Journal de la Société Française de Statistique
Vol. 158 No. 1 (2017)

Minimax and maximin space-filling designs: some properties and methods for construction

Titre: Plans d'expériences à remplissage d'espace minimax et maximin : quelques propriétés et méthodes de construction

Luc Pronzato[1]

Abstract: A few properties of minimax and maximin optimal designs in a compact subset of $\mathbb{R}^d$ are presented, and connections with other space-filling constructions are indicated. Several methods are given for the evaluation of the minimax-distance (or dispersion) criterion for a given n-point design. Various optimisation methods are proposed and their limitations, in particular in terms of dimension d, are indicated. A large majority of the results presented are not new, but their collection in a single document containing a respectable bibliography will hopefully be useful to the reader.

Résumé : Cet article présente quelques propriétés de plans d'expériences minimax et maximin optimaux dans un compact de $\mathbb{R}^d$, ainsi que leur relations avec d'autres plans "à remplissage d'espace". Plusieurs méthodes sont indiquées pour l'évaluation du critère de distance minimax (ou dispersion) pour un plan donné à n points. Diverses méthodes d'optimisation sont proposées et leurs limitations sont indiquées, en particulier en terme de dimension d. La grande majorité des résultats présentés ne sont pas nouveaux, mais nous espérons que le lecteur trouvera utile de les voir réunis dans un seul document accompagné d'une bibliographie conséquente.

Keywords: computer experiments, space-filling design, minimax-optimal design, maximin-optimal design, sphere covering, sphere packing
Mots-clés : expériences numériques, plans d'expériences à remplissage d'espace, plans minimax optimaux, plans maximin optimaux, sphères de recouvrement, empilement de sphères
AMS 2000 subject classifications: 62K99, 05B40

1. Introduction

In the common parametric framework with independent additive errors, optimal designs for linear or nonlinear models tend to concentrate observation sites at a few locations only. In particular, optimal design theory indicates that, for a model depending on p parameters, design measures maximising an information criterion based on the Fisher information matrix need no more than $p(p+1)/2+1$ support points, see, e.g., Fedorov (1972), Silvey (1980), Pukelsheim (1993), Atkinson et al. (2007), Pronzato and Pázman (2013), Fedorov and Leonov (2014). The situation is much different for nonparametric regression with correlated errors, where the absence of prior knowledge on the mean and variance functions yields optimal designs that are uniformly distributed, see Biedermann and Dette (2001). Also, the growing tendency in industry to replace physical experiments by simulations based on sophisticated mathematical models has generated a strong demand for designs dedicated to the specific situation where repeated observations at

[1] Université Côte d'Azur, CNRS, Laboratoire I3S, Sophia Antipolis
E-mail: Luc.Pronzato@cnrs.fr

 ISSN: 2102-6238

the same site (i.e., of several simulations with the same inputs) give the same response. This corresponds to the type of problems considered here: we want to predict the behavior of an unknown function f: $\mathbf{x} \in \mathscr{X} \mapsto f(\mathbf{x}) \in \mathbb{R}$, with $\mathscr{X}$ a compact subset of $\mathbb{R}^d$, $d \geq 2$ but reasonably small ($d \lesssim 10$, say); observations correspond to evaluations of f at a collection of sites $\mathbf{x}_i$, $i = 1, \ldots, n$, which form the n-point design $X_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$; they may possibly be corrupted by random errors, correlated or not. The absence of knowledge on f is an incitement to observe everywhere in $\mathscr{X}$, that is, to spread X_n over $\mathscr{X}$. The objective of the paper is to present techniques for the construction of designs having this property.

