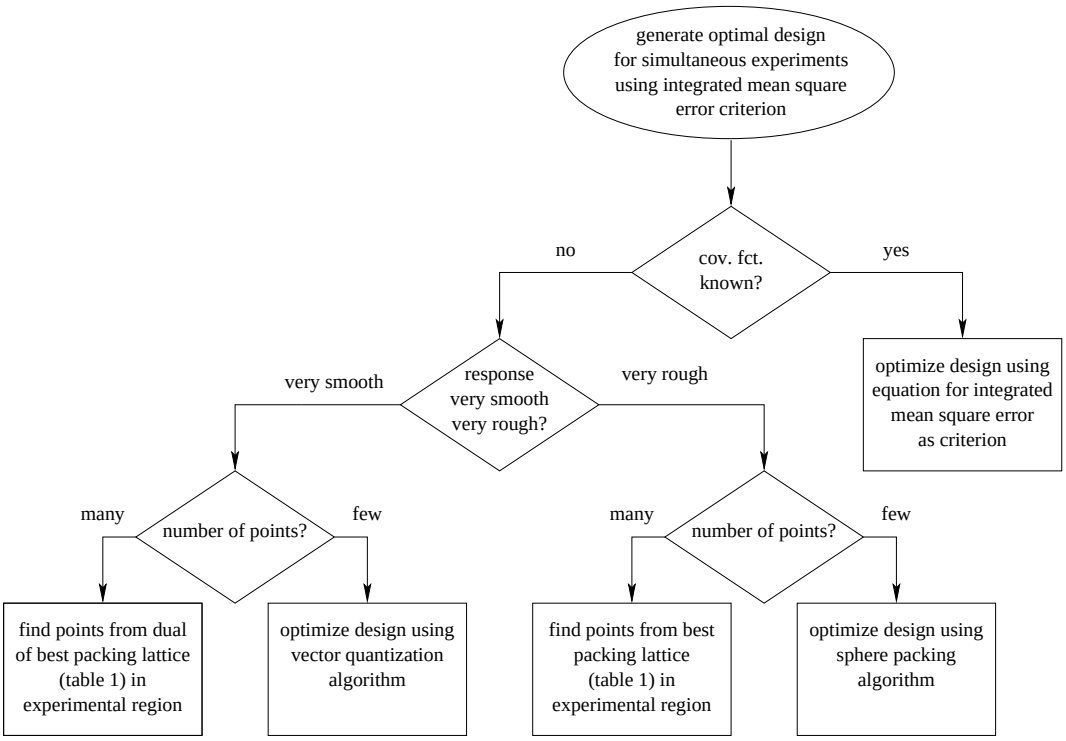


Figure 13.14: How to determine a suitable design method.

of experiments, still assuming that the number of experiments is large. If you need to make the most out of a limited number of experiments, you will be forced to invest time and expertise into numerical optimization of the design.

Appendix A Generator matrices for lattices

A.1 Generating lattice points

As discussed in section 13.5.1, each point in an infinite lattice can be specified by a linear combination of basis vectors, where all coefficients take integer values only. The basis vectors can be collected, row by row, into a generator matrix. In this Appendix, we give one generator matrix for each lattice mentioned in the chapter (the Z, A, D, and E families, and Λ_{16}).

To span a d-dimensional lattice, d basis vectors are required. However, each of these vectors can have more than d components. The two-dimensional hexagonal lattice A_2, for instance, can be defined as the intersection of the three-dimensional cubic lattice (x, y, z) with x, y, z integer, and of the plane $x + y + z = 0$. The resultant generator matrix can be written in the form

$$\begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

which contains, unlike the matrices given in sections 13.5.1 and A.2, no square root. In general, generator matrices are often easier to write down using basis

vectors with more than d components, and textbooks usually give these matrices. However, these lattices first need to be rotated, thus discarding the redundant dimensions, before they can be used in experimental design. All generator matrices tabulated in the following have exactly d columns for d-dimensional lattices. Most of them are taken from the appendix in [15], where further information can be found.

For every lattice, there is a dual (denoted by a star). One way to define it is through Fourier analysis. Consider a sum of Dirac impulses, one at each point of a particular lattice. The Fourier transform of the sum is another sum of Dirac impulses. The locations of these impulses is the dual of the original lattice. If B is a square generator matrix for a lattice, then $(B^{-1})^T$ is a generator matrix for its dual.

For any particular lattice, basis vectors can be chosen in a multitude of ways. Furthermore, lattices that can be obtained from each other just by rotation, scaling, and possibly reflection are usually considered *equivalent* (denoted by "$\cong$"), because even if such lattices do not contain the same points, they have analogous properties as far as Euclidean distance is concerned. Useful equivalence relations include $Z_1 \cong Z_1^* \cong A_1 \cong A_1^*$, $A_2 \cong A_2^*$ (hexagonal), $A_3 \cong D_3$ (fcc), $A_3^* \cong D_3^*$ (bcc), $D_4 \cong D_4^*$, $E_8 \cong E_8^*$, and $\Lambda_{16} \cong \Lambda_{16}^*$.

Once an appropriate lattice has been chosen, we must also be able to determine the points of the lattice that fall within a given experimental region $\mathcal{R}$. The general procedure can be outlined as follows.

