Design of computer experiments: space filling and beyond

Luc Pronzato · Werner G. Müller

Abstract When setting up a computer experiment, it has become a standard practice to select the inputs spread out uniformly across the available space. These so-called space-filling designs are now ubiquitous in corresponding publications and conferences. The statistical folklore is that such designs have superior properties when it comes to prediction and estimation of emulator functions. In this paper we want to review the circumstances under which this superiority holds, provide some new arguments and clarify the motives to go beyond space-filling. An overview over the state of the art of space-filling is introducing and complementing these results.

Keywords Kriging · Entropy · Design of experiments · Space-filling · Sphere packing · Maximin design · Minimax design

1 Introduction

Computer simulation experiments (see, e.g., Santner et al. 2003; Fang et al. 2005; Kleijnen 2009) have now become a popular substitute for real experiments when the latter are infeasible or too costly. In these experiments, a deterministic computer code, the simulator, replaces the real (stochastic) data generating process. This practice has generated a wealth of statistical questions, such as how well the simulator is able to mimic reality or which estimators are most suitable to adequately represent a system.

However, the foremost issue presents itself even before the experiment is started, namely how to determine the inputs for which the simulator is run? It has become standard practice to select these inputs such as to cover the available space as uniformly as possible, thus generating so called space-filling experimental designs. Naturally, in dimensions greater than one there are alternative ways to produce such designs. We will therefore in Sects. 2 and 3 briefly review the most common approaches to space-filling design, taking a purely model-free stance. We will then (Sect. 4) investigate how these designs can be motivated from a statistical modelers point of view and relate them to each other in a meaningful way. Eventually we will show that taking statistical modeling seriously will lead us to designs that go beyond space-filling (Sect. 5 and 6). Special attention is devoted to Gaussian process models and kriging. The only design objective considered corresponds to reproducing the behavior of a computer code over a given domain for its input variables. Some basic principles about algorithmic constructions are exposed in Sect. 7 and Sect. 8 briefly concludes.

The present paper can be understood as a survey focusing on the special role of space-filling designs and at the same time providing new illuminative aspects. It intends to bring the respective sections of Koehler and Owen (1996) up to date and to provide a more statistical point of view than Chen et al. (2006).

L. Pronzato (✉)
Laboratoire I3S, bâtiment Euclide, les Algorithmes, Université de Nice-Sophia Antipolis/CNRS, 2000 route des Lucioles, BP 121 06903, Sophia Antipolis cedex, France
e-mail: pronzato@i3s.unice.fr

W.G. Müller
Department of Applied Statistics, Johannes-Kepler-University, Linz Freistädter Straße 315, 4040 Linz, Austria
e-mail: werner.mueller@jku.at

2 State of the art on space-filling design

2.1 Geometric criteria

There is little ambiguity on what constitutes a space-filling design in one dimension. If we define an exact design $\xi = (x_1, \ldots, x_n)$ as a collection of n points and consider a section of the real line as the design space, say $\mathfrak{X} = [0, 1]$ after suitable renormalization, then, depending upon whether we are willing to exploit the edges or not, we have either $x_i = (i - 1)/(n - 1)$ or $x_i = (2i - 1)/(2n)$ respectively.

The distinction between those two basic cases comes from the fact that one may consider distances only amongst points in the design ξ or to all points in the set $\mathfrak{X}$. We can carry over this notion to the less straightforward higher dimensional case $d > 1$, with now $\xi = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Initially we need to define a proper norm $\|.\|$ on $\mathfrak{X} = [0, 1]^d$, Euclidean distances and normalization of the design space will not impede generality for our purposes. We shall denote

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$$

the distance between the two design points $\mathbf{x}_i$ and $\mathbf{x}_j$ of ξ. We shall not consider the case where there exist constraints that make only a subset of $[0, 1]^d$ admissible for design, see for instance Stinstra et al. (2003) for possible remedies; the construction of Latin hypercube designs (see Sect. 2.2) with constraints is considered in Petelet et al. (2010).

Let us first seek for a design that wants to achieve a high spread solely amongst its support points within the design region. One must then attempt to make the smallest distance between neighboring points in ξ as large as possible. That is ensured by the maximin-distance criterion (to be maximized)

$$\phi_{Mm}(\xi) = \min_{i \neq j} d_{ij}.$$

We call a design that maximizes $\phi_{Mm}(\cdot)$ a *maximin-distance design*, see Johnson et al. (1990). An example is given in Fig. 1 (left). This design can be motivated by setting up the tables in a restaurant such that one wants to minimize the chances to eavesdrop on another party's dinner talk.

In other terms, one wishes to maximize the radius of n non-intersecting balls with centers in $\mathfrak{X}$. When $\mathfrak{X}$ is a d-dimensional cube, this is equivalent to packing rigid spheres in $\mathfrak{X}$, see Melissen (1997, p. 78). The literature on sphere packing is rather abundant. In dimension $d = 2$, the best known results up to $n = 10{,}000$ for finding the maximum common radius of n circles which can be packed in a square are presented on http://www.packomania.com/ (the example on Fig. 1 (left) is taken from there, with $\phi_{Mm}(\xi) \simeq 0.5359$, indicating that the 7-point design in Johnson et al. (1990) is not a maximin-distance design); one may refer to Gensane (2004) for best-known results up to $n = 32$ for $d = 3$.
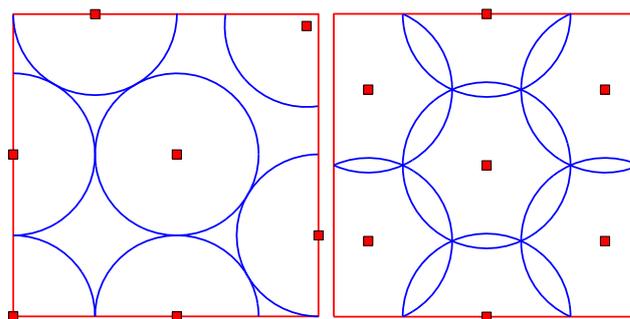
Fig. 1 Maximin (*left*, see http://www.packomania.com/ and minimax (*right*, see Johnson et al. 1990) distance designs for $n = 7$ points in $[0, 1]^2$. The *circles* have radius $\phi_{Mm}(\xi)/2$ on the *left panel* and radius $\phi_{mM}(\xi)$ on the *right one*

Among the set of maximin-distance designs (when there exist several), a maximin-optimal design ξ^*_{Mm} is such that the number of pairs of points $(\mathbf{x}_i, \mathbf{x}_j)$ at the distance $d_{ij} = \phi_{Mm}(\xi^*_{Mm})$ is minimum (several such designs can exist, and measures can be taken to remove draws, see Morris and Mitchell (1995), but this is not important for our purpose).

Consider now designs ξ that attempt to make the maximum distance from all the points in $\mathfrak{X}$ to their closest point in ξ as small as possible. This is achieved by minimizing the minimax-distance criterion

$$\phi_{mM}(\xi) = \max_{\mathbf{x} \in \mathfrak{X}} \min_{\mathbf{x}_i} \|\mathbf{x} - \mathbf{x}_i\|.$$

We call a design that minimizes $\phi_{mM}(\cdot)$ a *minimax-distance design*, see Johnson et al. (1990) and Fig. 1 (right) for an example. (Note the slight confusion in terminology as it is actually minimaximin.) These designs can be motivated by a table allocation problem in a restaurant, such that a waiter is as close as possible to a table wherever he is in the restaurant.

In other terms, one wishes to cover $\mathfrak{X}$ with n balls of minimum radius. Among the set of minimax-distance designs (in case several exist), a minimax-optimal design ξ^*_{mM} maximizes the minimum number of $\mathbf{x}_i$'s such that $\min_i \|\mathbf{x} - \mathbf{x}_i\| = \phi_{mM}(\xi^*_{mM})$ over all points $\mathbf{x}$ having this property.

2.2 Latin hypercubes

Note that pure space-filling designs such as ξ^*_{mM} and ξ^*_{Mm} may have very poor projectional properties; that is, they may be not space-filling on any of their meaningful subspaces, see Fig. 1. The opposite is desirable for computer experiments, particularly when some inputs are of no influence in the experiment, and this property was called non-collapsingness by some authors (cf. Stinstra et al. 2003). This requirement about projections is one of the reasons that researches have started to restrict the search for designs to the class of so-called Latin-hypercube (Lh) designs, see
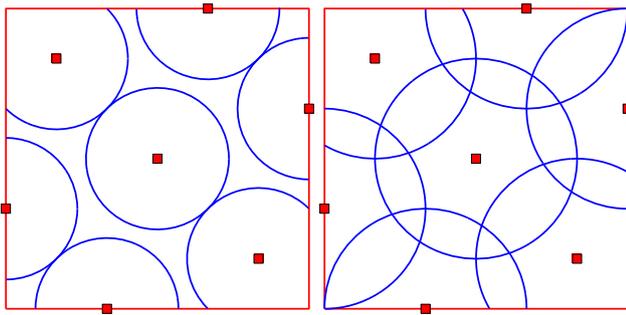
Fig. 2 Minimax-Lh and simultaneously maximin-Lh distance design for $n = 7$ points in $[0, 1]^2$, see http://www.spacefillingdesigns.nl/. The *circles* have radius $\phi_{Mm}(\xi)/2$ on the *left panel* and radius $\phi_{mM}(\xi)$ on the *right one*

McKay et al. (1979), which have the property that any of their one-dimensional projections yields the maximin distance sequence $x_i = (i - 1)/(n - 1)$. An additional advantage is that since the generation of Lh-designs as a finite class is computationally rather simple, it has become customary to apply a secondary, e.g. space-filling, criterion to them, sometimes by a mere brute-force enumeration as in van Dam (2007). An example of minimax and simultaneously maximin Lh design is presented in Fig. 2 (note that there is a slight inconsistency about minimax-Lh designs in that they are maximin rather than minimax on their one-dimensional projections).

Other distances than Euclidean could be considered; when working within the class of Lh designs the situation is easier with the L_1 or L_∞ norms than with the L_2 norm, at least for $d = 2$, see van Dam et al. (2007), van Dam (2007). The class of Lh designs is finite but large. It contains $(n!)^{d-1}$ different designs (not $(n!)^d$ since the order of the points is arbitrary and the first coordinates can be fixed to $\{x_i\}_1 = (i - 1)/(n - 1)$), and still $(n!)^{d-1}/(d - 1)!$ if we consider designs as equivalent when they differ by a permutation of coordinates. An exhaustive search is thus quickly prohibitive even for moderate values of n and d. Most algorithmic methods are of the exchange type, see Sect. 7. In order to remain in the class of Lh designs, one exchange-step corresponds to swapping the j-th coordinates of two points, which gives $(d - 1)n(n - 1)/2$ possibilities at each step (the first coordinates being fixed). Another approach that takes projectional properties into account but is not restricted to the class of Lh designs will be presented in Sect. 3.3. Note that originally McKay et al. (1979) have introduced Lh designs as random sampling procedures rather than candidates for providing fixed designs, those random designs being not guaranteed to have good space-filling properties. Tang (1993) has introduced orthogonal-array-based Latin hypercubes to improve projections on higher dimensional subspaces, the space-filling properties of which were improved by Leary et al. (2003). The usefulness of Lh designs in model-based

(as discussed in Sect. 4) examples was demonstrated in Pebesma and Heuvelink (1999). The algorithmic construction of Lh designs that optimize a discrepancy criterion (see Sect. 2.3) or an entropy based criterion (see Sect. 3.3) is considered respectively in Iooss et al. (2010) and Jourdan and Franco (2010); the algebraic construction of Lh designs that minimize the integrated kriging variance for the particular correlation structure $C(\mathbf{u}, \mathbf{v}; \nu) = \exp(-\nu\|\mathbf{u} - \mathbf{v}\|_1)$ (see Sect. 4.1) is considered in Pistone and Vicario (2010).

2.3 Other approaches to space-filling

There appear a number of alternative approaches to space-filling in the literature, most of which can be similarly distinguished by the stochastic nature of the inputs, i.e. whether ξ is to be considered random or fixed.

For the latter, natural simple designs are regular grids. Such designs are well suited for determining appropriate model responses and for checking whether assumptions about the errors are reasonably well satisfied. There seems little by which to choose between e.g. a square grid or a triangular grid; it is worth noting, however, that the former may be slightly more convenient from a practical standpoint (easier input determination) but that the latter seems marginally more efficient for purposes of model based prediction (cf. Yfantis et al. 1987).

Bellhouse and Herzberg (1984) have compared optimum designs and uniform grids (in a one-dimensional model based setup) and they come to the conclusion that (depending upon the model) predictions for certain output regions can actually be improved by regular grids. A comparison in a multi-dimensional setup including correlations can be found in Herzberg and Huda (1981).

For higher dimensional problems, Bates et al. (1996) recommend the use of (non-rectangular) lattices rather than grids (they also reveal connections to model-based approaches). In the two-dimensional setup (on the unit square $[-1, 1]^2$) the Fibonacci lattice (see Koehler and Owen 1996) proved to be useful. The advantage of lattices is that their projection on lower dimensions covers the design region more or less uniformly. Adaptations to irregular design regions may not be straightforward, but good enough approximations will suffice. This is not the case for many other systematic designs that are frequently proposed in the literature, such as central composite designs, the construction of which relies on the symmetry of the design region.

It is evident that randomization can be helpful for making designs more robust. On a finite grid $\mathfrak{X}$ with N candidate points we can think of randomization as drawing a single design ξ according to a pre-specified probability distribution $\pi(\cdot)$. The uniform distribution then corresponds to simple random sampling and more refined schemes (e.g., stratified random sampling, see Fedorov and Hackl 1997), can be devised by altering $\pi(\cdot)$. A comparison between deterministic

selection and random sampling is hard to make, since for a finite sample it is evident that for any single purpose it is possible to find a deterministic design that outperforms random sampling. Performance benchmarking for various space-filling designs can be found in Johnson et al. (2008) and Bursztyn and Steinberg (2006).

All of the methods presented above seem to ensure a reasonable degree of overall coverage of the study area. However, there have been claims (see e.g. Fang and Wang 1993), that the efficiency (with respect to coverage) of these methods may be poor when the number of design points is small. To allow for comparisons between designs in the above respect Fang (1980) (see also Fang et al. 2000) introduced some formal criteria, amongst them the so-called discrepancy

$$\mathcal{D}(\xi) = \max_{\mathbf{x} \in \mathfrak{X}} |F_n(\mathbf{x}) - U(\mathbf{x})|. \tag{1}$$

Here $U(\cdot)$ is the c.d.f. of the uniform distribution on $\mathfrak{X}$ and $F_n(\cdot)$ denotes the empirical c.d.f. for ξ. The discrepancy by this definition is just the Kolmogorov-Smirnov test statistic for the goodness-of-fit test for a uniform distribution. Based upon this definition, Fang and Wang (1993) suggest to find 'optimum' designs of given size n that minimize $\mathcal{D}(\xi)$, which they term the U-criterion. For $d = 1$ and $\mathfrak{X} = [0, 1]$, the minimax-optimal design ξ_{mM}^* with $x_i = (2i - 1)/(2n)$ is optimal for (1), with $\mathcal{D}(\xi_{mM}^*) = 1/(2n)$. Note, however, that $\mathcal{D}(\xi) \geq 0.06 \log(n)/n$ for any *sequence* of n points, see Niederreiter (1992, p. 24). It turns out that for certain choices of n lattice designs are U-optimum. Those lattice designs are also D-optimum for some specific Fourier regressions and this and other connections are explored by Riccomagno et al. (1997). An example in Santner et al. (2003, Chap. 5) shows that the measure of uniformity expressed by $\mathcal{D}(\xi)$ is not always in agreement with common intuition.

Niederreiter (1992) has used similar concepts for the generation of so called low discrepancy sequences. Originately devised for the use in Quasi Monte Carlo sampling, due to the Koksma-Hlawka inequality in numerical integration, their elaborate versions, like Faure, Halton and Sobol sequences, are increasingly used in computer experiments (see, e.g., Fang and Li 2006). Santner et al. (2003, Chap. 5) and Fang et al. (2005, Chap. 3) provide a good overview of the various types of the above discussed designs and their relations. Other norms than $\|\cdot\|_\infty$ can be used in the definition of discrepancy, yielding $\mathcal{D}_p(\xi) = (\int_{\mathfrak{X}} |F_n(\mathbf{x}) - U(\mathbf{x})|^p \, d\mathbf{x})^{1/p}$, and other types of discrepancy (centered, wrap-around) may also be considered. Low discrepancy sequences present the advantage that they can be constructed sequentially (which is not the case for Lh designs), although one should take care of the irregularity of distributions, see Niederreiter (1992, Chap. 3) and Fang et al. (2000) (a conjecture in number theory states that $\mathcal{D}(\xi_n) \geq$

$c_d [\log(n)]^{d-1}/n$ for any sequence ξ_n with c_d a constant depending on d). It seems, however, that designs obtained by optimizing a geometric space-filling criterion are preferable for moderate values of n and that, for n large and $d > 1$, the space-filling properties of designs corresponding to low-discrepancy sequences may not be satisfactory (the points presenting sometimes alignments along subspaces). Note that Bischoff and Miller (2006) and related work reveal (in the one-dimensional setup) relationships between uniform designs and designs that reserve a portion of the observations for detecting lack-of-fit for various classical design criteria.