After observations $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)$ have been collected, various methods can be used to predict the value $f(\mathbf{x})$ at any given $\mathbf{x} \in \mathscr{X}$, see for instance Sacks et al. (1989), Stein (1999), Santner et al. (2003) for kriging and computer experiments, and one may think of defining a design criterion especially adapted to the particular prediction method that will be used. For instance, it might be based on Maximum-Entropy Sampling (MES), see Shewry and Wynn (1987), or on the Maximum (respectively, Integrated) Mean-Squared prediction Error (MMSE, respectively IMSE) over $\mathscr{X}$, see (Santner et al., 2003, Chap. 6), Gauthier and Pronzato (2014, 2016). Here we shall focus our attention on "universal" (model-free) design criteria, not depending on the prediction method to be used and only based on geometrical properties of $\mathscr{X}$ and X_n. The relationship with model-based design constructions will be briefly considered in Section 2.5.2.

In all the paper $\|.\|$ denotes the l^2-norm, and for any set $\mathscr{A} \subseteq \mathscr{X}$ and any point $\mathbf{x} \in \mathscr{X}$ we write $d(\mathbf{x}, \mathscr{A}) = \min_{\mathbf{z} \in \mathscr{A}} \|\mathbf{x} - \mathbf{z}\|$; $\mathrm{vol}(\mathscr{A}) = \int_{\mathscr{A}} \mathrm{d}\mathbf{x}$ denotes the volume of $\mathscr{A}$ and $|\mathscr{A}|$ its cardinality when $\mathscr{A}$ is finite. The letter $\mathscr{X}$ will always denote a compact subset of $\mathbb{R}^d$ with strictly positive d-dimensional Lebesgue measure and equal to the closure of its interior. The set $\mathscr{X}$ will generally be convex. Typical examples correspond to the d-dimensional unit (hyper-)cube $[0,1]^d$ or the closed unit ball $\mathscr{B}(\mathbf{0}, 1)$, with $\mathbf{0}$ denoting the (here d-dimensional) null vector; $\mathbf{1}$ will denote a vector with all components equal to 1 and $\mathbf{e}_j$ will denote the j-th basis vector; $\mathbf{I}_p$ is the p-dimensional identity matrix.

The paper is organised around two different notions of dispersion that are fundamental in most space-filling strategies. The first one is based on distances between design points and corresponds to the maximin-distance criterion of Johnson et al. (1990),

$$\Phi_{Mm}(X_n) = \min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\| .$$

We shall denote $\Phi^*_{Mm,n} = \Phi^*_{Mm,n}(\mathscr{X}) = \max_{X_n \in \mathscr{X}^n} \Phi_{Mm}(X_n)$ and $X^*_{Mm,n}$ a maximin-optimal design such that $\Phi_{Mm}(X^*_{Mm,n}) = \Phi^*_{Mm,n}$. The index of $X^*_{Mm,n}$ is defined as the number of pairs of points being at distance $\Phi^*_{Mm,n}$, maximin-optimal designs of lowest index being preferable, see Johnson et al. (1990). See Figure 1-left for an example.

The second notion relies on the Hausdorff distance between X_n and $\mathscr{X}$,

$$\begin{aligned} \Phi_{mM}(X_n) &= d_H(X_n, \mathscr{X}) = \max \left\{ \max_{i=1,\ldots,n} d(\mathbf{x}_i, \mathscr{X}), \max_{\mathbf{x} \in \mathscr{X}} d(\mathbf{x}, X_n) \right\} = \max_{\mathbf{x} \in \mathscr{X}} d(\mathbf{x}, X_n) \\ &= \max_{\mathbf{x} \in \mathscr{X}} \min_{i=1,\ldots,n} \|\mathbf{x} - \mathbf{x}_i\| , \end{aligned} \tag{1}$$

since $d(\mathbf{x}_i, \mathscr{X}) = 0$ when $\mathbf{x}_i \in \mathscr{X}$. Φ_{mM} corresponds to the minimax-distance criterion of Johnson et al. (1990) and $\Phi_{mM}(X_n)$ is also called the dispersion of X_n, see (Niederreiter, 1992, Chap. 6).