(i) Project the region $\mathcal{R}$ onto the first basis vector. This defines an interval in the direction of the basis vector; find the endpoints of this interval. Repeat for all d basis vectors and call the intervals $J_1, \ldots, J_d$.

(ii) For $i = 1, \ldots, d$, generate all integer multiples of basis vector i that lie within J_i and call this set of vectors S_i.

(iii) Compute the set of all points that can be obtained as the sum of i vectors, one from each S_i. These points belong to a parallelepiped with sides parallel to the basis vectors. The parallelepiped encloses $\mathcal{R}$.

(iv) Discard all points in the set that lie outside $\mathcal{R}$.

A.2 Generator matrices

$$Z_d, \quad d \geq 1: \qquad \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$A_d, \quad d \geq 1: \qquad \begin{bmatrix} \alpha & 1 & \cdots & 1 \\ 1 & \alpha & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \alpha \end{bmatrix} \quad \text{with } \alpha = \sqrt{d+1} + 2$$

$A_d^*, \quad d \geq 1:$ as above with $\alpha = \sqrt{d+1} - d$

$D_d, \quad d \geq 3:$
$$\begin{bmatrix} 2 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$D_d^*, \quad d \geq 3:$
$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 1/2 & 1/2 & \cdots & 1/2 & 1/2 \end{bmatrix}$$

$E_6:$
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \alpha \\ 0 & 1 & 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & 0 & 0 & \alpha \\ 0 & 0 & 0 & 1 & 0 & \alpha \\ 0 & 0 & 0 & 0 & 1 & \alpha \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 3\alpha/2 \end{bmatrix} \text{ with } \alpha = \sqrt{3}$$

$E_6^*:$ as above with $\alpha = 1/\sqrt{3}$

$E_7:$
$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

$E_7^*:$
$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

$E_8:$
$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix}$$

$$\Lambda_{16}: \quad \begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Bibliography

[1] Hamprecht, F. A.; Thiel, W.; van Gunsteren, W. F. Journal of Chemical Information and Computer Science **2002**. *42*, 414-428.

[2] Fedorov, V. V. *Theory of Optimal Experiments*. Academic, London, 1972.

[3] Atkinson, A. C.; Donev, A. N. *Optimum Experimental Designs*. Oxford University Press, Oxford, 1992.

[4] Pebesma, E. J.; Heuveling, G. B. M. Technometrics **1999**. *41*, 303–312. http://www.gstat.org.

[5] Chilès, J.-P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*. Wiley series in probability and statistics. Wiley, New York, 1999.

[6] Cressie, N. A. C. Mathematical Geology **1990**. *22*, 239–252.

[7] Sacks, J.; Welch, W. J.; Mitchell, T. J.; Wynn, H. P. Statistical Science **1989**. *4*, 409–435.

[8] Agrell, E.; Hamprecht, F. A.; Künsch, H. in preparation.

[9] Petersen, D. P.; Middleton, D. Information and Control **1962**. *5*, 279–323.

[10] Johnson, M. E.; Moore, L. M.; Ylvisaker, D. Journal of Statistical Planning and Inference **1990**. *26*, 131–148.

[11] Gersho, A. IEEE Transactions on Information Theory **1982**. *IT-28*(2), 157–166.

[12] Conway, J. H.; Sloane, N. J. A. *Sphere Packings, Lattices and Groups*, vol. 290 of *Grundlehren der Mathematischen Wissenschaften*. Springer, New York, 1993, 2nd ed.

[13] Okabe, A.; Boots, B.; Sugihara, K. *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Wiley, Chichester, England, U.K., 1992.

[14] Agrell, E. *Voronoi-Based Coding*. Ph.D. thesis, Chalmers Univ. of Technology, Göteborg, Sweden, **1997**.

[15] Agrell, E.; Eriksson, T. IEEE Transactions on Information Theory **1998**. *44*(5), 1814–1828.

[16] Forney, Jr., G. D. In Calderbank, R.; Forney, Jr., G. D.; Moayeri, N., eds., *Coding and Quantization*, American Mathematical Society, Providence, RI, 1993, vol. 14 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 1–14.

[17] Agrell, E.; Eriksson, T.; Vardy, A.; Zeger, K. IEEE Transactions on Information Theory **2002**, to appear.

[18] Hardin, R. H.; Sloane, N. J. A. *Operating manual for Gosset: a general-purpose program for constructing experimental designs*. AT&T Bell Labs, Murray Hill, NJ, USA, 2nd ed., **1994**. http://www.research.att.com/∼njas/gosset/.

[19] Gray, R. M.; Neuhoff, D. L. IEEE Transactions on Information Theory **1998**. *44*(6), 2325–2383.

[20] Linde, Y.; Buzo, A.; Gray, R. M. IEEE Transactions on Communications **1980**. *COM-28*, 84–95.

[21] Yair, E.; Zeger, K.; Gersho, A. IEEE Transactions on Information Theory **1992**. *40*(2), 294–309.

[22] Jones, D. R.; Schonlau, M.; Welch, W. J. Journal of Global Optimization **1998**. *13*, 455–492.