2.4 Some properties of maximin and minimax optimal designs

Notice that for any design ξ, $\mathfrak{X} \subset \cup_{i=1}^n \mathfrak{B}(\mathbf{x}_i, \phi_{mM}(\xi))$, with $\mathfrak{B}(\mathbf{x}, R)$ the ball with center $\mathbf{x}$ and radius R. Therefore, $\phi_{mM}(\xi) > [\text{vol}(\mathfrak{X})/(n V_d)]^{1/d} = (n V_d)^{-1/d}$, with $V_d = \pi^{d/2}/\Gamma(d/2 + 1)$ the volume of the d-dimensional unit ball. One may also notice that for any ξ, $n \geq 2$, $\phi_{mM}(\xi) > \phi_{Mm}(\xi)/2$ since $\mathfrak{X}$ cannot be covered with non-overlapping balls. A sort of reverse inequality holds for maximin-optimal designs. Indeed, take a maximin-optimal design ξ_{Mm}^* and suppose that $\phi_{mM}(\xi_{Mm}^*) > \phi_{Mm}(\xi_{Mm}^*)$. It means that there exists a $\mathbf{x}^* \in \mathfrak{X}$ such that $\min_i \|\mathbf{x}^* - \mathbf{x}_i\| > \phi_{Mm}(\xi_{Mm}^*)$. By substituting $\mathbf{x}^*$ for a $\mathbf{x}_i$ in ξ_{Mm}^* such that $d_{ij} = \phi_{Mm}(\xi_{Mm}^*)$ for some j, one can then either increase the value of $\phi_{Mm}(\cdot)$, or decrease the number of pairs of design points at distance $\phi_{Mm}(\xi_{Mm}^*)$, which contradicts the optimality of ξ_{Mm}^*. Therefore, $\phi_{mM}(\xi_{Mm}^*) \leq \phi_{Mm}(\xi_{Mm}^*)$.

Both $\phi_{Mm}(\xi_{Mm}^*)$ and $\phi_{mM}(\xi_{mM}^*)$ are non-increasing functions of n when $\mathfrak{X} = [0, 1]^d$ (there may be equality for different values of n, for instance, $\phi_{Mm}(\xi_{Mm}^*) = \sqrt{2}$ for $n = 3, 4$ and $d = 3$, see Gensane 2004). This is no longer true, however, when working in the class of Lh designs (see e.g. van Dam 2007 who shows that $\phi_{mM}(\xi_{mM}^*)$ is larger for $n = 11$ than for $n = 12$ when $d = 2$ for Lh designs).

The value of $\phi_{Mm}(\cdot)$ is easily computed for any design ξ, even when n and the dimension d get large, since we only need to calculate distances between $n(n-1)/2$ points in $\mathbb{R}^d$.

The evaluation of the criterion $\phi_{mM}(\cdot)$ is more difficult, which explains why a discretization of $\mathfrak{X}$ is often used in the literature. It amounts at approximating $\phi_{mM}(\xi)$ by $\tilde{\phi}_{mM,N}(\xi) = \max_{\mathbf{x} \in \mathfrak{X}_N} \min_i \|\mathbf{x} - \mathbf{x}_i\|$, with $\mathfrak{X}_N$ a finite grid of N points in $\mathfrak{X}$. Even so, the calculation of $\tilde{\phi}_{mM,N}(\xi)$ quickly becomes cumbersome when N increases (and N should increase fast with d to have a fine enough grid). It happens, however, that basic tools from computational geometry permit to reduce the calculation of $\max_{\mathbf{x} \in \mathfrak{X}} \min_i \|\mathbf{x} - \mathbf{x}_i\|$ to the evaluation of $\min_i \|\mathbf{z}_j - \mathbf{x}_i\|$ for a finite collection of points $\mathbf{z}_j \in \mathfrak{X}$, provided that $\mathfrak{X}$ is the d-dimensional cube $[0, 1]^d$. This does not seem to be much used and we detail the idea hereafter.

Consider the Delaunay tessellation of the points of ξ, see, e.g., Okabe et al. (1992), Boissonnat and Yvinec (1998). Each simplex has its $d+1$ vertices at design points in the tessellation and has the property that its circumscribed sphere does not contain any design point in its interior. We shall call those circumscribed spheres Delaunay spheres. When a solution $\mathbf{x}_*$ of the problem $\max_{\mathbf{x} \in \mathfrak{X}} \min_i \|\mathbf{x} - \mathbf{x}_i\|$ is in the interior of $[0, 1]^d$, it must be the center of some Delaunay sphere.

There a slight difficulty when $\mathbf{x}_*$ is on the boundary of $\mathfrak{X}$, since the tessellation directly constructed from the $\mathbf{x}_i$ does not suffice. However, $\mathbf{x}_*$ is still the center of a Delaunay sphere if we construct the tessellation not only from the points in ξ but also from their symmetric with respect to all $(d-1)$-dimensional faces of $\mathfrak{X}$, see Appendix A.

The Delaunay tessellation is thus constructed on a set of $(2d+1)n$ points. (One may notice that $\mathfrak{X}$ is not necessarily included in the convex hull of these points for $d \geq 3$, but this is not an issue.) Once the tessellation is calculated, we collect the radii of Delaunay spheres having their center in $\mathfrak{X}$ (boundary included); the value of $\phi_{mM}(\xi)$ is given by the maximum of these radii (see Appendix A for the computation of the radius of the circumscribed sphere to a simplex).

Efficient algorithms exist for the computation of Delaunay tessellations, see Okabe et al. (1992), Boissonnat and Yvinec (1998), Cignoni et al. (1998) and the references therein, which make the computation of $\phi_{mM}(\xi)$ affordable for reasonable values of d and n (the number of simplices in the Delaunay tessellation of M points in dimension d is bounded by $O(M^{\lceil d/2 \rceil})$). Clearly, not all $2dn$ symmetric points are useful in the construction, leaving open the possibility to reduce the complexity of calculations by using less than $(2d+1)n$ points.

Figure 3 presents the construction obtained for a 5-point Latin-hypercube design in dimension 2: 33 triangles are constructed, 11 centers of circumscribed circles belong to $\mathfrak{X}$, with some redundancy so that only 8 distinct points are candidate for being solution of the maximization problem $\max_{\mathbf{x} \in \mathfrak{X}} \min_i \|\mathbf{x} - \mathbf{x}_i\|$. The solution is at the origin and gives $\phi_{mM}(\xi) = \min_i \|\mathbf{x}_i\| \simeq 0.5590$.

3 Model-free design

We continue for the moment to consider the situation when we are not able, or do no want, to make an assumption about a suitable model for the emulator. We investigate the properties of some geometric and other model-free design criteria more closely and make connections between them.

3.1 L_q-regularization of the maximin-distance criterion

Following the approach in Appendix B, one can define regularized forms of the maximin-distance criterion, valid when
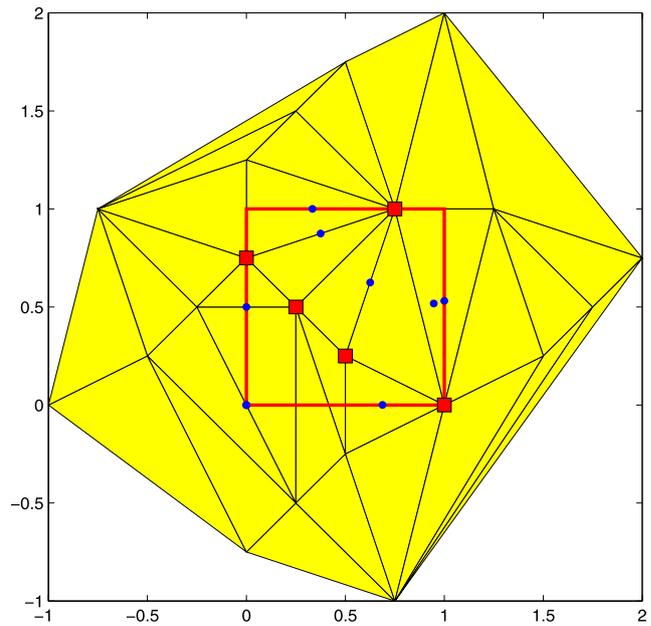
Fig. 3 Delaunay triangulation for a 5-point Lh design (squares), the 8 candidate points for being solution of $\max_{\mathbf{x} \in \mathfrak{X}} \min_i \|\mathbf{x} - \mathbf{x}_i\|$ are indicated by *dots*

$q > 0$ for any ξ such that $\phi_{Mm}(\xi) > 0$:

$$\underline{\phi}_{[q]}(\xi) = \left[\sum_{i<j} d_{ij}^{-q} \right]^{-1/q},$$

$$\overline{\phi}_{[q]}(\xi) = \left[\sum_{i<j} \mu_{ij} d_{ij}^{-q} \right]^{-1/q},$$

with $\mu_{ij} > 0$ for all i and $\sum_{i<j} \mu_{ij} = 1$, see (33), (34). The criterion $\underline{\phi}_{[q]}(\cdot)$ satisfies $\underline{\phi}_{[q]}(\xi) \leq \phi_{Mm}(\xi) \leq \overline{\phi}_{[q]}(\xi) \leq \underline{\mu}^{-1/q} \underline{\phi}_{[q]}(\xi)$, $q > 0$, with $\underline{\mu} = \min_{i<j} \mu_{ij}$, and the convergence to $\phi_{Mm}(\xi)$ is monotonic in q from both sides as $q \to \infty$. Taking μ as the uniform measure, i.e., $\mu_{ij} = \underline{\mu} = \binom{n}{2}^{-1}$ for all $i < j$, gives $\overline{\phi}_{[q]}(\cdot) = \underline{\mu}^{-1/q} \underline{\phi}_{[q]}(\cdot)$ and

$$\underline{\phi}_{[q]}(\xi) \leq \phi_{Mm}(\xi) \leq \binom{n}{2}^{1/q} \underline{\phi}_{[q]}(\xi). \tag{2}$$

It also yields the best lower bound on the maximin efficiency of an optimal design $\underline{\xi}^*_{[q]}$ for $\underline{\phi}_{[q]}(\cdot)$,

$$\frac{\phi_{Mm}(\underline{\xi}^*_{[q]})}{\phi_{Mm}(\xi^*_{Mm})} \geq \binom{n}{2}^{-1/q}, \tag{3}$$

where ξ^*_{Mm} denotes any maximin-distance design, see Appendix B. One may define $\overline{\phi}_{[0]}(\xi)$ as

$$\overline{\phi}_{[0]}(\xi) = \exp\left\{ \binom{n}{2}^{-1} \left[\sum_{i<j} \log(d_{ij}) \right] \right\} \tag{4}$$

and $\underline{\phi}_{[2]}(\cdot)$ corresponds to a criterion initially proposed by Audze and Eglais (1977). Morris and Mitchell (1995) use $\underline{\phi}_{[q]}(\cdot)$ with different values of q and make the observation that for moderate values of q (say, $q \precsim 5$) the criterion is easier to optimize than $\phi_{Mm}(\cdot)$ in the class of Lh designs. They also note that, depending on the problem, one needs to take q in the range 20–50 to make the two criteria $\underline{\phi}_{[q]}(\cdot)$ and $\phi_{Mm}(\cdot)$ agree about the designs considered best. Their observation is consistent with the efficiency bounds given above. According to the inequality (3), to ensure that the maximin efficiency of an optimal design for $\underline{\phi}_{[q]}(\cdot)$ is larger than $1 - \epsilon$ one should take approximately $q > 2 \log(n)/\epsilon$ (independently of the dimension d). Note that the use of $\underline{\phi}_{[q]}(\xi) = [\sum_{i \neq j} d_{ij}^{-q}]^{-1/q}$ would worsen the maximin efficiency bounds by a factor $2^{-1/q} < 1$ (but leaves $\overline{\phi}_{[q]}(\cdot)$ unchanged when the uniform measure $\mu_{ij} = [n(n-1)]^{-1}$ is used).

We may alternatively write $\phi_{Mm}(\xi)$ as

$$\phi_{Mm}(\xi) = \min_i d_i^*, \qquad (5)$$

where $d_i^* = \min_{j \neq i} d_{ij}$ denotes the nearest-neighbor (NN) distance from $\mathbf{x}_i$ to another design point in ξ. Following the same technique as above, a L_q-regularization applied to the min function in (5) then gives

$$\underline{\phi}_{[NN,q]}(\xi) \leq \phi_{Mm}(\xi) \leq n^{1/q} \underline{\phi}_{[NN,q]}(\xi) = \overline{\phi}_{[NN,q]}(\xi) \quad (6)$$

with

$$\underline{\phi}_{[NN,q]}(\xi) = \left[\sum_{i=1}^n (d_i^*)^{-q} \right]^{-1/q}. \qquad (7)$$

The reason for not constructing $\underline{\phi}_{[NN,q]}(\xi)$ from the decomposition $\phi_{Mm}(\xi) = \min_i \min_{j > i} d_{ij}$ is that the resulting criterion $[\sum_{i=1}^n (\min_{j>i} d_{ij})^{-q}]^{-1/q}$ depends on the ordering of the design points. One may also define $\overline{\phi}_{[NN,0]}(\xi)$ as

$$\overline{\phi}_{[NN,0]}(\xi) = \exp\left\{ \frac{1}{n} \left[\sum_{i=1}^n \log(d_i^*) \right] \right\}, \qquad (8)$$

see Appendix B. One can readily check that using the generalization (36) with $\phi(t) = \log(t)$ and $q = -1$ also gives $\overline{\phi}_{[NN,-1,\log]}(\xi) = \overline{\phi}_{[NN,0]}(\xi)$. Not surprisingly, $\underline{\phi}_{[NN,q]}(\cdot)$ gives a better approximation of $\phi_{Mm}(\cdot)$ than $\underline{\phi}_{[q]}(\xi)$: an optimal design $\underline{\xi}_{[NN,q]}^*$ for $\underline{\phi}_{[NN,q]}(\cdot)$ satisfies

$$\frac{\phi_{Mm}(\underline{\xi}_{[NN,q]}^*)}{\phi_{Mm}(\xi_{Mm}^*)} \geq n^{-1/q}.$$

which is larger than $1 - \epsilon$ when $q > \log(n)/\epsilon$, compare with (3). Exploiting the property that, for a given i,

$$\left(\sum_{j \neq i} d_{ij}^{-q} \right)^{-1/q} \leq d_i^* \leq (n-1)^{1/q} \left(\sum_{j \neq i} d_{ij}^{-q} \right)^{-1/q},$$

see (35), we obtain that

$$2^{-1/q} \underline{\phi}_{[q]}(\xi) \leq \underline{\phi}_{[NN,q]}(\xi) \leq \phi_{Mm}(\xi)$$

$$\phi_{Mm}(\xi) \leq n^{1/q} \underline{\phi}_{[NN,q]}(\xi) \leq \binom{n}{2}^{1/q} \underline{\phi}_{[q]}(\xi).$$

Note that the upper bounds on $\phi_{Mm}(\cdot)$ are sharp (think of a design with $n = d + 1$ points, all at equal distance from each other, i.e., such that $d_{ij} = d_i^*$ is constant).

Figure 4 presents the bounds (2) (dashed lines, top) and (6) (dashed lines, bottom) on the value $\phi_{Mm}(\xi)$ (solid line) for the 7-point maximin-distance design of Fig. 1 (left). Notice the accuracy of the upper bound $n^{1/q} \underline{\phi}_{[NN,q]}(\xi)$ (note the different scales between the top and bottom panels); the situation is similar for other maximin-distance designs since $d_i^* = \phi_{Mm}(\xi_{Mm}^*)$ for many i.

Oler (1961) indicates that for $d = 2$ $\phi_{Mm}(\xi_{Mm}^*) \leq [1 + \sqrt{1 + 2(n-1)/\sqrt{3}}]/(n-1)$. The equivalence with sphere-packing gives $\phi_{Mm}(\xi_{Mm}^*) < [(n V_d)^{1/d}/2 - 1]^{-1}$ with V_d the volume of the d-dimensional unit ball; this bound becomes quite loose for large d and can be improved by using results on packing densities of densest known packings (which may be irregular for some $d > 3$), yielding $\phi_{Mm}(\xi_{Mm}^*) \leq (3^{1/4}\sqrt{n/2} - 1)^{-1}$ for $d = 2$ and $\phi_{Mm}(\xi_{Mm}^*) \leq [(n/\sqrt{2})^{1/3} - 1]^{-1}$ for $d = 3$. Bounds for maximin Lh designs in dimension d can be found in van Dam et al. (2009).