 ISSN: 2102-6238

We shall denote $\Phi^*_{mM,n} = \Phi^*_{mM,n}(\mathscr{X}) = \min_{X_n \in \mathscr{X}^n} \Phi_{mM}(X_n)$ and $X^*_{mM,n}$ a minimax-optimal design such that $\Phi_{mM}(X^*_{mM,n}) = \Phi^*_{mM,n}$. Let $\mathscr{X}^*(X_n)$ denote the set of most distant points $\mathbf{x} \in \mathscr{X}$, i.e., such that $d(\mathbf{x}, X_n) = \Phi^*_{mM,n}$. The index of $X^*_{mM,n}$ is defined as $\min_{x \in \mathscr{X}^*(X_n)} N(\mathbf{x})$, with $N(\mathbf{x}) = \left|\{\mathbf{x}_i \in X_n : \|\mathbf{x} - \mathbf{x}_i\| = \Phi^*_{mM,n}\}\right|$, and minimax-optimal designs of highest index are preferable. An example is given in Figure 1-right.

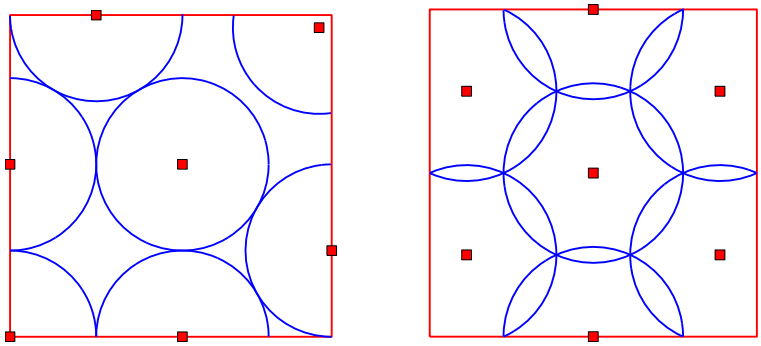

FIGURE 1. *Maximin (left) and minimax (right) optimal designs in a square for $n = 7$. The circles have radius $\Phi^*_{Mm,7}/2$ on the left and $\Phi^*_{mM,7}$ on the right; the maximin design on the left has index 8; the minimax optimal design $X^*_{mM,7}$ on the right has index 1, with $|\mathscr{X}^*(X^*_{mM,7})| = 16$.*

Clearly, maximin-optimal design will tend to push design points to the boundary of $\mathscr{X}$ whereas minimax-optimal design will do that to a much lesser extent, see Section 4.3.1 for an exploitation of this property. For a given n, these different behaviours become more striking as the dimension d increases. We shall only consider designs of given size n and not *sequences of nested designs* (one may refer to Niederreiter, 1992, Chap. 6, for the construction of low-dispersion sequences; see also Section 2.5.1 for discrepancy and its link with dispersion). One exception is the algorithm of Section 4.1 that generates design sequences with 50% minimax and maximin efficiency.

2. Justification and properties

2.1. *Motivation for minimising Φ_{mM}*

In the absence of precise knowledge on f, choosing a collection X_n of sites $\mathbf{x}_i \in \mathscr{X}$ that minimise $\Phi_{mM}(X_n)$ is intuitively appealing for the prediction of the value of f at unsampled locations $\mathbf{x} \in \mathscr{X}$ based on error-free evaluations $f(\mathbf{x}_i)$. Indeed, for any $\mathbf{z} \in \mathscr{X}$ and $X_n \in \mathscr{X}^n$, denote by $\mathbf{x}_{i^*(\mathbf{z})}$ (any of) the nearest neighbour(s) of $\mathbf{z}$ in X_n; that is, $\mathbf{x}_{i^*(\mathbf{z})} \in \operatorname{Arg\,min}_{i=1,\ldots,n} \|\mathbf{z} - \mathbf{x}_i\|$. If f is Lipschitz continuous, with $|f(\mathbf{z}) - f(\mathbf{x})| \le L_0 \|\mathbf{z} - \mathbf{x}\|$ for all $\mathbf{x}$ and $\mathbf{z} \in \mathscr{X}$, then $\max_{\mathbf{z} \in \mathscr{X}} |f(\mathbf{z}) - f(\mathbf{x}_{i^*(\mathbf{z})})| \le L_0 \Phi_{mM}(X_n)$. If f is Lipschitz differentiable, with $\|\nabla_f(\mathbf{z}) - \nabla_f(\mathbf{x})\| \le L_1 \|\mathbf{z} - \mathbf{x}\|$ for all $\mathbf{x}$ and $\mathbf{z} \in \mathscr{X}$, then $\max_{\mathbf{z} \in \mathscr{X}} |f(\mathbf{z}) - f(\mathbf{x}_{i^*(\mathbf{z})}) - \nabla_f^\top(\mathbf{x}_{i^*(\mathbf{z})})(\mathbf{z} - \mathbf{x}_{i^*(\mathbf{z})})| \le L_1 [\Phi_{mM}(X_n)]^2$. These obvious constructions are rather naive; they indicate, however, that the major difficulty for approximating an unknown f, including the famous curse of dimensionality, can in fact be formulated as a design problem: good designs with small Φ_{mM} value yield good approximations, even with constructions as simple as above. One may refer to Sukharev (1992, Chap. 3) for more precise developments.

 ISSN: 2102-6238

Suppose now that f belong to a RKHS $\mathscr{H}$ with kernel $K(\cdot,\cdot)$, see, e.g., Berlinet and Thomas-Agnan (2004). Let $\hat{\eta}_n$ denote the linear predictor of f based on the $f(\mathbf{x}_i)$, $i=1,\ldots,n$, such that $\hat{\eta}_n(\mathbf{x}_i)=f(\mathbf{x}_i)$ for all i ($\hat{\eta}_n$ interpolates f on X_n) and having minimum norm $\|\hat{\eta}_n\|_{\mathscr{H}}$. This best-linear interpolator is the orthogonal projection of f on $\mathrm{span}\{K_{\mathbf{x}_1},\ldots,K_{\mathbf{x}_n}\}$, where $K_{\mathbf{x}}:\mathbf{z}\in\mathscr{X}\mapsto K(\mathbf{z},\mathbf{x})$. It satisfies, for all $\mathbf{x}\in\mathscr{X}$, $|f(\mathbf{x})-\hat{\eta}_n(\mathbf{x})|\le\|f\|_{\mathscr{H}}\,\rho_n(\mathbf{x})$, where the first term $\|f\|_{\mathscr{H}}$ does not depend on X_n and the second one $\rho_n(\mathbf{x})$ does not depend on f, see, e.g., Auffray et al. (2012). In the framework of random-field modeling and kriging, $\rho_n^2(\mathbf{x})$ corresponds to the kriging variance at $\mathbf{x}$. There, the unknown function is considered as the realisation of a random field $Z_{\mathbf{x}}$ on $\mathscr{X}$, with zero mean and covariance $\mathsf{E}\{Z_{\mathbf{x}}Z_{\mathbf{x}'}\}=\sigma^2K(\mathbf{x},\mathbf{x}')$. The best linear unbiased predictor of $Z_{\mathbf{x}}$ based on the observations $\mathbf{z}_n=(Z_{\mathbf{x}_1},\ldots,Z_{\mathbf{x}_n})^\top$ is given by $\hat{\eta}_n(\mathbf{x})=\mathbf{k}_n^\top(\mathbf{x})\mathbf{K}_n^{-1}\mathbf{z}_n$ and satisfies $\mathsf{E}\{[Z_{\mathbf{x}}-\hat{\eta}_n(\mathbf{x})]^2\}=\sigma^2\rho_n^2(\mathbf{x})=\sigma^2\,[K(\mathbf{x},\mathbf{x})-\mathbf{k}_n^\top\mathbf{K}_n^{-1}\mathbf{k}_n(\mathbf{x})]$, with $\mathbf{k}_n(\mathbf{x})=(K_{\mathbf{x}_1}(\mathbf{x}),\ldots,K_{\mathbf{x}_n}(\mathbf{x}))^\top$ and $\{\mathbf{K}_n\}_{ij}=K(\mathbf{x}_i,\mathbf{x}_j)$, $i,j=1,\ldots,n$. (The fact that $\rho_n^2(\mathbf{x})$ does not depend on $\mathbf{z}_n$ has the important consequence that MES, or the optimisation of any design criterion based on ρ_n, see Section 2.5.2, cannot be improved if performed sequentially, by adapting the choice of the k-th design point to all current observations; see Vazquez and Bect (2011) for a thorough discussion.) Schaback (1995) shows that for many isotropic kernels (such that $K(\mathbf{x},\mathbf{z})$ depends on $\|\mathbf{x}-\mathbf{z}\|$), $\rho_n^2(\mathbf{x})$ satisfies $\sup_{\mathbf{x}\in\mathscr{X}}\rho_n(\mathbf{x})\le S[\Phi_{mM}(X_n)]$, with S an increasing function depending on K, hence, again, the interest of using designs with a small Φ_{mM} value.