3.2 L_q-regularization of the minimax-distance criterion

The same type of relaxation can be applied to the criterion $\phi_{mM}(\xi)$. First, $\phi(\mathbf{x}) = \min_{\mathbf{x}_i} \|\mathbf{x} - \mathbf{x}_i\|$ is approximated by $\phi_q(\mathbf{x}) = (\sum_i \|\mathbf{x} - \mathbf{x}_i\|^{-q})^{-1/q}$ with $q > 0$. Second, when $\mathfrak{X}$ is discretized into a finite grid $\mathfrak{X}_N = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$, $\max_{\mathbf{x} \in \mathfrak{X}_N} \phi_q(x) = [\min_{\mathbf{x} \in \mathfrak{X}_N} \phi_q^{-1}(\mathbf{x})]^{-1}$ can be approximated by $[\sum_{j=1}^N \phi_q^p(\mathbf{x}^{(j)})]^{1/p}$ with $p > 0$. This gives the following substitute for $\phi_{mM}(\xi)$,

$$\phi_{[p,q]}(\xi) = \left\{ \sum_{j=1}^N \left[\sum_{i=1}^n \|\mathbf{x}^{(j)} - \mathbf{x}_i\|^{-q} \right]^{-p/q} \right\}^{1/p}$$

with $p, q > 0$, see Royle and Nychka (1998). Note that the $\mathbf{x}_i$ are usually elements of $\mathfrak{X}_N$. When $\mathfrak{X}$ is not discretized, the sum over $\mathbf{x}^{(j)} \in \mathfrak{X}_N$ should be replaced by an integral over $\mathfrak{X}$, which makes the evaluation of $\phi_{[p,q]}(\xi)$ rather cumbersome.
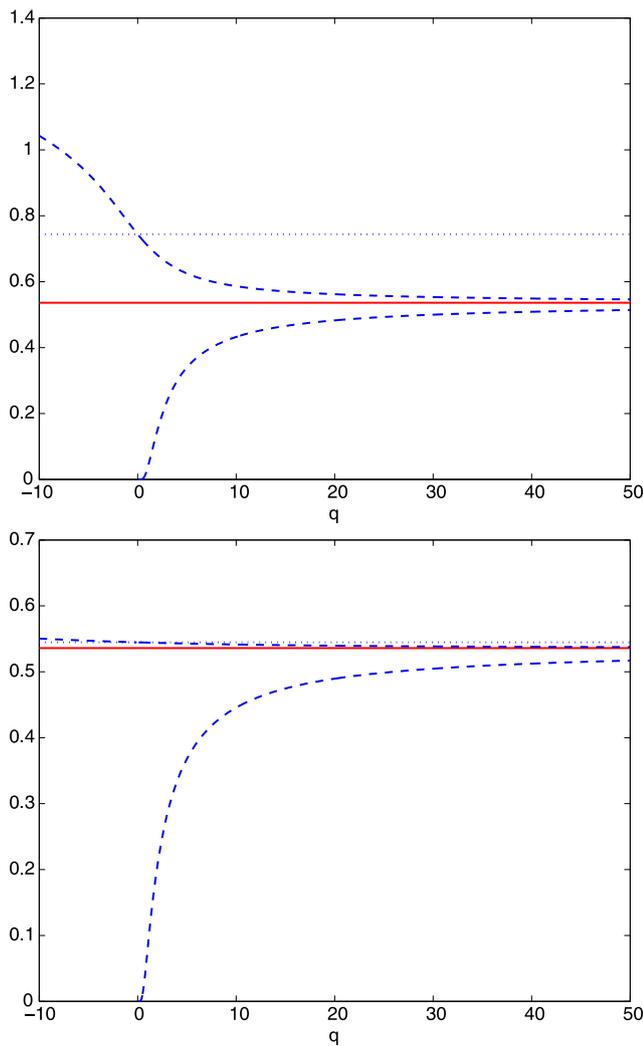
Fig. 4 Upper and lower bounds (*dashed lines*) on the value $\phi_{Mm}(\xi)$ for the 7-point maximin-distance design of Fig. 1 (left): (2) on the top, (6) on the bottom; the value of $\phi_{Mm}(\xi)$ is indicated by a solid line, $\overline{\phi}_{[0]}(\xi)$ (4) and $\overline{\phi}_{[NN,0]}(\xi)$ (8) are in *dotted lines*, respectively on the *top* and *bottom panels*

3.3 From maximin-distance to entropy maximization

Suppose that the n points $\mathbf{x}_i$ in ξ form n i.i.d. samples of a probability measure with density φ with respect to the Lebesgue measure on $\mathfrak{X}$. A natural statistical approach to measure of the quality of ξ in terms of its space-filling properties is to compare it in some way with samples from the uniform measure on $\mathfrak{X}$. Using discrepancy is a possibility, see Sect. 2.3. Another one relies on the property that the uniform distribution has maximum entropy among all distributions with finite support. This is the approach followed in this section.

The Rényi (1961) entropy of a random vector of $\mathbb{R}^d$ having the p.d.f. φ (that we shall call the Rényi entropy of φ) is

defined by

$$H_\alpha^*(\varphi) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} \varphi^\alpha(\mathbf{x})\,d\mathbf{x}, \quad \alpha \neq 1. \tag{9}$$

The Havrda–Charvát (1967) entropy (also called Tsallis 1988 entropy) of φ is defined by

$$H_\alpha(\varphi) = \frac{1}{\alpha - 1}\left(1 - \int_{\mathbb{R}^d} \varphi^\alpha(\mathbf{x})\,d\mathbf{x}\right), \quad \alpha \neq 1. \tag{10}$$

When α tends to 1, both H_α and H_α^* tend to the (Boltzmann-Gibbs-) Shannon entropy

$$H_1(\varphi) = -\int_{\mathbb{R}^d} \varphi(\mathbf{x}) \log[\varphi(\mathbf{x})]\,d\mathbf{x}. \tag{11}$$

Note that $H_\alpha^* = \log[1 - (\alpha-1)H_\alpha]/(1-\alpha)$ so that, for any α, $d(H_\alpha^*)/d(H_\alpha) > 0$ and the maximizations of H_α^* and H_α are equivalent; we can thus speak indifferently of α-entropy maximizing distributions.

The entropy H_α is a concave function of the density φ for $\alpha > 0$ (and convex for $\alpha < 0$). Hence, α-entropy maximizing distributions, under some specific constraints, are uniquely defined for $\alpha > 0$. In particular, the α-entropy maximizing distribution is uniform under the constraint that the distribution is finitely supported. The idea, suggested by Franco (2008), is thus to construct an estimator of the entropy of the design points $\mathbf{x}_i$ in ξ, considering them as if independently drawn with some probability distribution, and use this entropy estimator as a design criterion to be maximized. Note that this use of entropy (for a distribution in the space of input factors) is not directly connected Maximum-Entropy Sampling of Sect. 4.3 (for a distribution in the space of responses).

Many methods exist for the estimation of the entropy of a distribution from i.i.d. samples, and one may refer for instance to the survey papers (Hall and Morton 1993; Beirlant et al. 1997) for an overview. We shall consider three, because they have either already been used in the context of experimental design or are directly connected with other space-filling criteria. In a fourth paragraph, entropy decomposition is used to avoid the collapsing of design points when considering lower dimensional subspaces.

Plug-in method based on kernel density estimation The approach is in two steps. First, one construct an estimator of the p.d.f. φ by a kernel method as

$$\hat{\varphi}_n(\mathbf{x}) = \frac{1}{n\,h_n^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right), \tag{12}$$

where $K(\cdot)$ denotes the kernel and h_n the window width. The choices of $K(\cdot)$ and h_n are important issues when the objective is to obtain an accurate estimation of φ and there

exists a vast literature on that topic. However, this should not be too critical here since we only need to get an entropy estimator that yields a reasonable space-filling criterion. A common practice in density estimation is to take h_n decreasing with n, e.g. as $n^{-1/(d+4)}$, see Scott (1992, p. 152), and to use a p.d.f. for $K(\cdot)$, e.g. that of the standard normal distribution in $\mathbb{R}^d$. A kernel with bounded support could be more indicated since $\mathfrak{X}$ is bounded, but the choice of the window width might then gain importance. When a kernel-based prediction method is to be used, it seems natural to relate $K(\cdot)$ and h_n to the kernel used for prediction (to the correlation function in the case of kriging); this will be considered in Sect. 4.3.

In a second step, the entropy H_α^* or H_α is estimated by replacing the unknown φ by the estimate $\hat{\varphi}_n$ in the definition. In order to avoid the evaluation of multidimensional integrals, a Monte-Carlo estimator can be used, namely $\hat{H}_1^n = -\sum_{i=1}^n \log[\hat{\varphi}_n(\mathbf{x}_i)]$ for Shannon entropy, and

$$\hat{H}_\alpha^n = \frac{1}{\alpha - 1}\left[1 - \sum_{i=1}^n \hat{\varphi}_n^{\alpha-1}(\mathbf{x}_i)\right] \tag{13}$$

for H_α with $\alpha \neq 1$. A surprising result about normal densities is that when $K(\cdot)$ is the p.d.f. of the normal $\mathcal{N}(\mathbf{0}, \mathbf{I})$, then

$$\int_{\mathbb{R}^d} \hat{\varphi}_n^2(\mathbf{x})\, d\mathbf{x} = \frac{1}{2^d \pi^{d/2} n^2 h_n^d} \sum_{i,j} \exp\left[-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4h_n^2}\right];$$

that is, a Monte-Carlo evaluation gives the exact value of the integral in (9, 10) for $\varphi = \hat{\varphi}_n$ when $\alpha = 2$. This is exploited in Bettinger et al. (2008, 2009) for the sequential construction of an experiment with the objective of inverting an unknown system.

Nearest-neighbor (NN) distances The following estimator of $H_\alpha(\varphi)$ is considered in Leonenko et al. (2008)

$$\hat{H}_{n,k,\alpha} = \frac{1 - \frac{[(n-1)C_k V_d]^{1-\alpha}}{n}\sum_{i=1}^n (d_{k,i}^*)^{d(1-\alpha)}}{\alpha - 1}, \tag{14}$$

where $V_d = \pi^{d/2}/\Gamma(d/2 + 1)$ is the volume of the unit ball $\mathfrak{B}(0,1)$ in $\mathbb{R}^d$, $C_k = [\Gamma(k)/\Gamma(k+1-\alpha)]^{1/(1-\alpha)}$ and $d_{k,i}^*$ is the k-th nearest-neighbor distance from $\mathbf{x}_i$ to some other $\mathbf{x}_j$ in the sample (that is, from the $n-1$ distances d_{ij}, $j \neq i$, we form the order statistics $d_{1,i}^* = d_i^* \leq d_{2,i}^* \leq \cdots \leq d_{n-1,i}^*$). The L_2-consistency of this estimator is proved in Leonenko et al. (2008) for any $\alpha \in (1, (k+1)/2)$ when $k \geq 2$ (respectively $\alpha \in (1, 1 + 1/[2d])$ when $k = 1$) if f is bounded. For $\alpha < 1$, one may refer to Penrose and Yukich (2011) for the a.s. and L_2 convergence of $\hat{H}_{n,k,\alpha}$ to $H_\alpha(\varphi)$; see also the results of Yukich (1998) on the subadditivity of Euclidean functionals.

For $\alpha = 1$ (Shannon entropy), the following estimator is considered in Kozachenko and Leonenko (1987), Leonenko et al. (2008)

$$\hat{H}_{n,k,1} = \frac{d}{n}\sum_{i=1}^n \log d_{k,i}^* + \log(n-1) + \log(V_d) - \Psi(k),$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

Maximizing $\hat{H}_{n,1,\alpha}$ for $\alpha > 1$ thus corresponds to maximizing $\underline{\phi}_{[NN,q]}(\xi)$ with $q = d(\alpha - 1)$, see (7). For $1 - 1/d \leq \alpha \leq 1$, the criterion $\hat{H}_{n,1,\alpha}$ is still eligible for space-filling, its maximization is equivalent to that of $\underline{\phi}_{[NN,q]}(\xi)$ with $q \in [-1, 0]$; for instance, the maximization of $\hat{H}_{N,1,1}$ is equivalent to the maximization of $\overline{\phi}_{[NN,0]}(\xi)$, see (8).

Several comments should be made, however, that will temper the feeling that L_q-regularization of maximin-distance design and maximization of NN-estimates of entropy are equivalent.

First, these estimators rely on the assumption that the $\mathbf{x}_i$ are i.i.d. with some p.d.f. φ. However, optimizing the locations of points with respect to some design criterion makes the corresponding sample completely atypical. The associated value of the estimator is therefore atypical too. Consider for instance the maximin-distance design ξ_{Mm}^* on $[0, 1]$, defined by $x_i = (i - 1)/(n - 1)$, $i = 1, \ldots, n$. Direct calculation gives $\hat{H}_{n,1,\alpha}(\xi_{Mm}^*) = [1 - 2^{1-\alpha}/\Gamma(2 - \alpha)]/(\alpha - 1)$, which is greater than 1 for $0 < \alpha < 2$, with a maximum $\gamma + \log(2) \simeq 1.2704$ when α tends to 1. On the other hand, the maximum value of $H(\varphi)$ for φ a p.d.f. on $[0, 1]$ is obtained for the uniform distribution $\varphi^*(x) = 1$ for all x, with $H(\varphi^*) = 0$.

Second, even if the design points in ξ are generated randomly, using k-th NN distances with $k > 1$ does not make much sense in terms of measuring the space-filling performance. Indeed, when using $\hat{H}_{n,k,\alpha}$ with $k > 1$, a design obtained by fusing sets of k points will show a higher entropy than a design with all points separated. This is illustrated by the simple example of a maximin-distance design on the real line. For the design ξ_{Mm}^* with n points we have

$$\hat{H}_{n,2,\alpha}(\xi_{Mm}^*) = \frac{1 - \frac{2^{1-\alpha}}{\Gamma(3-\alpha)}\left[1 + \frac{2(2^{1-\alpha}-1)}{n}\right]}{\alpha - 1}.$$

Suppose that $n = 2m$ and consider the design $\tilde{\xi}_{Mm}^*$ obtained by duplicating the maximin-distance design with m points; that is, $x_i = (i - 1)/(m - 1)$, $i = 1, \ldots, m$, and $x_i = (i - m - 1)/(m - 1)$, $i = m + 1, \ldots, 2m$. We get

$$\hat{H}_{n,2,\alpha}(\tilde{\xi}_{Mm}^*) = \frac{1 - \frac{2^{1-\alpha}}{\Gamma(3-\alpha)}\left[2 + \frac{1}{m-1}\right]^{1-\alpha}}{\alpha - 1}$$

and $\hat{H}_{n,2,\alpha}(\tilde{\xi}_{Mm}^*) > \hat{H}_{n,2,\alpha}(\xi_{Mm}^*)$ for $\alpha \in (0, 3)$. We should thus restrict our attention to $\hat{H}_{n,k,\alpha}$ with $k = 1$. The range of

values of α for which the strong consistency of the estimator is ensured is then restricted to $\alpha < 1 + 1/[2d]$. Strictly speaking, it means that the maximization of $\underline{\phi}_{[NN,q]}(\xi)$ can be considered as the maximization of a *NN* entropy estimator for $q < 1/2$ only.

Minimum-spanning-tree Redmond and Yukich (1996), Yukich (1998) use the subadditivity of some Euclidean functionals on graphs to construct strongly consistent estimators of $H_\alpha^*(\varphi)$ (9) for $0 < q < 1$, up to some bias term independent of φ and related to the graph properties. Their approach covers the case of the graph of k-th *NN* (where the bias constant depends on the value of k through C_k, see (14)), but also the graphs corresponding to the solution of a traveling salesman problem, or the minimum spanning tree (MST). In each case, the entropy estimate is based on $\sum_{i=1}^M d_i^{d(1-\alpha)}$, where the d_i denote the lengths of the M edges of the graph, with $M = n - 1$ for the MST and $M = n$ for the traveling-salesman tour and NN graphs.

The MST constructed from the $\mathbf{x}_i$ has already been advocated as a useful tool to assess the quality of designs in terms of their space-filling properties: in Franco et al. (2009), the empirical mean and variance of the lengths of edges d_i of the MST are used to characterize classes of designs (such as random, low discrepancy sequences, maximin-distance and minimax-distance designs); designs with large empirical means are considered preferable. With the same precautions as above for NN entropy estimation, the maximization of the function $(\sum_{i=1}^{n-1} d_i^{-q})^{-1/q}$ in the MST constructed from the $\mathbf{x}_i$ is related to the maximization of an entropy estimator of the distribution of the $\mathbf{x}_i$; in particular, the maximization of the empirical mean of the edge lengths ($q = -1$) forms a reasonable objective.

Entropy decomposition to avoid collapsing on projections Let $\mathbf{u}$ and $\mathbf{v}$ be two independent random vectors respectively in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$. Define $\mathbf{x} = (\mathbf{u}^\top, \mathbf{v}^\top)^\top$ and let $\varphi(\mathbf{u}, \mathbf{v})$ denote the joint density for $\mathbf{x}$. Let $\varphi_1(\mathbf{u})$ and $\varphi_2(\mathbf{v})$ be the marginal densities for $\mathbf{u}$ and $\mathbf{v}$ respectively, so that $\varphi(\mathbf{u}, \mathbf{v}) = \varphi_1(\mathbf{u})\varphi_2(\mathbf{v})$. It is well known that the Shannon and Rényi entropies (11) and (9) satisfy the additive property $H_\alpha^*(\varphi) = H_\alpha^*(\varphi_1) + H_\alpha^*(\varphi_2)$, $\alpha \in \mathbb{R}$ (extensivity property of Shannon and Rényi entropies) while for the Tsallis entropy (10) one has $H_\alpha(\varphi) = H_\alpha(\varphi_1) + H_\alpha(\varphi_2) + (1 - \alpha)H_\alpha(\varphi_1)H_\alpha(\varphi_2)$ (non-extensivity Tsallis entropy, with α the parameter of non-extensivity).

Now, when φ is the p.d.f. of the uniform distribution on the unit cube $\mathfrak{X} = [0, 1]^d$, one can consider all one-dimensional projections $\{\mathbf{x}\}_i$, $i = 1, \ldots, d$, and $H_\alpha^*(\varphi) = \sum_{i=1}^d H_\alpha^*(\varphi_i)$ with φ_i the density of the i-th projection $\{\mathbf{x}\}_i$. This can be used to combine a criterion related to space-filling in $\mathfrak{X}$ with criteria related to space-filling along one-

dimensional projections. Consider for instance the *NN* estimator of $H_\alpha^*(\varphi)$ of Leonenko et al. (2008) (for $\alpha \neq 0$),

$$\hat{H}_{n,k,\alpha}^* = \frac{\log\left\{ \frac{[(n-1)\,C_k\,V_d]^{1-\alpha}}{n} \sum_{i=1}^n (d_{k,i}^*)^{d(1-\alpha)} \right\}}{1-\alpha}. \tag{15}$$

For $k = 1$ ($k > 1$ does not fit with the space-filling requirement, see the discussion above), we have

$$\hat{H}_{n,1,\alpha}^* = \frac{1}{1-\alpha} \log\left[\sum_{i=1}^n (d_i^*)^{d(1-\alpha)} \right] + A(\alpha, d, n),$$

where $A(\alpha, d, n)$ is a constant that does not depend on ξ. A suitable criterion (to be maximized) that simultaneously takes into account the space-filling objectives in $\mathfrak{X}$ and along all one-dimensional projections is thus

$$\frac{1}{1-\alpha} \left\{ (1 - \gamma) \log\left[\sum_{i=1}^n (d_i^*)^{d(1-\alpha)} \right] \right.$$
$$\left. + \gamma \sum_{j=1}^d \log\left[\sum_{i=1}^n (d_{j_i}^*)^{(1-\alpha)} \right] \right\},$$

with $\gamma \in (0, 1)$ and $d_{j_i}^* = \min_{k \neq i} |\{\mathbf{x}_i\}_j - \{\mathbf{x}_k\}_j|$, or equivalently, setting $q = d(\alpha - 1)$,

$$\phi_{q,1P}(\xi) = (1 - \gamma) \log[\underline{\phi}_{[NN,q]}(\xi)]$$
$$+ \frac{\gamma}{d} \sum_{j=1}^d \log[\underline{\phi}_{[NN,q/d,j]}(\xi)],$$

where $\underline{\phi}_{[NN,q]}(\xi)$ is given by (7) and $\underline{\phi}_{[NN,q,j]}(\xi) = [\sum_{i=1}^n (d_{j_i}^*)^{-q}]^{-1/q}$. Letting q tend to infinity, we get the following compromise between maximin-distance designs on $\mathfrak{X}$ and on its one-dimensional projections

$$\phi_{\infty,1P}(\xi) = (1 - \gamma) \log[\phi_{Mm}(\xi)] + \frac{\gamma}{d} \sum_{j=1}^d \log[\phi_{Mm_j}(\xi)],$$

with $\phi_{Mm_j}(\xi) = \min_i d_{j_i}^* = \min_{k \neq i} |\{\mathbf{x}_i\}_j - \{\mathbf{x}_k\}_j|$. One should note that there exists a threshold $\gamma^* = \gamma^*(d, n)$ such that the optimal design associated with any $\gamma \geq \gamma^*$ is a maximin Lh design.