2.2. *Relations between Φ_{mM} and Φ_{Mm}, and bounds on their optimum values*

Relations between Φ_{mM} and Φ_{Mm}. As noticed in (Gonzalez, 1985), any $(n+1)$-point design in $\mathscr{X}$ ($n\ge1$) satisfies

$$\frac{1}{2}\,\Phi_{Mm}(X_{n+1})\le\Phi^*_{mM,n}\,. \tag{2}$$

Indeed, from the pigeonhole principle, one of the n balls $\mathscr{B}(\mathbf{z}_i,\Phi^*_{mM,n})$, with $\mathbf{z}_i\in X^*_{mM,n}$, must contain two points $\mathbf{x}_i,\mathbf{x}_j$ of X_{n+1}, implying that $\Phi_{Mm}(X_{n+1})\le\|\mathbf{x}_i-\mathbf{x}_j\|\le2\Phi^*_{mM,n}$.

Consider now a design X_n with n distinct points $\mathbf{x}_i$. Any pair of (closed) balls $\mathscr{B}(\mathbf{x}_i,\Phi_{Mm}(X_n)/2)$ and $\mathscr{B}(\mathbf{x}_j,\Phi_{Mm}(X_n)/2)$, $i\ne j$, do not intersect, and since a convex set $\mathscr{X}$ cannot be covered by two or more non-overlapping balls all having their centres in $\mathscr{X}$, we obtain that any design X_n in a convex set $\mathscr{X}$, with $n\ge2$, satisfies $\Phi_{mM}(X_n)>\Phi_{Mm}(X_n)/2$. On the other hand,

$$\Phi^*_{mM,n}\le\Phi^*_{Mm,n} \tag{3}$$

for any $\mathscr{X}$ and any n. The proof is by contradiction, see Auffray et al. (2012), and is reproduced hereafter. Suppose that $\Phi^*_{mM,n}>\Phi^*_{Mm,n}$ and consider a maximin-optimal design $X^*_{Mm,n}$ with lowest index. The property $\Phi_{mM}(X^*_{Mm,n})\ge\Phi^*_{mM,n}>\Phi^*_{Mm,n}$ implies the existence of a $\mathbf{x}^*$ in $\mathscr{X}$ such that $d(\mathbf{x}^*,X^*_{Mm,n})>\Phi^*_{Mm,n}$, and by substituting $\mathbf{x}^*$ for any point of $X^*_{Mm,n}$ in a pair of points at respective distance $\Phi^*_{Mm,n}$, we would obtain a design X'_n with lower index than $X^*_{Mm,n}$ or such that $\Phi_{Mm}(X'_n)>\Phi^*_{Mm,n}$. One may notice that we have in fact proved that $\Phi_{mM}(X^*_{Mm,n})\le\Phi_{Mm}(X^*_{Mm,n})$ when $X^*_{Mm,n}$ is a maximin-optimal design with lowest index. Minimising Φ_{mM} is more difficult than maximising Φ_{Mm}, see Sections 2.4, 3.2 and 4, and this inequality explains why one is often satisfied with a maximin-optimal design.