When $\alpha = 1$ (Shannon entropy), identical developments lead to the same criterion $\phi_{q,1P}(\xi)$ as above with q set to zero, $\overline{\phi}_{[NN,0]}(\xi)$ defined by (8) and $\overline{\phi}_{[NN,0,j]}(\xi) = \exp\{[\sum_{i=1}^n \log(d_{j_i}^*)]/n\}$.

Other combinations of criteria are possible; one may for instance maximize a space-filling criterion in $\mathfrak{X}$ under constraints on the space-filling properties along one-dimensional projections. Also, projections on higher dimensional subspaces can be taken into account in a similar way using the appropriate decomposition of the entropy of joint densities.

4 Model-based design: the case of kriging

In the following we assume that we have a reasonable simplified model (the so called emulator) for the unknown function $f(\cdot)$, whose evaluation at a given point $\mathbf{x}$ relies on a computer code (evaluations at the design points in ξ form a computer experiment).

4.1 Gaussian-process model and kriging

In particular, consider the following spatial random field

$$Y(\mathbf{x}) = f(\mathbf{x}) = \eta(\mathbf{x}, \beta) + Z(\mathbf{x}), \qquad (16)$$

where β is an unknown vector of parameters in $\mathbb{R}^p$ and the random term $Z(\mathbf{x})$ has zero mean, (unknown) variance σ_Z^2 and a parameterized spatial error correlation structure such that $\mathbb{E}\{Z(\mathbf{u})Z(\mathbf{v})\} = \sigma_Z^2 C(\mathbf{u}, \mathbf{v}; \nu)$. It is often assumed that the deterministic term has a linear structure, that is, $\eta(\mathbf{x}, \beta) = \mathbf{r}^\top(\mathbf{x})\beta$, and that the random field $Z(\mathbf{x})$ is Gaussian, allowing the estimation of β, σ_Z and ν by Maximum Likelihood. This setup is used in such diverse areas of spatial data analysis (see Cressie 1993) as mining, hydrogeology, natural resource monitoring and environmental science, etc., and has become the standard modeling paradigm in computer simulation experiments, following the seminal paper of Sacks et al. (1989). Here, $\lim_{\mathbf{v} \to \mathbf{u}} C(\mathbf{u}, \mathbf{v}; \nu) = C(\mathbf{u}, \mathbf{u}; \nu) = 1$ for all $\mathbf{u} \in \mathfrak{X}$.

Denote by $\hat{Y}(\mathbf{x}|\xi)$ the Best Linear Unbiased Predictor (BLUP) of $Y(\mathbf{x})$ based on the design points in ξ and associated observations $\mathbf{y}(\xi) = [Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n)]^\top$. Optimal design in this context is usually performed by minimizing a functional of $\text{var}[\hat{Y}(\mathbf{x}|\xi)] = \mathbb{E}[(\hat{Y}(\mathbf{x}|\xi) - Y(\mathbf{x}))^2]$ at $\mathbf{x}$, the unconditional Mean-Squared Prediction Error (MSPE), also called the kriging variance. Keeping ν fixed, then in the linear setting (universal kriging, with $\eta(\mathbf{x}, \beta) = \mathbf{r}^\top(\mathbf{x})\beta$, generally a polynomial in $\mathbf{x}$), the BLUP takes the form

$$\hat{Y}(\mathbf{x}|\xi) = \mathbf{r}^\top(\mathbf{x})\hat{\beta} + \mathbf{c}_\nu^\top(\mathbf{x})\mathbf{C}_\nu^{-1}[\mathbf{y}(\xi) - \mathbf{R}\hat{\beta}], \qquad (17)$$

where $\{\mathbf{c}_\nu(\mathbf{x})\}_i = C(\mathbf{x}, \mathbf{x}_i; \nu)$, $\{\mathbf{C}_\nu\}_{ij} = C(\mathbf{x}_i, \mathbf{x}_j; \nu)$, $i, j = 1, \ldots, n$, and $\hat{\beta} = \hat{\beta}_\nu$ is the weighted least-squares estimator of β in the linear regression model, that is,

$$\hat{\beta}_\nu = [\mathbf{R}^\top\mathbf{C}_\nu^{-1}\mathbf{R}]^{-1}\mathbf{R}^\top\mathbf{C}_\nu^{-1}\mathbf{y}(\xi),$$

with $\mathbf{R} = [\mathbf{r}(\mathbf{x}_1), \ldots, \mathbf{r}(\mathbf{x}_n)]^\top$. Notice that $\hat{Y}(\mathbf{x}|\xi)$ does not depend on σ_Z and that $\hat{Y}(\mathbf{x}_i|\xi) = Y(\mathbf{x}_i)$ for all i (the predictor is a perfect interpolator). We can write

$$\hat{Y}(\mathbf{x}|\xi) = \mathbf{v}_\nu^\top(\mathbf{x})\mathbf{y}(\xi)$$

where

$$\mathbf{v}_\nu(\mathbf{x}) = \mathbf{C}_\nu^{-1}[\mathbf{I}_n - \mathbf{R}(\mathbf{R}^\top\mathbf{C}_\nu^{-1}\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{C}_\nu^{-1}]\mathbf{c}_\nu(\mathbf{x})$$
$$+ \mathbf{C}_\nu^{-1}\mathbf{R}(\mathbf{R}^\top\mathbf{C}_\nu^{-1}\mathbf{R})^{-1}\mathbf{r}(\mathbf{x}) \qquad (18)$$

with $\mathbf{I}_n$ the n-dimensional identity matrix. The MSPE is given by

$$\text{MSPE}_\xi(\mathbf{x}, \sigma_Z^2, \nu) = \sigma_Z^2 \Big\{ 1 - \mathbf{c}_\nu^\top(\mathbf{x})\mathbf{C}_\nu^{-1}\mathbf{c}_\nu(\mathbf{x})$$
$$+ \mathbf{g}_\nu^\top(\mathbf{x})[\mathbf{R}^\top\mathbf{C}_\nu^{-1}\mathbf{R}]^{-1}\mathbf{g}_\nu(\mathbf{x}) \Big\}$$

with $\mathbf{g}_\nu(\mathbf{x}) = \mathbf{r}(\mathbf{x}) - \mathbf{R}^\top\mathbf{C}_\nu^{-1}\mathbf{c}_\nu(\mathbf{x})$. Note that the MSPE depends on (σ_Z^2, ν), with σ_Z^2 intervening only as a multiplicative factor. We shall denote by $\rho^2(\mathbf{x}) = \rho_\xi^2(\mathbf{x}, \nu)$ the normalized kriging variance,

$$\rho_\xi^2(\mathbf{x}, \nu) = \text{MSPE}_\xi(\mathbf{x}, \sigma_Z^2, \nu)/\sigma_Z^2 \qquad (19)$$

and omit the dependence in ξ and ν when it does not lead to ambiguities. Note that $\rho_\xi^2(\mathbf{x}_i, \nu) = 0$ for all i.

We suppose for the moment that ν is known (the investigation of the (more realistic) situation where ν is unknown is postponed to Sect. 5) and omit the dependence on ν in the notations. It is sufficient in many circumstances to take $\eta(\mathbf{x}, \beta) = \beta$, that is, to model the unknown function as the realization of a stochastic process with unknown mean value. In that case, the normalized kriging variance is simply

$$\rho^2(\mathbf{x}) = 1 - \mathbf{c}^\top(\mathbf{x})\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}) + \frac{[1 - \mathbf{c}^\top(\mathbf{x})\mathbf{C}^{-1}\mathbf{1}]^2}{\mathbf{1}^\top\mathbf{C}^{-1}\mathbf{1}}, \qquad (20)$$

with $\mathbf{1}$ the n-dimensional vector of ones.

A natural approach for designing an experiment is to choose ξ that minimizes a functional of the kriging variance, for instance its integrated value $\phi_A(\xi) = \int_{\mathfrak{X}} \rho^2(\mathbf{x}) \, d\mathbf{x}$ (generally evaluated by a discrete sum over a finite grid) or the G-optimality criterion (by analogy with G-optimal design for regression models, see Kiefer and Wolfowitz 1960)

$$\phi_G(\xi) = \max_{\mathbf{x} \in \mathfrak{X}} \rho^2(\mathbf{x}). \qquad (21)$$

Johnson et al. (1990) show that a minimax-optimal design is asymptotically G-optimal when the correlation function has the form $C^k(\cdot)$ with k tending to infinity (i.e., it tends to be G-optimal for weak correlations). See also Joseph (2006) who motivates the use of minimax-optimal designs for his limit-kriging approach. The evaluation of $\phi_G(\xi)$ at any given ξ requires the solution of a maximization problem over $\mathfrak{X}$, which makes the optimization of $\phi_G(\cdot)$ a rather exhausting task. Replacing the optimization over $\mathfrak{X}$ by a grid search over a finite subset $\mathfrak{X}_N \subset \mathfrak{X}$ is often used; another option is to perform a Delaunay tessellation of the points in ξ plus the vertices of $\mathfrak{X} = [0, 1]^d$ and initialize a local search for the maximum of $\rho^2(\mathbf{x})$ at the center of each Delaunay simplex (see Sect. 2.4). A third option, considered below, consists in using an upper bound on $\phi_G(\xi)$.

4.2 Upper bounds on the kriging variance

We only consider isotropic processes, with correlation depending on the Euclidean distance between points, i.e. satisfying $\mathbb{E}\{Z(\mathbf{u})Z(\mathbf{v})\} = \sigma_Z^2 C(\|\mathbf{u} - \mathbf{v}\|; \nu)$, $(\mathbf{u}, \mathbf{v}) \in \mathfrak{X}^2$. The extension to the non-isotropic case should not raise major difficulties through an appropriate change of metric in $\mathfrak{X}$. We suppose that the radial correlation function $C(\cdot; \nu)$ is non-increasing and non-negative on $\mathbb{R}^+$. Denote $C_{Mm} = C(\phi_{Mm})$, $C_{mM} = C(\phi_{mM})$ (we omit the dependence in ξ where there is no ambiguity) and

$$\rho_0^2(\mathbf{x}) = 1 - \mathbf{c}^\top(\mathbf{x})\mathbf{C}^{-1}\mathbf{c}(\mathbf{x})$$

the (normalized) kriging variance when β is known. The objective of this section is to construct upper bounds on $\max_{\mathbf{x} \in \mathfrak{X}} \rho_0^2(\mathbf{x})$ and $\max_{\mathbf{x} \in \mathfrak{X}} \rho^2(\mathbf{x})$, see (20).

From the developments given in Appendix C, we obtain the bound

$$\rho_0^2(\mathbf{x}) \leq 1 - \frac{C_{mM}^2}{\lambda_{\max}(\mathbf{C})},$$

for the case where β is known and, for a weak enough correlation, the approximate bound

$$\rho^2(\mathbf{x}) \leq 1 - \frac{C_{mM}^2}{\lambda_{\max}(\mathbf{C})} + \frac{(1 - C_{mM}\,\underline{u})^2}{\mathbf{1}^\top\mathbf{C}^{-1}\mathbf{1}},$$

where $\underline{u} = \min_i\{\mathbf{C}^{-1}\mathbf{1}\}_i$ when β is unknown.

Using further approximations, one can obtain bounds that depend on C_{mM} and C_{Mm} but not on $\mathbf{C}$, see Appendix C. We obtain

$$\rho_0^2(\mathbf{x}) \leq \bar{\rho}_0^2(\mathbf{x}) = 1 - \frac{\bar{c}(\mathbf{x})^2}{1 + (n-1)C_{Mm}}, \tag{22}$$

where $\bar{c}(\mathbf{x}) = \max_i\{\mathbf{c}(\mathbf{x})\}_i$, and thus

$$\max_{\mathbf{x} \in \mathfrak{X}} \rho_0^2(\mathbf{x}) \leq \bar{\rho}_0^2 = 1 - \frac{C_{mM}^2}{1 + (n-1)C_{Mm}}. \tag{23}$$

Also, when the correlation is weak enough,

$$\rho^2(\mathbf{x}) \leq \bar{\rho}^2(\mathbf{x}) = \bar{\rho}_0^2(\mathbf{x}) + \frac{1 + (n-1)C_{Mm}}{n}R^2(\mathbf{x}) \tag{24}$$

with $\bar{\rho}_0^2(\mathbf{x})$ given by (22) and

$$R^2(\mathbf{x}) = \left[1 - \bar{c}(\mathbf{x})\frac{1 - (n-1)C_{Mm}}{1 - (n-1)C_{Mm}^2}\right]^2,$$

which gives

$$\max_{\mathbf{x} \in \mathfrak{X}} \rho^2(\mathbf{x}) \leq \bar{\rho}^2 = \bar{\rho}_0^2 + \left[1 - C_{mM}\frac{1 - (n-1)C_{Mm}}{1 - (n-1)C_{Mm}^2}\right]^2$$
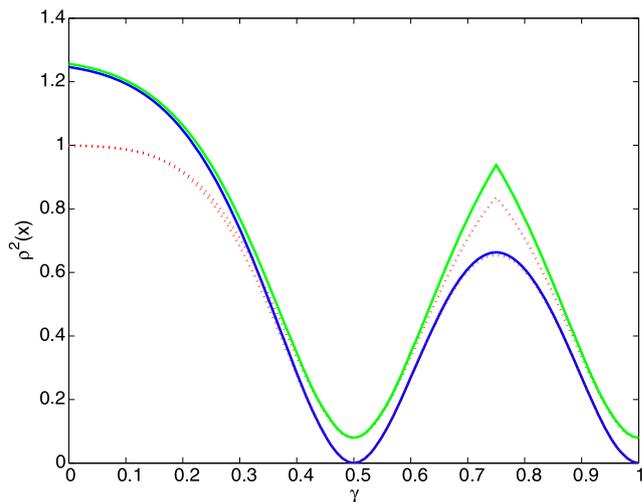
Fig. 5 Kriging variance (normalized) and bounds with, in *solid lines* from top to bottom, $\bar{\rho}^2(\mathbf{x})$ given by (24) and the exact (normalized) kriging variance $\rho^2(\mathbf{x})$; the values of $\rho_0^2(\mathbf{x})$ and of its upper bound $\bar{\rho}_0^2(\mathbf{x})$ (22) are indicated in *dotted lines*; $\mathbf{x} = (0, 0)^\top + \gamma(1, 1)^\top$

$$\times \frac{1 + (n-1)C_{Mm}}{n}, \tag{25}$$

with $\bar{\rho}_0^2$ given by (23). More accurate bounds are given in Griffith (2003) when the points in ξ follow a regular pattern. Similar ideas could be applied to the limit kriging predictor of Joseph (2006).

Example 1 We consider a two-dimensional example with four design points, three at the corners $(1, 0)$, $(1, 1)$, $(0, 1)$ and one in the center $(1/2, 1/2)$ of $\mathfrak{X} = [0, 1]^2$. Prediction is considered along the diagonal going from the origin to the corner $(1, 1)$, with $\mathbf{x} = (0, 0)^\top + \gamma(1, 1)^\top$, $\gamma \in [0, 1]$. The correlation function is $C(t) = (1 - t)^4(1 + 4t)$ with $C(t) = 0$ for $t \geq 1$, see Wendland (2005). Notice that $\mathbf{C}$ has the form (41) with $C_{Mm} = C(\sqrt{2}/2) \simeq 0.0282$. Figure 5 presents the (normalized) kriging variances $\rho_0^2(\mathbf{x})$ and $\rho^2(\mathbf{x})$ together with the bounds constructed above. We have $\rho^2(\mathbf{x}) = \rho_0^2(\mathbf{x}) = 0$ at the design points $(1/2, 1/2)$ and $(1, 1)$. Note that the bounds $\bar{\rho}^2(\mathbf{x})$ and $\bar{\rho}_0^2(\mathbf{x})$ although not tight everywhere (in particular, they are pessimistic at the design points) give a reasonable approximation of the behavior of $\rho^2(\mathbf{x})$ and $\rho_0^2(\mathbf{x})$ respectively. Also note that the global bounds (23) and (25) (reached at $\mathbf{x} = (0, 0)$) are rather tight.