 ISSN: 2102-6238

Lower bound. For any n and any $X_n \in \mathscr{X}^n$, the n balls $\mathscr{B}(\mathbf{x}_i, \Phi_{mM}(X_n))$ cover $\mathscr{X}$, and the determination of a minimax-optimal design $X^*_{mM,n}$ corresponds to a *sphere covering* problem. As a consequence, denoting $V_d = \mathsf{vol}[\mathscr{B}(0,1)] = \pi^{d/2}/\Gamma(d/2+1)$ the volume of the d-dimensional unit ball $\mathscr{B}(\mathbf{0},1)$, we get $nV_d\,(\Phi^*_{mM})^d \geq \mathsf{vol}(\mathscr{X})$, which yields the lower bound

$$\Phi^*_{Mm,n} \geq \Phi^*_{mM,n} \geq \underline{R}^*_n = \left(\frac{\mathsf{vol}(\mathscr{X})}{nV_d}\right)^{1/d}. \tag{4}$$

Upper bound. By construction, the n balls $\mathscr{B}(\mathbf{x}_i, \Phi^*_{Mm,n}/2)$, $n \geq 2$, do not overlap (more precisely, their intersection has zero volume) and have their centres in $\mathscr{X}$, so that

$$\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, \Phi^*_{Mm,n}/2) \subset \{\mathscr{X} \oplus \mathscr{B}(\mathbf{0}, \Phi^*_{Mm,n}/2)\} \subset \{\mathscr{X} \oplus [-\Phi^*_{Mm,n}/2, \Phi^*_{Mm,n}/2]^d)\}, \tag{5}$$

where $\oplus$ denotes Minkowski sum (here, a dilation of $\mathscr{X}$). When $\mathscr{X}$ *is convex* and has a simple enough shape, this induces a close relationship between maximin-optimal design and *sphere packing*. To investigate this relationship we shall need to expand and renormalise $\mathscr{X}$, and it is simpler for that purpose to assume, without any loss of generality, that $\mathscr{X}$ has been first normalised, so that the largest ball included in $\mathscr{X}$ is the unit ball $\mathscr{B}(\mathbf{0},1)$.

For any $r \geq 0$, denote $\mathscr{X}_r^- = [1/(1+r)]\,[\mathscr{X} \oplus \mathscr{B}(\mathbf{0},r)] = \{(\mathbf{x}+\mathbf{z})/(1+r) : \mathbf{x} \in \mathscr{X}, \|\mathbf{z}\| \leq r\}$. Any $\mathbf{y} \in \mathscr{X}_r^-$ can be written as $\alpha\mathbf{x} + (1-\alpha)\mathbf{z}$, with $\alpha = 1/(1+r)$, $\mathbf{x} \in \mathscr{X}$ and $\mathbf{z} \in \mathscr{B}(\mathbf{0},1) \subset \mathscr{X}$. Therefore, $\mathscr{X}_r^- \subset \mathscr{X}$ when $\mathscr{X}$ is convex. Consider an n-point maximin-optimal design $X^*_{Mm,n}$, with maximin value $\Phi^*_{Mm,n}$; take $r = \Phi^*_{Mm,n}/2$. The n balls $\mathscr{B}(\mathbf{x}_i, r)$ do not overlap and are contained in $\mathscr{X} \oplus \mathscr{B}(\mathbf{0},r)$, the non-overlapping balls $\mathscr{B}(\mathbf{x}_i/(1+r), r/(1+r))$ are thus contained in $\mathscr{X}$, which implies that

$$r/(1+r) = \Phi^*_{Mm,n}/(2+\Phi^*_{Mm,n}) \leq r^*_{SP,n}, \tag{6}$$

with $r^*_{SP,n}$ the maximum radius of n identical non-overlapping spheres packed in $\mathscr{X}$.