Example 2 We consider a one-dimensional example with the 5-point minimax-optimal design $\xi_{mM}^* = (0.1, 0.3, 0.5, 0.7, 0.9)$ in $\mathfrak{X} = [0, 1]$ for the correlation $C(t) = \exp(-10t)$. Figure 6 presents the (normalized) kriging variances $\rho_0^2(x)$ and $\rho^2(x)$ together with the bounds constructed above as x varies in $\mathfrak{X}$. The bounds $\bar{\rho}_0(x)$ given by (22) and $\bar{\rho}(x)$ given by (24) are nowhere tight (neither are the global bounds
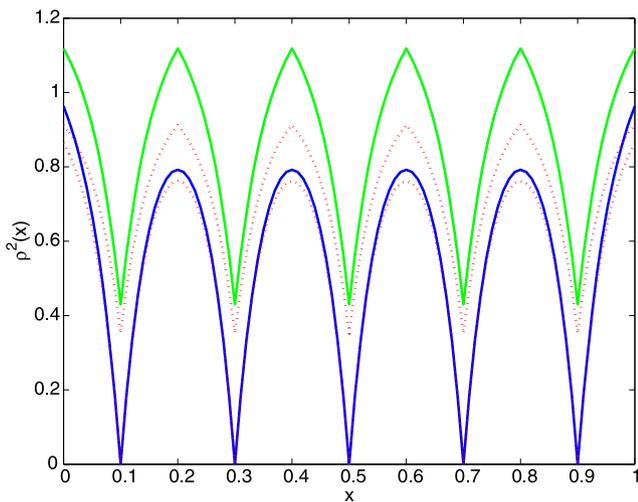
Fig. 6 Kriging variance (normalized) and bounds for the 5-point minimax-optimal design with, in *solid lines* from top to bottom, $\bar{\rho}(x)$ given by (24) and the exact (normalized) kriging variance $\rho^2(x)$; the values of $\rho_0^2(x)$ and of its upper bound $\bar{\rho}_0(x)$ (22) are indicated in *dotted lines*

Fig. 8 $\max_{x \in \mathfrak{X}} \rho^2(x)$ (*solid line*) and $\bar{\rho}^2$ (25) (*dashed line*) as functions of α for the design $\xi(\alpha)$ when $\nu = 7$ (*left*) and $\nu = 40$ (*right*); the maximin-optimal design corresponds to $\alpha = 0$, the minimax-optimal design to $\alpha = 0.1$ (*dotted vertical line*)

Fig. 7 Kriging variance (normalized) and bounds for the 5-point maximin-optimal design with, in *solid lines* from top to bottom, $\bar{\rho}(x)$ given by (24) and the exact (normalized) kriging variance $\rho^2(x)$; the values of $\rho_0^2(x)$ and of its upper bound $\bar{\rho}_0(x)$ (22) are indicated in *dotted lines*

$\bar{\rho}^2$ and $\bar{\rho}_0$ given by (23) and (25)), but the behavior of the kriging variance as a function of x is satisfactorily reproduced. Figure 7 presents the same information for the 5-point maximin-optimal design $\xi_{Mm}^* = (0, 0.25, 0.5, 0.75, 1)$.

For a small enough correlation, a minimax-optimal design ensures a smaller value for $\max_{\mathbf{x} \in \mathfrak{X}} \rho^2(\mathbf{x})$ than a maximin-optimal design, see Johnson et al. (1990). One might hope that this tendency will also be observed when using the upper bound $\bar{\rho}^2$ given by (25). This seems to be the case, as the following continuation of Example 2 illustrates.

Example 2 (Continued) We consider the following family of 5-point designs: $\xi(\alpha) = (\alpha, \alpha + (1 - 2\alpha)/4, \alpha + (1 - 2\alpha)/2, \alpha + 3(1 - 2\alpha)/4, 1 - \alpha)$, which includes ξ_{Mm}^* (for $\alpha = 0$) and ξ_{mM}^* (for $\alpha = 0.1$). The correlation function is $C(t) = \exp(-\nu t)$. Figure 8 presents $\max_{x \in \mathfrak{X}} \rho^2(x)$ and $\bar{\rho}^2$ given by (25) as functions of α in the strong (left, $\nu = 7$) and weak (right, $\nu = 40$) correlation cases. Although the curves do not reach their minimum value for the same α, they indicate the same preference between ξ_{Mm}^* and ξ_{mM}^*.

4.3 Maximum-entropy sampling

Suppose that $\mathfrak{X}$ is discretized into the finite set $\mathfrak{X}_N$ with N points. Consider $\mathbf{y}_N$, the vector formed by $Y(\mathbf{x})$ for $\mathbf{x} \in \mathfrak{X}_N$, and $\mathbf{y}(\xi)$, the vector obtained for $\mathbf{x} \in \xi$. For any random $\mathbf{y}$ with p.d.f. $\varphi(\cdot)$ denote ent($\mathbf{y}$) the (Shannon) entropy of φ, see (11). Then, from a classical theorem in information theory (see, e.g., Ash 1965, p. 239),

$$\text{ent}(\mathbf{y}_N) = \text{ent}(\mathbf{y}_\xi) + \text{ent}(\mathbf{y}_{N \setminus \xi} | \mathbf{y}_\xi) \tag{26}$$

where $\mathbf{y}_{N \setminus \xi}$ denotes the vector formed by $Y(\mathbf{x})$ for $\mathbf{x} \in \mathfrak{X}_N \setminus \xi$ and ent($\mathbf{y}|\mathbf{w}$), the conditional entropy, is the expectation with respect to $\mathbf{w}$ of the entropy of the conditional p.d.f. $\varphi(\mathbf{y}|\mathbf{w})$, that is,

$$\text{ent}(\mathbf{y}|\mathbf{w}) = -\int \varphi(\mathbf{w}) \left(\int \varphi(\mathbf{y}|\mathbf{w}) \log[\varphi(\mathbf{y}|\mathbf{w})] \, d\mathbf{y} \right) d\mathbf{w}.$$

The argumentation in Shewry and Wynn (1987) is as follows: since ent($\mathbf{y}_N$) in (26) is fixed, the natural objective of minimizing ent($\mathbf{y}_{N \setminus \xi} | \mathbf{y}_\xi$) can be fulfilled by maximizing ent($\mathbf{y}_\xi$). When $Z(\mathbf{x})$ in (16) is Gaussian, ent($\mathbf{y}_\xi$) =

$(1/2) \log \det(\mathbf{C}) + (n/2)[1 + \log(2\pi)]$, and Maximum-Entropy-Sampling corresponds to maximizing $\det(\mathbf{C})$, which is called D-optimal design (by analogy with optimum design in a parametric setting). One can refer to Wynn (2004) for further developments.

Johnson et al. (1990) show that a maximin-optimal design is asymptotically D-optimal when the correlation function has the form $C^k(\cdot)$ with k tending to infinity (i.e., it tends to be D-optimal for weak correlations). We have considered in Sect. 3.3 the design criterion (to be maximized) given by a plug-in kernel estimator $\hat{H}_\alpha^n$ of the distribution of the $\mathbf{x}_i$, see (13) and (12). When $\alpha = 2$, the maximization of $\hat{H}_\alpha^n$ is equivalent to the minimization of

$$\phi(\xi) = \sum_{i,j} K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{h_n}\right).$$

A natural choice in the case of prediction by kriging is $K[(\mathbf{u} - \mathbf{v})/h_n] = C(\|\mathbf{u} - \mathbf{v}\|)$, which yields $\phi(\xi) = \sum_{i,j} \{\mathbf{C}\}_{ij}$. Since $\mathbf{C}$ has all its diagonal elements equal to 1, its determinant is maximum when the off-diagonal elements are zero, that is when $\phi(\xi) = n$. Also note that

$$1 - (n-1)C_{Mm} \leq \lambda_{\min}(\mathbf{C}) \leq \frac{\phi(\xi)}{n} = \frac{\mathbf{1}^\top \mathbf{C} \mathbf{1}}{n}$$
$$\leq \lambda_{\max}(\mathbf{C}) \leq 1 + (n-1)C_{Mm}.$$

The upper bound on $\lambda_{\max}(\mathbf{C})$ is derived in Appendix C. The lower bound is obtained from $\lambda_{\min}(\mathbf{C}) \geq t - s\sqrt{n-1}$ with $t = \mathrm{tr}(\mathbf{C})/n$ and $s^2 = \mathrm{tr}(\mathbf{C}^2)/n - t^2$, see Wolkowicz and Styan (1980). Since $\{\mathbf{C}\}_{ij} = \{\mathbf{C}\}_{ji} \leq C_{Mm}$ for all $i \neq j$, we get $\mathrm{tr}(\mathbf{C}) = n$ and $\mathrm{tr}(\mathbf{C}^2) \leq n[1 + (n-1)C_{Mm}^2]$ which gives the lower bound above. Note that bounds on $\lambda_{\min}(\mathbf{C})$ have been derived in the framework of interpolation with radial basis functions, see Narcowich (1991), Ball (1992), Sun (1992) for lower bounds and Schaback (1994) for upper bounds. A maximin-distance design minimizes C_{Mm} and thus minimizes the upper bound above on $\phi(\xi)$.

5 Design for estimating covariance parameters

We now consider the case where the covariance $\mathbf{C}$ used for kriging (Sect. 4.1) depends upon unknown parameters ν that need to be estimated (by Maximum Likelihood) from the dataset $\mathbf{y}(\xi)$.

5.1 The Fisher information matrix

Under this assumption, a first step towards good prediction of the spatial random field may be the precise estimation of both sets of parameters β and ν. The information on them is contained in the so-called Fisher information matrix, which can be derived explicitly when the process $Z(\cdot)$ is Gaussian.

In this case the (un-normalized) information matrix for β and $\theta = (\sigma_Z^2, \nu^\top)^\top$ is block diagonal. Denoting $\mathbf{C}_\theta = \sigma_Z^2 \mathbf{C}_\nu$, we get

$$\mathbf{M}_{\beta,\theta}(\xi;\beta,\theta) = \begin{pmatrix} \mathbf{M}_\beta(\xi;\theta) & \mathbf{O} \\ \mathbf{O} & \mathbf{M}_\theta(\xi;\theta) \end{pmatrix}, \qquad (27)$$

where, for the model (16) with $\eta(\mathbf{x},\beta) = \mathbf{r}^\top(\mathbf{x})\beta$,

$$\mathbf{M}_\beta(\xi;\theta) = \frac{1}{\sigma_Z^2} \mathbf{R}^\top \mathbf{C}^{-1} \mathbf{R}$$

with $\mathbf{R} = [\mathbf{r}(\mathbf{x}_1), \ldots, \mathbf{r}(\mathbf{x}_n)]^\top$ and

$$\{\mathbf{M}_\theta(\xi;\theta)\}_{ij} = \frac{1}{2} \mathrm{tr}\left\{\mathbf{C}_\theta^{-1} \frac{\partial \mathbf{C}_\theta}{\partial \theta_i} \mathbf{C}_\theta^{-1} \frac{\partial \mathbf{C}_\theta}{\partial \theta_j}\right\}.$$

Since $\hat{Y}(\mathbf{x}|\xi)$ does not depend on σ_Z and σ_Z^2 only intervenes as a multiplicative factor in the MSPE, see Sect. 4.1, we are only interested in the precision of the estimation of β and ν. Note that

$$\mathbf{M}_\theta(\xi;\theta) = \begin{pmatrix} n/(2\sigma_Z^4) & \mathbf{z}_\nu^\top(\xi;\theta) \\ \mathbf{z}_\nu(\xi;\theta) & \mathbf{M}_\nu(\xi;\nu) \end{pmatrix}$$

with

$$\{\mathbf{z}_\nu(\xi;\theta)\}_i = \frac{1}{2\sigma_Z^2} \mathrm{tr}\left(\mathbf{C}_\nu^{-1} \frac{\partial \mathbf{C}_\nu}{\partial \nu_i}\right)$$

$$\{\mathbf{M}_\nu(\xi;\nu)\}_{ij} = \frac{1}{2} \mathrm{tr}\left\{\mathbf{C}_\nu^{-1} \frac{\partial \mathbf{C}_\nu}{\partial \nu_i} \mathbf{C}_\nu^{-1} \frac{\partial \mathbf{C}_\nu}{\partial \nu_j}\right\}.$$

Denote

$$\mathbf{M}_\theta^{-1}(\xi;\theta) = \begin{pmatrix} a(\xi;\theta) & \mathbf{b}_\nu^\top(\xi;\theta) \\ \mathbf{b}_\nu(\xi;\theta) & \mathbf{A}_\nu(\xi;\nu) \end{pmatrix}.$$

The block of $\mathbf{A}_\nu(\xi;\nu)$ then characterizes the precision of the estimation of ν (note that $\mathbf{A}_\nu(\xi;\nu) = [\mathbf{M}_\nu(\xi;\nu) - 2\sigma_Z^4 \mathbf{z}_\nu(\xi;\theta)\mathbf{z}_\nu^\top(\xi;\theta)/n]^{-1}$ does not depend on σ_Z). The matrix $\mathbf{A}_\nu(\xi;\nu)$ is often replaced by $\mathbf{M}_\nu^{-1}(\xi;\nu)$ and $\mathbf{M}_{\beta,\theta}(\xi;\beta,\theta)$ by

$$\mathbf{M}_{\beta,\nu}(\xi;\beta,\theta) = \begin{pmatrix} \mathbf{M}_\beta(\xi;\theta) & \mathbf{O} \\ \mathbf{O} & \mathbf{M}_\nu(\xi;\nu) \end{pmatrix},$$

which corresponds to the case when σ_Z is known. This can sometimes be justified from estimability considerations concerning the random-field parameters σ_Z and ν. Indeed, under the infill design framework (i.e., when the design space is compact) typically not all parameters are estimable, only some of them, or suitable functions of them, being micro-ergodic, see e.g. Stein (1999), Zhang and Zimmerman (2005). In that case, a reparametrization can be used, see e.g. Zhu and Zhang (2006), and one may sometimes set σ_Z to an arbitrary value. When both σ_Z and ν are estimable, there is usually no big difference between $\mathbf{A}_\nu(\xi;\nu)$ and $\mathbf{M}_\nu^{-1}(\xi;\nu)$.

Following traditional optimal design theory, see, e.g., Fedorov (1972), it is common to choose designs that maximize a scalar function of $\mathbf{M}_{\beta,\nu}(\xi;\beta,\theta)$, such as its determinant (D-optimality). Müller and Stehlík (2010) have suggested to maximize a compound criterion with weighing factor α,

$$\Phi_D[\xi|\alpha] = \left(\det[\mathbf{M}_\beta(\xi;\theta)]\right)^\alpha \left(\det[\mathbf{M}_\nu(\xi;\nu)]\right)^{1-\alpha}. \quad (28)$$

Some theoretical results for special situations showing that $\alpha \to 1$ leads to space-filling have been recently given in Kiseľák and Stehlík (2008), Zagoraiou and Antognini (2009) and Dette et al. (2008); Irvine et al. (2007) motivate the use of designs with clusters of points.

5.2 The modified kriging variance

G-optimal designs based on the (normalized) kriging variance (19) are space filling (see, e.g., van Groenigen 2000); however, they do not reflect the resulting additional uncertainty due to the estimation of the covariance parameters. We thus require an updated design criterion that takes that uncertainty into account. Even if this effect is asymptotically negligible, see Putter and Young (2001), its impact in finite samples may be decisive, see Müller et al. (2010).

Various proposals have been made to correct the kriging variance for the additional uncertainty due to the estimation of ν. One approach, based on Monte-Carlo sampling from the asymptotic distribution of the estimated parameters $\hat{\nu}^n$, is proposed in Nagy et al. (2007). Similarly, Sjöstedt-De-Luna and Young (2003) and den Hertog et al. (2006) have employed bootstrapping techniques for assessing the effect. Harville and Jeske (1992) use a first-order expansion of the kriging variance for $\hat{\nu}^n$ around its true value, see also Abt (1999) for more precise developments and Zimmerman and Cressie (1992) for a discussion and examples. This has the advantage that we can obtain an explicit correction term to augment the (normalized) kriging variance, which gives the approximation

$$\tilde{\rho}_\xi^2(\mathbf{x},\nu) = \rho_\xi^2(\mathbf{x},\nu)$$
$$+ \mathrm{tr}\left\{\mathbf{M}_\nu^{-1}(\xi;\nu)\frac{\partial\mathbf{v}_\nu^\top(\mathbf{x})}{\partial\nu}\mathbf{C}_\nu(\nu)\frac{\partial\mathbf{v}_\nu(\mathbf{x})}{\partial\nu^\top}\right\}, \quad (29)$$

with $\mathbf{v}_\nu(\mathbf{x})$ given by (18) (note that $\tilde{\rho}_\xi^2(\mathbf{x}_i,\nu) = 0$ for all i). Consequently, Zimmerman (2006) constructs designs by minimizing

$$\tilde{\phi}_G(\xi) = \max_{\mathbf{x}\in\mathcal{X}}\tilde{\rho}_\xi^2(\mathbf{x},\nu) \quad (30)$$

for some nominal ν, which he terms EK-(empirical kriging-)optimality (see also Zhu and Stein (2005) for a similar criterion). The objective here is to take the dual effect of the design into account (obtaining accurate predictions at

unsampled sites and improving the accuracy of the estimation of the covariance parameters, those two objectives being generally conflicting) through the formulation of a single criterion. One should notice that $\tilde{\rho}_\xi^2(\mathbf{x},\nu)$ may seriously overestimate the MSPE at $\mathbf{x}$ when the correlation is excessively weak. Indeed, for very weak correlation the BLUP (17) approximately equals $\mathbf{r}^\top(\mathbf{x})\hat{\beta}$ excepted in the neighborhood of the $\mathbf{x}_i$ due to the interpolating property $\hat{Y}(\mathbf{x}_i|\xi) = Y(\mathbf{x}_i)$ for all i; $\mathbf{v}(\mathbf{x})$ then shows rapid variations in the neighborhood of the $\mathbf{x}_i$ and $\|\partial\mathbf{v}(\mathbf{x})/\partial\nu\|$ may become very large. In that case, one may add a nugget effect to the model and replace (16) by $Y(\mathbf{x}) = \eta(\mathbf{x},\beta) + Z(\mathbf{x}) + \varepsilon(\mathbf{x})$ where the $\varepsilon(\mathbf{x}_i)$ are i.i.d. errors, also independent from the random process $Z(\mathbf{x})$, with zero mean and constant variance σ_ε^2. The BLUP then no longer interpolates the data which renders $\mathbf{v}(\mathbf{x})$ more stable; see e.g. Gramacy and Lee (2010) for other motivations concerning the introduction of a nugget effect.

The minimization of (30) is a difficult task, even for moderate d, due to the required maximizations of $\tilde{\rho}^2(\mathbf{x})$. Similarly to Sect. 4.2, the derivation of an upper bound on $\tilde{\rho}^2(\mathbf{x})$ could be used to form a simpler criterion. (Notice that in the case considered in Sect. 4.2 where $\eta(\mathbf{x},\beta) = \beta$ we have $\mathbf{v}_\nu(\mathbf{x}) = \mathbf{C}_\nu^{-1}\mathbf{P}_\nu\mathbf{c}_\nu(\mathbf{x}) + \mathbf{w}_\nu$ with $\mathbf{P}_\nu$ the projector $\mathbf{P}_\nu = \mathbf{I}_n - \mathbf{1}\mathbf{1}^\top\mathbf{C}_\nu^{-1}/(\mathbf{1}^\top\mathbf{C}_\nu^{-1}\mathbf{1})$ and $\mathbf{w}_\nu = \mathbf{C}_\nu^{-1}\mathbf{1}/(\mathbf{1}^\top\mathbf{C}_\nu^{-1}\mathbf{1})$ not depending on $\mathbf{x}$.) An alternative approach that also takes the effect of the unknown ν on predictions into account would be to place Maximum-Entropy-Sampling of Sect. 4.3 into a Bayesian framework, setting a prior distribution on β and ν. Also, the relationship between the criteria (28) and (30) is explored by Müller et al. (2010) who show that, although a complete equivalence can not be reached by a mere selection of α in general, respective efficiencies are usually quite high.

The strategy proposed in the next section tries to combine space-filling model-free design, to be used in a first stage, with estimation-oriented design, based on a model, to be used in a second stage. The objective is to reach good performance in terms of the modified kriging variance (29) but keep the computational burden as low as possible.

6 Combining space-filling and estimation designs

A rather straightforward method for combining estimation and prediction-based designs is suggested in Müller (2007). First a design consisting of a certain number of points n_0 is selected to maximize a criterion for the estimation of β, e.g. $\det[\mathbf{M}_\beta(\xi,\theta)]$; it is then optimally augmented by n_1 points for a criterion related to the estimation of ν, e.g. $\det[\mathbf{M}_\nu(\xi,\nu)]$, to yield a complete design with $n = n_0 + n_1$ points.

A similar idea can be exploited here and we suggest the following strategy.

Fig. 9 Contour plot for $\tilde{\phi}_G(\xi)$ on the 7-point minimax and maximin Lh-design of Fig. 2 (*left*) and with the central point shifted (*right*)
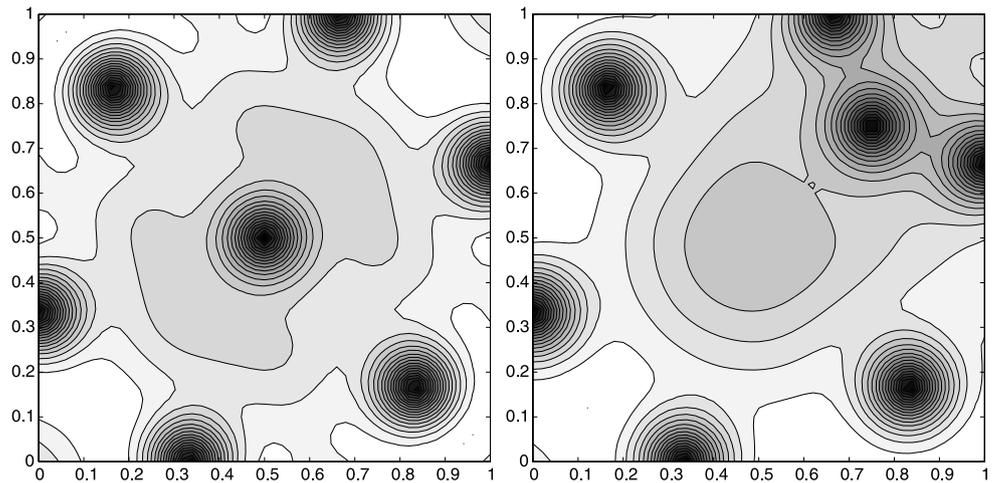
(1) Choose a space-filling, model-free, criterion $\phi_0(\cdot)$, e.g., $\phi_{mM}(\cdot)$, $\phi_{Mm}(\cdot)$ or a criterion from Sect. 3.

(2) Determine a n-point optimal design $\xi_{n,0}^*$ for $\phi_0(\cdot)$, compute $\tilde{\phi}_G(\xi_{n,0}^*)$. Set $n_1 = 1$.

(3) Determine a n_0-point optimal design $\xi_{n_0}^*$ for $\phi_0(\cdot)$, with $n_0 = n - n_1$; augment $\xi_{n_0}^*$ to a n-point design ξ_{n_0,n_1}^* by choosing n_1 points optimally for the criterion $\det[\mathbf{M}_\nu(\xi_{n_0,n_1}; \nu)]$, compute $\tilde{\phi}_G(\xi_{n_0,n_1}^*)$ (30).

(4) If $\tilde{\phi}_G(\xi_{n_0,n_1}^*) \geq \tilde{\phi}_G(\xi_{n_0+1,n_1-1}^*)$, select the design ξ_{n_0+1,n_1-1}^*, and stop; otherwise, increment n_1 by 1 and return to step 3).

Asymptotically, EK-optimal designs approach the typically space-filling G-optimal designs since the correcting term in (29) vanishes, see Putter and Young (2001). For n large and $\phi_0(\cdot)$ in good agreement with G-optimality (e.g., $\phi_0 = \phi_{mM}$) we can thus expect the value of n_1 in ξ_{n_0+1,n_1-1}^* selected by the strategy above to be relatively small.

Example 3 That step 3 of the above procedure makes sense can be demonstrated on a simple example. Take the 7 point Lh design from Fig. 2, which is simultaneously minimax and maximin optimal. Setting $\nu = 7$, we get $\tilde{\phi}_G(\xi) \simeq 1.913$, obtained for $\mathbf{x} \simeq (0.049, 0.951)$ or $(0.951, 0.049)$, see Fig. 9 (left), with $M_\nu(\xi; \nu) \simeq 2.41 \times 10^{-3}$ (and $A_\nu^{-1}(\xi; \nu) \simeq 2.40 \times 10^{-3}$). If we now shift the central point away from the center towards the upper right (or lower left) corner, say to the coordinate $(3/4, 3/4)$ as in the right panel of Fig. 9, the criterion is improved to a value of $\tilde{\phi}_G(\xi) \simeq 1.511$, attained for $\mathbf{x}$ at the opposite corner, and $M_\nu(\xi; \nu)$ is increased to 4.79×10^{-3} (and $A_\nu^{-1}(\xi; \nu)$ to 4.71×10^{-3}). This effect is enhanced as we move the central point closer to one of the non-central points and as we increase the value of ν and clearly shows the need to go beyond space-filling in this scenario.

7 Algorithms

Sequential and adaptive design In the sequential construction of a design, points can be introduced one-at-a-time (full sequential design) or by batches of given size $m > 1$. At each stage, all points introduced previously are fixed, which means in particular that the associated observations can be used to better select new points. The design is called adaptive when this information is used. This is especially useful in model-based design, when the criterion depends on some unknown parameters, for instance ν in the kriging variance (19) or (29). Suppose that $\phi(\cdot)$ is a criterion to be minimized. At stage k of a full-sequential design one constructs $\xi_{k+1} = (\xi_k, \mathbf{x}_{k+1})$, with $\xi_k = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ already determined, by choosing

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x} \in \mathfrak{X}} \phi[(\xi_k, \mathbf{x})]. \tag{31}$$

It should be noticed that such a construction is not always suitable. In particular, for G-optimality (21) a $\mathbf{x}_{k+1}$ chosen in this way will usually be in the close vicinity of a point already present in ξ_k, due to the fact that $\phi_G(\xi)$ only depends on a local characteristic (the value of $\rho^2(\mathbf{x})$ at its maximizer). The much simpler construction

$$\mathbf{x}_{k+1} = \arg\max_{\mathbf{x} \in \mathfrak{X}} \rho_{\xi_k}^2(\mathbf{x}) \tag{32}$$

is often used for that reason. Note that when the correlation tends to zero, this $\mathbf{x}_{k+1}$ tends to be as far as possible from the points already present, similarly to the greedy algorithm for maximin-optimal design for which $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x} \in \mathfrak{X}} \min_{\mathbf{x}_i \in \xi_k} \|\mathbf{x} - \mathbf{x}_i\|$. The design obtained by (32) will thus tend to be of the maximin-distance rather than minimax-distance type, and thus different from a G-optimal design (in particular, (32) tends to push points to the boundary of $\mathfrak{X}$). When $\phi(\xi) = \phi(\xi; \nu)$ depends on some unknown parameters ν, the construction (31) is easily made adaptive by using a forced-certainty-equivalence

adaptation rule (see, e.g., Pronzato 2008) that replaces at stage k the unknown ν by $\hat{\nu}^k$ estimated from ξ_k and the associated observations $\mathbf{y}(\xi_k)$. One then chooses $\mathbf{x}_{k+1} = \arg\min_{\mathbf{x} \in \mathfrak{X}} \phi[(\xi_k, \mathbf{x}); \hat{\nu}^k]$; the adaptive version of (32) is simply $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x} \in \mathfrak{X}} \rho_{\xi_k}^2(\mathbf{x}, \hat{\nu}^k)$.

Non-sequential design The direct minimization of a function $\phi(\cdot)$ with respect to $\xi = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is a rather formidable task even for moderate values of n and d when $\phi(\cdot)$ is not convex and local minimizers exist, which is always the case for the criteria considered here. Instead of performing a direct optimization with respect to $\xi \in \mathbb{R}^{nd}$ (or over a finite class in the case of Lh designs, see Sect. 2.2), most approaches combine heuristics with an exchange algorithm. The methods are abundant, ranging from genetic algorithms and tabu search (see e.g., Glover et al. 1995) to simulated annealing (Morris and Mitchell 1995). Some are more adapted to combinatorial search (and thus useful when working in the class of Lh designs, see van Dam et al. (2007) for a branch-and-bound algorithm for maximin Lh designs and Jin et al. (2005) for a stochastic evolutionary method). One may refer to Husslage et al. (2006) for a recent survey on optimization methods, including numerical constructions of Lh maximin designs (up to $d = 10$ and $n = 300$). Recent methods suggested for purely geometric problems, see, e.g., Cortés and Bullo (2009), could be transferred to more statistically-based space-filling criteria. Software can be obtained for instance from Royle and Nychka (1998), Walvoort et al. (2010), see also the packages DiceDesign (http://www.dice-consortium.fr/) by Franco, Dupuy and Roustant or lhs by Carnell (2009). We simply indicate below a prototype algorithm and mention its main ingredients.

One of the simplest, but not very efficient, algorithm is as follows: generate a random sequence of designs ξ_k, select the best one among them in terms of $\phi(\cdot)$. Note that one does not need to store all designs, only the best one found so far, say ξ_k^* after k designs have been generated, and its associated criterion value $\phi(\xi_k^*)$ have to be remembered. This procedure is often used for Lh designs, for instance to generate good designs in terms of minimax or maximin distance. Note that it may help to generate the random sequence $\{\xi_k\}$ according to a particular process, see Franco (2008), Franco et al. (2008) who use the Strauss point process, accounting for repulsion between points and thus favoring maximin-distance designs. Also note that each ξ_k generated can be used as starting point for a local search algorithm (using generalized gradients when $\phi(\cdot)$ is not differentiable). A more sophisticated version of this idea is as follows.

Let ξ_k denote the design at iteration k and $\phi(\cdot)$ be a criterion to be minimized. Consider an algorithm for which at iteration k one exchanges one point $\mathbf{x}_i$ of ξ_k with a new one

$\mathbf{x}^*$, see Fedorov (1972), Mitchell (1974) for exchange algorithms originally proposed for optimal design in a parametric setting. Three elements must then be defined: (i) how to select $\mathbf{x}_i$? (ii) how to construct $\mathbf{x}^*$? (iii) what to do if the substitution of $\mathbf{x}^*$ for $\mathbf{x}_i$ in ξ_k, possibly followed by a local search, does not improve the value of $\phi(\cdot)$?

Typically, the answer to (i) involves some randomness, possibly combined with some heuristics (for instance, for maximin-optimal design, it seems reasonable to select $\mathbf{x}_i$ among the pairs of points at respective distance $\phi_{Mm}(\xi)$). For (ii), the choice of $\mathbf{x}^*$ can be purely random (e.g., a random walk in $\mathfrak{X}$ originated at $\mathbf{x}_i$), or based on a deterministic construction, or a mixture of both. Finally, for (iii), a simulated-annealing method is appropriate: denote ξ_k^+ the design obtained by substituting $\mathbf{x}^*$ for $\mathbf{x}_i$ in ξ_k, set $\xi_{k+1} = \xi_k^+$ when $\phi(\xi_k^+) < \phi(\xi_k)$ (improvement) but also accept this move with probability $P_k = \exp\{-[\phi(\xi_k^+) - \phi(\xi_k)/T_k]\}$ when $\phi(\xi_k^+) > \phi(\xi_k)$. Here T_k denotes a 'temperature' that should decrease with k. Note that keeping trace of the best design encountered, ξ_k^* at step k, facilitates the proof of convergence to an optimal design: indeed, $\limsup_{k \to \infty} \phi(\xi_k)$ may not converge to its minimal value $\phi^* = \min_\xi \phi(\xi)$ (and there might not be a reversible measure for the transition kernel from ξ_k to ξ_{k+1}), but it is easy to prove that $\lim_{k \to \infty} \phi(\xi_k^*) = \phi^*$ when there is enough randomness in (i) and (ii). One may refer to (Morris and Mitchell 1995) for indications on how to choose the initial temperature T_0 and make it decrease with k. See e.g. Schilling (1992) and Angelis et al. (2001) for the use of a similar algorithm in a related framework.

8 Concluding remarks

The design of computer experiments has been a rapidly growing field in the last few years, with special emphasis put on the construction of criteria quantifying how spread out a design is: geometric measures, related to sphere covering and sphere packing (minimax and maximin distance designs), statistical measures of uniformity (discrepancy and more recently entropy). Much work remains to be done to determine which approaches are more suitable for computer experiments and to construct efficient algorithms tailored to specific criteria (some being easier to optimize than the others).

The paper has also drawn attention on the importance of going beyond space filling. The estimation of parameters in a model-based approach calls for designs that are not uniformly spread out. A simple procedure has been proposed (Sect. 6), but here also much remains to be done. We conclude the presentation by mentioning some recent results that might help reconciliating the space-filling and non-space-filling points of view. When a model selection strategy

using a spatial information criterion is employed, Hoeting et al. (2006) note that clustered designs perform the best. When different levels of accuracy are required, nested space-filling designs have shown to be useful (cf. Qian et al. 2009; Rennen et al. 2010) and non-space fillingness arises naturally in some sequentially designed experiments (see Gramacy and Lee 2009). Picheny et al. (2010) induce it by focusing their attention towards particular target regions.

More generally, Dette and Pepelyshev (2010) observe that the kriging variance for a uniform design on $[0, 1]$ is (in general) larger near the boundaries than near the center, so that a G-optimal design tends to put more points near the boundaries (note, however, that this is not true for the exponential covariance function, see Figs. 6 and 7, due to the particular Markovian structure of the Ornstein-Uhlenbeck process on the real line). Similarly, an optimal experiment for polynomial regression (D-optimal for instance) puts more points near the boundary of the domain as the degree of the polynomial increases; in the limit, the design points are distributed with the arc-sine distribution. These observations speak for space-filling designs that do not fill the space uniformly, but rather put more points near its boundaries. Since such designs will place some points close together, they may help the estimation of covariance parameters in kriging, and thus perhaps kill two birds with one stone.

Appendix A: Computation of the minimax distance criterion via Delaunay tessellation

Let $\mathbf{x}_*$ be a point of $\mathfrak{X}$ satisfying $\min_i \|\mathbf{x} - \mathbf{x}_i\| = \phi(\xi_{mM}^*)$. If $\mathbf{x}_*$ is in the interior of $\mathfrak{X}$, it is at equal distance of $d+1$ points of $\mathfrak{X}$ that form a non-degenerate simplex; it is therefore the center of a circumscribed sphere to a simplex in the Delaunay tessellation of the points of ξ (which we call a Delaunay sphere).

Suppose now that $\mathbf{x}_*$ lies on the boundary of $\mathfrak{X}$. It then belongs to some $(d-q)$-dimensional face $\mathfrak{H}_q$ of $\mathfrak{X}$, $1 \leq q \leq d$ (a 0-dimensional face being a vertex of $\mathfrak{X}$, a 1-dimensional face an edge, etc., a $(d-1)$-dimensional face is a $(d-1)$-dimensional hyperplane). Also, it must be at equal distance D^* from $m = d+1-q$ points of ξ and no other point from ξ can be closer. Consider now the symmetric points of those m points with respect to the q different $(d-1)$-dimensional faces of $\mathfrak{X}$ that define $\mathfrak{H}_q$; we obtain in this way $m(q+1)$ points that are all at distance D^* from $\mathbf{x}_*$. No other point from ξ, or any symmetric of a point of ξ with respect to a $(d-1)$-face of $\mathfrak{X}$, is at distance from $\mathbf{x}_*$ less than D^*. Since $m(q+1) = (d+1-q)(q+1) \geq d+1$ (with equality when

$q = d$, that is when $\mathbf{x}_*$ if one of the 2^d vertices of $\mathfrak{X}$), $\mathbf{x}_*$ is always at the center of a Delaunay sphere obtained from the tessellation of the points in ξ and their $2d+1$ symmetric points with respect to the $(d-1)$-dimensional faces of $\mathfrak{X}$. (The tessellation obtained is not unique in general due to the enforced symmetry among the set of points constructed, but this is not an issue.)

The center $\mathbf{z}_*$ and radius R of the circumscribed sphere to a simplex defined by $d+1$ vectors $\mathbf{z}_1, \ldots, \mathbf{z}_{d+1}$ of $\mathbb{R}^d$ is easily computed as follows. Since $(\mathbf{z}_* - \mathbf{z}_i)^\top (\mathbf{z}_* - \mathbf{z}_i) = R^2$ for all i, $2\mathbf{z}_i^\top \mathbf{z}_* - \mathbf{z}_i^\top \mathbf{z}_i = \mathbf{z}_*^\top \mathbf{z}_* - R^2$ is a constant. Denote this constant by γ and the vector formed by the squared norm of the $\mathbf{z}_i$ by $\mathbf{w}$, so that $\{\mathbf{w}\}_i = \mathbf{z}_i^\top \mathbf{z}_i$ for all i. We thus have, in matrix form, $[2\mathbf{Z}^\top - \mathbf{1}] (\mathbf{z}_*^\top \ \gamma)^\top = \mathbf{w}$ with $\mathbf{Z}$ the $d \times (d+1)$ matrix $(\mathbf{z}_1, \ldots, \mathbf{z}_{d+1})$ and $\mathbf{1}$ the $(d+1)$-dimensional vector of ones. Note that the singularity of the matrix $[2\mathbf{Z}^\top - \mathbf{1}]$ would imply that all $\mathbf{z}_i$ lie in a $(d-1)$-dimensional hyperplane; the matrix is thus of full rank when the $\mathbf{z}_i$ form a non-degenerate simplex. The values of $\mathbf{z}_*$ and γ are directly obtained from the equation above and R is then given by $\sqrt{\mathbf{z}_*^\top \mathbf{z}_* - \gamma}$.

Appendix B: Regularization through L_q norms

Consider a design criterion $\phi(\cdot)$ which can be written as the minimum of a set of criteria, $\phi(\xi) = \min_i \phi_i(\xi)$.

Suppose that this set of criteria $\phi_i(\cdot)$ is finite (extensions to infinite sets and generalized classes of criteria indexed by a continuous parameters are possible but useless here for our purpose), so that $i \in \{1, \ldots, M\}$ for some finite M. The min function makes $\phi(\cdot)$ non smooth even when the $\phi_i(\cdot)$ are. Different regularization methods can be used in that case to construct a smooth approximation of $\phi(\cdot)$.

Suppose that $\phi(\xi) > 0$ and define

$$\underline{\phi}_{[q]}(\xi) = \left[\sum_{i=1}^M \phi_i^{-q}(\xi) \right]^{-1/q}. \tag{33}$$

From a property of L_q norms, $\underline{\phi}_{[q_2]}(\xi) \leq \underline{\phi}_{[q_1]}(\xi)$ for any $q_1 > q_2 > 0$, so that $\underline{\phi}_{[q]}(\xi)$ with $q > 0$ forms a lower bound on $\phi(\xi)$ which tends to $\phi(\xi)$ as $q \to \infty$. $\underline{\phi}_{[q]}(\xi)$ is also an increasing function of q for $q < 0$ but is not defined at $q = 0$ (with $\lim_{q \to 0^-} \underline{\phi}_{[q]}(\xi) = +\infty$ and $\lim_{q \to 0^+} \underline{\phi}_{[q]}(\xi) = 0$). Note that $\underline{\phi}_{[-1]}(\xi) = \sum_{i=1}^M \phi_i(\xi) \geq \phi(\xi)$.

Consider now the criterion

$$\overline{\phi}_{[q]}(\xi) = \left[\sum_{i=1}^M \mu_i \, \phi_i^{-q}(\xi) \right]^{-1/q}, \tag{34}$$

where $\mu_i > 0$ for all i and $\sum_{i=1}^M \mu_i = 1$ (the μ_i define a probability measure on the index set $\{1, \ldots, M\}$). Again, for

any ξ such that $\phi(\xi) > 0$, $\overline{\phi}_{[q]}(\xi) \to \phi(\xi)$ as q tends to ∞. The computation of the derivative $\partial \overline{\phi}_{[q]}(\xi)/\partial q$ gives

$$\frac{\partial \overline{\phi}_{[q]}(\xi)}{\partial q} = \frac{\overline{\phi}_{[q]}(\xi)}{q^2 \sum_{i=1}^{M} \mu_i \phi_i^{-q}(\xi)}$$
$$\times \left\{ \left[\sum_{i=1}^{M} \mu_i \phi_i^{-q}(\xi) \right] \log \left[\sum_{i=1}^{M} \mu_i \phi_i^{-q}(\xi) \right] \right.$$
$$\left. - \sum_{i=1}^{M} \mu_i \phi_i^{-q}(\xi) \log[\phi_i^{-q}(\xi)] \right\} \le 0 \text{ for any } q,$$

where the inequality follows from Jensen's inequality (the function $x \to x \log x$ being convex). The inequality is strict when the $\phi_i(\xi)$ take at least two different values and $\overline{\phi}_{[q]}(\xi)$ then decreases monotonically to $\phi(\xi)$ as $q \to \infty$. Similarly to the case of $\underline{\phi}_{[q]}(\cdot)$, we have $\overline{\phi}_{[-1]}(\xi) = \sum_{i=1}^{M} \mu_i \phi_i(\xi) \ge \phi(\xi)$. Moreover,

$$\lim_{q \to 0} \overline{\phi}_{[q]}(\xi) = \exp \left\{ \sum_{i=1}^{M} \mu_i \log[\phi_i(\xi)] \right\},$$

which, by continuity can be defined as being $\overline{\phi}_{[0]}(\xi)$.

Define $\underline{\mu} = \min\{\mu_i, \ i = 1, \ldots, M\}$. We have, for any ξ such that $\phi(\xi) > 0$ and any $q > 0$,

$$\underline{\phi}_{[q]}(\xi) \le \phi(\xi) \le \overline{\phi}_{[q]}(\xi) \le \underline{\mu}^{-1/q} \underline{\phi}_{[q]}(\xi), \tag{35}$$

so that

$$0 \le \phi(\xi) - \underline{\phi}_{[q]}(\xi) \le (\underline{\mu}^{-1/q} - 1)\phi(\xi^*),$$
$$0 \le \overline{\phi}_{[q]}(\xi) - \phi(\xi) \le (\underline{\mu}^{-1/q} - 1)\phi(\xi^*),$$

where ξ^* is optimal for $\phi(\cdot)$ and $\underline{\mu}^{-1/q}$ tends to 1 as $q \to \infty$. The convergence of $\underline{\phi}_{[q]}(\cdot)$ and $\overline{\overline{\phi}}_{[q]}(\cdot)$ to $\phi(\cdot)$, respectively from below and from above, is thus uniform over any set of designs such that $\phi(\xi)$ is bounded away from zero. Moreover, we can directly deduce from (35) that the ϕ-efficiency of optimal designs optimal for $\overline{\phi}_{[q]}(\cdot)$ or $\overline{\phi}_{[q]}(\cdot)$ is at least $\underline{\mu}^{1/q}$. Indeed, let ξ^*, $\underline{\xi}^*_{[q]}$ and $\overline{\xi}^*_{[q]}$ respectively denote an optimal design for ϕ, $\underline{\phi}_{[q]}(\cdot)$ and $\overline{\phi}_{[q]}(\cdot)$; (35) implies that

$$\phi(\underline{\xi}^*_{[q]}) \ge \underline{\phi}_{[q]}(\underline{\xi}^*_{[q]}) \ge \underline{\phi}_{[q]}(\xi^*) \ge \underline{\mu}^{1/q} \phi(\xi^*)$$
$$\phi(\overline{\xi}^*_{[q]}) \ge \underline{\mu}^{1/q} \overline{\phi}_{[q]}(\overline{\xi}^*_{[q]}) \ge \underline{\mu}^{1/q} \overline{\phi}_{[q]}(\xi^*) \ge \underline{\mu}^{1/q} \phi(\xi^*).$$

The best efficiency bounds are obtained when $\underline{\mu}$ is maximal, that is, when μ is the uniform measure and $\underline{\mu} = 1/M$. In that case, $\overline{\phi}_{[q]}(\xi) = M^{1/q} \underline{\phi}_{[q]}(\xi)$ and $\overline{\phi}_{[0]}(\xi) = [\prod_{i=1}^{M} \phi_i(\xi)]^{1/M}$.

An obvious generalization of the regularization by L_q norm is as follows. Let $\psi(\cdot)$ be a strictly increasing

function and $\psi^{\leftarrow}(\cdot)$ denote its inverse. Then, $\phi(\xi) = \psi^{\leftarrow}\{\min_i \psi[\phi_i(\xi)]\}$, and, applying the L_q regularizations above to the min function, we can define

$$\overline{\phi}_{[q,\psi]}(\xi) = \psi^{\leftarrow}\left\{ \left[\sum_{i=1}^{M} \mu_i \{\psi[\phi_i(\xi)]\}^{-q} \right]^{-1/q} \right\}, \tag{36}$$

$$\underline{\phi}_{[q,\psi]}(\xi) = \psi^{\leftarrow}\left\{ \left[\sum_{i=1}^{M} \{\psi[\phi_i(\xi)]\}^{-q} \right]^{-1/q} \right\}. \tag{37}$$

A case of special interest is $\psi(t) = \exp(t)$, which gives

$$\underline{\phi}_{[q,\exp]}(\xi) = -\frac{1}{q} \log \left\{ \sum_{i=1}^{M} \exp[-q\phi_i(\xi)] \right\},$$

and is appealing in situations where one may have $\phi_i(\xi) \le 0$, see Li and Fang (1997).

Appendix C: Derivation of bounds on the kriging variance

β is known   We have $\rho_0^2(\mathbf{x}) \le 1 - \max_i \{\mathbf{c}(\mathbf{x})\}_i^2 / \lambda_{\max}(\mathbf{C})$, with $\lambda_{\max}(\mathbf{C})$ the maximum eigenvalue of $\mathbf{C}$. Since $C(\cdot)$ is non-increasing, $\{\mathbf{C}\}_{ij} \le C_{Mm}$ for all $i \ne j$. We denote by $\mathfrak{C}$ the set of matrices $\mathbf{C}$ satisfying $0 \le \{\mathbf{C}\}_{ij} = \{\mathbf{C}\}_{ji} \le C_{Mm}$ for $i \ne j$. A classical inequality on matrix norms gives $\lambda_{\max}(\mathbf{C}) = \|\mathbf{C}\|_2 \le (\|\mathbf{C}\|_1 \|\mathbf{C}\|_\infty)^{1/2}$, where $\|\mathbf{C}\|_1 = \max_j \sum_i |\{\mathbf{C}\}_{ij}| = \|\mathbf{C}\|_\infty$. Therefore, any $\mathbf{C} \in \mathfrak{C}$ satisfies $\lambda_{\max}(\mathbf{C}) \le 1 + (n-1)C_{Mm}$ and

$$\rho_0^2(\mathbf{x}) \le \bar{\rho}_0^2(\mathbf{x}) = 1 - \frac{\bar{c}(\mathbf{x})^2}{1 + (n-1)C_{Mm}}, \tag{38}$$

where $\bar{c}(\mathbf{x}) = \max_i \{\mathbf{c}(\mathbf{x})\}_i$. Since $\min_i \|\mathbf{x} - \mathbf{x}_i\| \le \phi_{mM}$ for all $\mathbf{x} \in \mathfrak{X}$, we have $\bar{c}(\mathbf{x}) \ge C_{mM}$ for all $\mathbf{x}$ and

$$\max_{\mathbf{x} \in \mathfrak{X}} \rho_0^2(\mathbf{x}) \le \bar{\rho}_0^2 = 1 - \frac{C_{mM}^2}{1 + (n-1)C_{Mm}}. \tag{39}$$

Note that the bound (39) will become worse as n increases since the bound $1 + (n-1)C_{Mm}$ on $\lambda_{\max}(\mathbf{C})$ becomes more and more pessimistic. Also, (38) can be tight only for those $\mathbf{x}$ such that $\mathbf{c}(\mathbf{x})$ corresponds to the direction of an eigenvector associated with $\lambda_{\max}(\mathbf{C})$.

β is unknown   We need to bound the second term in $\rho^2(\mathbf{x})$ given by (20). Our first step is to enclose the feasible set for $\mathbf{c}(\mathbf{x})$ into a set $\mathcal{C}$ of simple description. Notice that $\|\mathbf{x} - \mathbf{x}_i\| < \phi_{Mm}/2$ for some i implies that $\|\mathbf{x} - \mathbf{x}_j\| > \phi_{Mm}/2$ for all $i \ne j$. Therefore,

$$\mathcal{C} \subset [0, 1]^n \setminus \mathfrak{P}(\phi_{Mm})$$
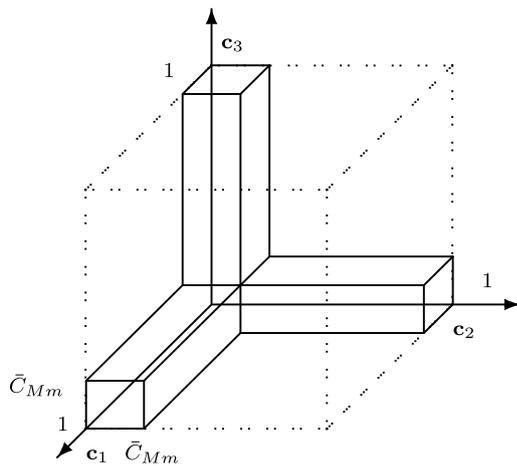
Fig. 10 The set $[0, 1]^3 \setminus \mathfrak{P}(\phi_{Mm})$

with $\mathfrak{P}(\phi_{Mm}) = \{\mathbf{c} \in [0, 1]^n : \exists i \neq j \text{ with } \{\mathbf{c}\}_i > \bar{C}_{Mm} \text{ and } \{\mathbf{c}\}_j > \bar{C}_{Mm}\}$, where $\bar{C}_{Mm} = C(\phi_{Mm}/2)$, see Fig. 10 for an illustration when $n = 3$. Notice that $\phi_{mM} > \phi_{Mm}/2$ implies that $\bar{C}_{Mm} = C(\phi_{Mm}/2) \geq C_{mM}$ (with also $\bar{C}_{Mm} \geq C_{Mm}$). Next, since $\bar{c}(\mathbf{x}) = \max_i\{\mathbf{c}(\mathbf{x})\}_i \geq C_{mM}$, we have

$$\mathcal{C} \subset [0, 1]^n \setminus [0, \bar{c}(\mathbf{x})]^n \subset [0, 1]^n \setminus [0, C_{mM}]^n. \tag{40}$$

Consider $T(\mathbf{x}) = \mathbf{c}^\top(\mathbf{x})\mathbf{C}^{-1}\mathbf{1}$. Notice that $\{\mathbf{c}(\mathbf{x})\}_i = 1$ for some i implies that $\mathbf{x} = \mathbf{x}_i$, and that $\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}_i) = \mathbf{e}_i$, the i-th basis vector, so that $T(\mathbf{x}_i) = 1$. Also, if $\{\mathbf{c}(\mathbf{x})\}_i = 1$ for some i, then $\|\mathbf{x} - \mathbf{x}_j\| \geq \phi_{Mm}$ and thus $\{\mathbf{c}(\mathbf{x})\}_j \leq C_{Mm}$ for all $j \neq i$. When the correlation is weak enough, $0 < \{\mathbf{C}^{-1}\mathbf{1}\}_i \leq 1$ for all i (which is true for some processes whatever the importance of the correlation, it is the case for instance for the one-dimensional Ornstein-Uhlenbeck process). This gives $T(\mathbf{x}) \leq \mathbf{c}^\top(\mathbf{x})\mathbf{1} \leq 1 + (n - 1)\bar{C}_{Mm}$. Also, (40) implies that the minimum of $T(\mathbf{x})$ is larger than $\bar{c}(\mathbf{x})\mathbf{e}_1^\top\mathbf{C}_*^{-1}\mathbf{1}$ with

$$\mathbf{C}_* = \begin{pmatrix} 1 & C_{Mm}\mathbf{1}_{n-1}^\top \\ C_{Mm}\mathbf{1}_{n-1} & \mathbf{I}_{n-1} \end{pmatrix} \tag{41}$$

where $\mathbf{I}_{n-1}$ and $\mathbf{1}_{n-1}$ respectively denote the $(n - 1)$-dimensional identity matrix and vector of ones, which gives

$$T(\mathbf{x}) \geq \bar{c}(\mathbf{x}) \frac{1 - (n - 1)C_{Mm}}{1 - (n - 1)C_{Mm}^2}.$$

Since $\mathbf{1}^\top\mathbf{C}^{-1}\mathbf{1} \geq n/\lambda_{\max}(\mathbf{C}) \geq n/[1 + (n - 1)C_{Mm}]$, we finally obtain

$$\rho^2(\mathbf{x}) \leq \bar{\rho}^2(\mathbf{x}) = \bar{\rho}_0^2(\mathbf{x}) + \frac{1 + (n - 1)C_{Mm}}{n}R^2(\mathbf{x}) \tag{42}$$

with $\bar{\rho}_0^2(\mathbf{x})$ given by (38) and $R^2(\mathbf{x}) = \max[R_a^2(\mathbf{x}), R_b^2]$ where

$$R_a^2(\mathbf{x}) = \left[1 - \bar{c}(\mathbf{x})\frac{1 - (n - 1)C_{Mm}}{1 - (n - 1)C_{Mm}^2}\right]^2$$

and $R_b^2 = (n - 1)^2\bar{C}_{Mm}^2$. It should be noticed that the upper bound R_b^2 is very pessimistic. In fact, $\max_{\mathbf{x}} T(\mathbf{x})$ seldom exceeds one (it may do so marginally when $C(t)$ is concave at $t = 0$), see for instance Joseph (2006), and for that reason it is sufficient to use $R^2(\mathbf{x}) = R_a^2(\mathbf{x})$.

References

Abt, M.: Estimating the prediction mean squared error in Gaussian stochastic processes with exponential correlation structure. Scand. J. Stat. **26**(4), 563–578 (1999)

Angelis, L., Senta, E.B., Moyssiadis, C.: Optimal exact experimental designs with correlated errors through a simulated annealing algorithm. Comput. Stat. Data Anal. **37**(3), 275–296 (2001)

Ash, R.: Information Theory. Wiley, New York (1965) (republished by Dover, New York, 1990)

Audze, P., Eglais, V.: New approach for planning out experiments. Probl. Dyn. Strengths **35**, 104–107 (1977)

Ball, K.: Eigenvalues of Euclidean distance matrices. J. Approx. Theory **68**, 74–82 (1992)

Bates, R.A., Buck, R.J., Riccomagno, E., Wynn, H.P.: Experimental design and observation for large systems. J. R. Stat. Soc. B **58**(1), 77–94 (1996)

Beirlant, J., Dudewicz, E., Györfi, L., van der Meulen, E.: Nonparametric entropy estimation; an overview. Int. J. Math. Stat. Sci. **6**(1), 17–39 (1997)

Bellhouse, D.R., Herzberg, A.M.: Equally spaced design points in polynomial regression: a comparison of systematic sampling methods with the optimal design of experiments. Can. J. Stat. **12**(2), 77–90 (1984)

Bettinger, R., Duchêne, P., Pronzato, L., Thierry, E.: Design of experiments for response diversity. In: Proc. 6th International Conference on Inverse Problems in Engineering (ICIPE). J. Phys. Conf. Ser., Dourdan (Paris) (2008)

Bettinger, R., Duchêne, P., Pronzato, L.: A sequential design method for the inversion of an unknown system. In: Proc. 15th IFAC Symposium on System Identification, Saint-Malo, France, pp. 1298–1303 (2009)

Bischoff, W., Miller, F.: Optimal designs which are efficient for lack of fit tests. Ann. Stat. **34**(4), 2015–2025 (2006)

Boissonnat, J.D., Yvinec, M.: Algorithmic Geometry. Cambridge University Press, Cambridge (1998)

Bursztyn, D., Steinberg, D.: Comparison of designs for computer experiments. J. Stat. Plan. Inference **136**(3), 1103–1119 (2006)

Carnell, R.: lhs: latin hypercube samples. R package version 0.5 (2009)

Chen, V.C.P., Tsui, K.L., Barton, R.R., Meckesheimer, M.: A review on design, modeling and applications of computer experiments. AIIE Trans. **38**(4), 273–291 (2006)

Cignoni, P., Montani, C., Scopigno, R.: DeWall: a fast divide and conquer Delaunay triangulation algorithm in E^d. Comput. Aided Des. **30**(5), 333–341 (1998)

Cortés, J., Bullo, F.: Nonsmooth coordination and geometric optimization via distributed dynamical systems. SIAM Rev. **51**(1), 163–189 (2009)

Cressie, N.: Statistics for Spatial Data. Wiley Series in Probability and Statistics. Wiley-Interscience, New York (1993). Rev. Sub. Ed.

den Hertog, D., Kleijnen, J.P.C., Siem, A.Y.D.: The correct kriging variance estimated by bootstrapping. J. Oper. Res. Soc. **57**(4), 400–409 (2006)

Dette, H., Pepelyshev, A.: Generalized Latin hypercube design for computer experiments. Technometrics **52**(4), 421–429 (2010)

Dette, H., Kunert, J., Pepelyshev, A.: Exact optimal designs for weighted least squares analysis with correlated errors. Stat. Sin. **18**(1), 135–154 (2008)

Fang, K.T.: The uniform design: application of number theoretic methods in experimental design. Acta Math. Appl. Sin. **3**, 363–372 (1980)

Fang, K.T., Li, R.: Uniform design for computer experiments and its optimal properties. Int. J. Mater. Prod. Technol. **25**(1), 198–210 (2006)

Fang, K.T., Wang, Y.: Number-Theoretic Methods in Statistics, 1st edn. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall/CRC, London/Boca Raton (1993)

Fang, K.T., Lin, D.K.J., Winker, P., Zhang, Y.: Uniform design: Theory and application. Technometrics **42**(3), 237–248 (2000)

Fang, K.T., Li, R., Sudjianto, A.: Design and Modeling for Computer Experiments. Chapman & Hall/CRC, London/Boca Raton (2005)

Fedorov, V.: Theory of Optimal Experiments. Academic Press, New York (1972)

Fedorov, V.V., Hackl, P.: Model-Oriented Design of Experiments. Lecture Notes in Statistics, vol. 125. Springer, Berlin (1997)

Franco, J.: Planification d'expériences numériques en phase exploratoire pour la simulation de phénomènes complexes. Ph.D. thesis, École Nationale Supérieure des Mines de Saint Etienne (2008)

Franco, J., Bay, X., Corre, B., Dupuy, D.: Planification d'expériences numériques à partir du processus ponctuel de Strauss. Preprint, Département 3MI, École Nationale Supérieure des Mines de Saint-Etienne. http://hal.archives-ouvertes.fr/hal-00260701/fr/ (2008)

Franco, J., Vasseur, O., Corre, B., Sergent, M.: Minimum Spanning Tree: a new approach to assess the quality of the design of computer experiments. Chemom. Intell. Lab. Syst. **97**, 164–169 (2009)

Gensane, T.: Dense packings of equal spheres in a cube. Electron. J. Comb. 11 (2004)

Glover, F., Kelly, J., Laguna, M.: Genetic algorithms and tabu search: hybrids for optimization. Comput. Oper. Res. **22**(1), 111–134 (1995)

Gramacy, R., Lee, H.: Cases for the nugget in modeling computer experiments. Tech. rep., http://arxiv.org/abs/1007.4580 (2010)

Gramacy, R.B., Lee, H.K.: Adaptive design and analysis of supercomputer experiments. Technometrics **51**(2), 130–144 (2009)

Griffith, D.: Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization. Springer, Berlin (2003)

Hall, P., Morton, S.: On the estimation of entropy. Ann. Inst. Stat. Math. **45**(1), 69–88 (1993)

Harville, D.A., Jeske, D.R.: Mean squared error of estimation or prediction under a general linear model. J. Am. Stat. Assoc. **87**(419), 724–731 (1992)

Havrda, M., Charvát, F.: Quantification method of classification processes: concept of structural α-entropy. Kybernetika **3**, 30–35 (1967)

Herzberg, A.M., Huda, S.: A comparison of equally spaced designs with different correlation structures in one and more dimensions. Can. J. Stat. **9**(2), 203–208 (1981)

Hoeting, J.A., Davis, R.A., Merton, A.A., Thompson, S.E.: Model selection for geostatistical models. Ecol. Appl. **16**(1), 87–98 (2006)

Husslage, B., Rennen, G., van Dam, E., den Hertog, D.: Space-filling Latin hypercube designs for computer experiments. Discussion paper 2006-18, Tilburg University, Center for Economic Research (2006)

Iooss, B., Boussouf, L., Feuillard, V., Marrel, A.: Numerical studies of the metamodel fitting and validation processes. International Journal on Advances in Systems and Measurements **3**(1–2), 11–21 (2010)

Irvine, K., Gitelman, A., Hoeting, J.: Spatial designs and properties of spatial correlation: effects on covariance estimation. J. Agric. Biol. Environ. Stat. **12**(4), 450–469 (2007)

Jin, R., Chen, W., Sudjianto, A.: An efficient algorithm for constructing optimal design of computer experiments. J. Stat. Plan. Inference **134**(1), 268–287 (2005)

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

Johnson, R.T., Montgomery, D.C., Jones, B., Fowler, J.W.: Comparing designs for computer simulation experiments. In: WSC '08: Proceedings of the 40th Conference on Winter Simulation, pp. 463–470 (2008)

Joseph, V.: Limit kriging. Technometrics **48**(4), 458–466 (2006)

Jourdan, A., Franco, J.: Optimal Latin hypercube designs for the Kullback-Leibler criterion. AStA Adv. Stat. Anal. **94**, 341–351 (2010)

Kiefer, J., Wolfowitz, J.: The equivalence of two extremum problems. Can. J. Math. **12**, 363–366 (1960)

Kiseľák, J., Stehlík, M.: Equidistant and d-optimal designs for parameters of Ornstein–Uhlenbeck process. Stat. Probab. Lett. **78**(12), 1388–1396 (2008)

Kleijnen, J.P.C.: Design and Analysis of Simulation Experiments. Springer, New York (2009)

Koehler, J., Owen, A.: Computer experiments. In: Ghosh, S., Rao, C.R. (eds.) Handbook of Statistics: Design and Analysis of Experiments, vol. 13, pp. 261–308. North-Holland, Amsterdam (1996)

Kozachenko, L., Leonenko, N.: On statistical estimation of entropy of random vector. Probl. Inf. Transm. **23**(2), 95–101 (1987) (translated from Problemy Peredachi Informatsii, in Russian, vol. 23, No. 2, pp. 9–16, 1987)

Leary, S., Bhaskar, A., Keane, A.: Optimal orthogonal-array-based Latin hypercubes. J. Appl. Stat. **30**(5), 585–598 (2003)

Leonenko, N., Pronzato, L., Savani, V.: A class of Rényi information estimators for multidimensional densities. Ann. Stat. **36**(5), 2153–2182 (2008) (correction in Ann. Stat. **38**(6), 3837–3838, 2010)

Li, X.S., Fang, S.C.: On the entropic regularization method for solving min-max problems with applications. Math. Methods Oper. Res. **46**, 119–130 (1997)

McKay, M., Beckman, R., Conover, W.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**(2), 239–245 (1979)

Melissen, H.: Packing and covering with circles. Ph.D. thesis, University of Utrecht (1997)

Mitchell, T.: An algorithm for the construction of "*D*-optimal" experimental designs. Technometrics **16**, 203–210 (1974)

Morris, M., Mitchell, T.: Exploratory designs for computational experiments. J. Stat. Plan. Inference **43**, 381–402 (1995)

Müller, W.G.: Collecting Spatial Data: Optimum Design of Experiments for Random Fields, 3rd edn. Springer, Heidelberg (2007)

Müller, W.G., Stehlík, M.: Compound optimal spatial designs. Environmetrics **21**(3–4), 354–364 (2010)

Müller, W.G., Pronzato, L., Waldl, H.: Relations between designs for prediction and estimation in random fields: an illustrative case. In: E. Porcu, J.M. Montero, M. Schlather (eds.) Advances and Challenges in Space-Time Modelling of natural Events. Springer Lecture Notes in Statistics (2011, to appear)

Nagy, B., Loeppky, J.L., Welch, W.J.: Fast Bayesian inference for Gaussian process models. Tech. rep., The University of British Columbia, Department of Statistics (2007)

Narcowich, F.: Norms of inverses and condition numbers for matrices associated with scattered data. J. Approx. Theory **64**, 69–94 (1991)

Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods (CBMS-NSF Regional Conference Series in Applied Mathematics). SIAM, Philadelphia (1992)

Okabe, A., Books, B., Sugihama, K.: Spatial Tessellations. Concepts and Applications of Voronoi Diagrams. Wiley, New York (1992)

Oler, N.: A finite packing problem. Can. Math. Bull. **4**, 153–155 (1961)

Pebesma, E.J., Heuvelink, G.B.M.: Latin hypercube sampling of Gaussian random fields. Technometrics **41**(4), 303–312 (1999)

Penrose, M., Yukich, J.: Laws of large numbers and nearest neighbor distances. In: Wells, M., Sengupta, A. (eds.) Advances in Directional and Linear Statistics. A Festschrift for Sreenivasa Rao Jammalamadaka, Chap. 13, pp. 189–199 (2011). arXiv:0911.0331v1

Petelet, M., Iooss, B., Asserin, O., Loredo, A.: Latin hypercube sampling with inequality constraints. AStA Adv. Stat. Anal. **94**, 325–339 (2010)

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R.T., Kim, N.H.: Adaptive designs of experiments for accurate approximation of a target region. J. Mech. Des. **132**(7), 071,008 (2010)

Pistone, G., Vicario, G.: Comparing and generating Latin hypercube designs in Kriging models. AStA Adv. Stat. Anal. **94**, 353–366 (2010)

Pronzato, L.: Optimal experimental design and some related control problems. Automatica **44**(2), 303–325 (2008)

Putter, H., Young, A.: On the effect of covariance function estimation on the accuracy of kriging predictors. Bernoulli **7**(3), 421–438 (2001)

Qian, P.Z.G., Ai, M., Wu, C.F.J.: Construction of nested space-filling designs. Ann. Stat. **37**(6A), 3616–3643 (2009)

Redmond, C., Yukich, J.: Asymptotics for Euclidian functionals with power-weighted edges. Stoch. Process. Appl. **61**, 289–304 (1996)

Rennen, G., Husslage, B., van Dam, E., den Hertog, D.: Nested maximin Latin hypercube designs. Struct Multidisc Optiml **41**, 371–395 (2010)

Rényi, A.: On measures of entropy and information. In: Proc. 4th Berkeley Symp. on Math. Statist. and Prob., pp. 547–561 (1961)

Riccomagno, E., Schwabe, R., Wynn, H.P.: Lattice-based D-optimum design for Fourier regression. Ann. Stat. **25**(6), 2313–2327 (1997)

Royle, J., Nychka, D.: An algorithm for the construction of spatial coverage designs with implementation in SPLUS. Comput. Geosci. **24**(5), 479–488 (1998)

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

Santner, T., Williams, B., Notz, W.: The Design and Analysis of Computer Experiments. Springer, Heidelberg (2003)

Schaback, R.: Lower bounds for norms of inverses of interpolation matrices for radial basis functions. J. Approx. Theory **79**, 287–306 (1994)

Schilling, M.F.: Spatial designs when the observations are correlated. Commun. Stat., Simul. Comput. **21**(1), 243–267 (1992)

Scott, D.: Multivariate Density Estimation. Wiley, New York (1992)

Shewry, M., Wynn, H.: Maximum entropy sampling. Appl. Stat. **14**, 165–170 (1987)

Sjöstedt-De-Luna, S., Young, A.: The bootstrap and kriging prediction intervals. Scand. J. Stat. **30**(1), 175–192 (2003)

Stein, M.: Interpolation of Spatial Data: Some Theory for Kriging. Springer, Heidelberg (1999)

Stinstra, E., den Hertog, D., Stehouwer, P., Vestjens, A.: Constrained maximin designs for computer experiments. Technometrics **45**(4), 340–346 (2003)

Sun, X.: Norm estimates for inverses of Euclidean distance matrices. J. Approx. Theory **70**, 339–347 (1992)

Tang, B.: Orthogonal array-based latin hypercubes. J. Am. Stat. Assoc. **88**(424), 1392–1397 (1993)

Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. **52**(1/2), 479–487 (1988)

van Dam, E.: Two-dimensional minimax Latin hypercube designs. Discrete Appl. Math. **156**(18), 3483–3493 (2007)

van Dam, E., Hussage, B., den Hertog, D., Melissen, H.: Maximin Latin hypercube designs in two dimensions. Oper. Res. **55**(1), 158–169 (2007)

van Dam, E., Rennen, G., Husslage, B.: Bounds for maximin Latin hypercube designs. Oper. Res. **57**(3), 595–608 (2009)

van Groenigen, J.: The influence of variogram parameters on optimal sampling schemes for mapping by kriging. Geoderma **97**(3–4), 223–236 (2000)

Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J.: An *R* package for spatial coverage sampling and random sampling from compact geographical strata by *k*-means. Comput. Geosci. **36**, 1261–1267 (2010)

Wendland, H.: Scattered Data Approximation. Cambridge University Press, Cambridge (2005)

Wolkowicz, H., Styan, G.: Bounds for eigenvalues using traces. Linear Algebra Appl. **29**, 471–506 (1980)

Wynn, H.: Maximum entropy sampling and general equivalence theory. In: Di Bucchianico, A., Läuter, H., Wynn, H. (eds.) mODa'7—Advances in Model–Oriented Design and Analysis, Proceedings of the 7th Int. Workshop, Heeze, Netherlands, pp. 211–218. Physica Verlag, Heidelberg (2004)

Yfantis, E., Flatman, G., Behar, J.: Efficiency of kriging estimation for square, triangular, and hexagonal grids. Math. Geol. **19**(3), 183–205 (1987)

Yukich, J.: Probability Theory of Classical Euclidean Optimization Problems. Springer, Berlin (1998)

Zagoraiou, M., Antognini, A.B.: Optimal designs for parameter estimation of the Ornstein-Uhlenbeck process. Appl. Stoch. Models Bus. Ind. **25**(5), 583–600 (2009)

Zhang, H., Zimmerman, D.: Towards reconciling two asymptotic frameworks in spatial statistics. Biometrika **92**(4), 921–936 (2005)

Zhu, Z., Stein, M.: Spatial sampling design for parameter estimation of the covariance function. J. Stat. Plan. Inference **134**(2), 583–603 (2005)

Zhu, Z., Zhang, H.: Spatial sampling design under the infill asymptotic framework. Environmetrics **17**(4), 323–337 (2006)

Zimmerman, D.L.: Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics **17**(6), 635–652 (2006)

Zimmerman, D., Cressie, N.: Mean squared prediction error in the spatial linear model with estimated covariance parameters. Ann. Inst. Stat. Math. **44**(1), 27–43 (1992)