Consider now the set $\mathscr{X}_r^+ = \{\mathbf{x} : \forall \mathbf{z} \in \mathscr{B}(\mathbf{0},r),\ (1-r)\mathbf{x} + \mathbf{z} \in \mathscr{X}\}$, $r \in [0,1)$. Since $\mathscr{B}(\mathbf{0},1) \subset \mathscr{X}$ and $\mathscr{X}$ is assumed to be convex, $\mathscr{X} \subset \mathscr{X}_r^+$. On the other hand, when the erosion of $\mathscr{X}$ by $\mathscr{B}(\mathbf{0},r)$ satisfies

$$\mathscr{X} \ominus \mathscr{B}(\mathbf{0},r) = \{\mathbf{x} : \forall \mathbf{z} \in \mathscr{B}(\mathbf{0},r),\ \mathbf{x}+\mathbf{z} \in \mathscr{X}\} \text{ is homothetic to } \mathscr{X}, \tag{7}$$

with balls and regular polytopes of $\mathbb{R}^d$ as typical examples, then $\mathscr{X}_r^+ = \mathscr{X}$. In that case, any n non-overlapping spheres packed in $\mathscr{X}$ with radius $r^*_{SP,n}$, $n \geq 2$, are such that their centres $\mathbf{c}_i$ belong to $\mathscr{X} \ominus \mathscr{B}(\mathbf{0}, r^*_{SP,n})$, and therefore $\mathbf{c}_i/(1-r^*_{SP,n}) \in \mathscr{X}$ for all i. This implies that $2r^*_{SP,n}/(1-r^*_{SP,n}) \leq \min_{i\neq j}\|\mathbf{c}_i - \mathbf{c}_j\|/(1-r^*_{SP,n}) \leq \Phi^*_{Mm,n}$. Together with (6), this gives $\Phi^*_{Mm,n}/(2+\Phi^*_{Mm,n}) = r^*_{SP,n}$ and shows the equivalence between maximin-optimal design and sphere packing when $\mathscr{X}$ is a convex set satisfying (7); see Figure 2 for an illustration.

When the shape of $\mathscr{X}$ is simple enough, (5) and the equivalence with sphere packing yield upper bounds on $\mathsf{vol}\left[\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, \Phi^*_{Mm,n}/2)\right] = nV_d(\Phi^*_{Mm,n}/2)^d$, i.e., upper bounds on $\Phi^*_{Mm,n}$. In particular, two special cases provide a simple bound. When $\mathscr{X}$ is the unit ball $\mathscr{B}(\mathbf{0},1)$, we obtain $\Phi^*_{Mm,n} \leq 2/(n^{1/d}-1)$. When $\mathscr{X}$ is the hypercube $[0,1]^d$ and $n \leq n_* = \left\lceil [2(1+\sqrt{d})]^d/(V_d d^{d/2}) \right\rceil$,

 ISSN: 2102-6238

we have the trivial bound $\Phi^*_{Mm,n} \leq \overline{R}^*_n = \sqrt{d}$ (consider $X^*_{Mm,2}$ formed by two opposite vertices of $\mathscr{X}$). For $n > n_*$ (and therefore $(nV_d)^{1/d} > 2$), we get

$$\Phi^*_{mM,n} \leq \Phi^*_{Mm,n} \leq \overline{R}^*_n = \frac{2}{(nV_d)^{1/d} - 2}\,. \tag{8}$$

(When $n \leq \left\lceil [2(2+\sqrt{d})]^d/(V_d d^{d/2}) \right\rceil$, we also have the trivial bound $\Phi^*_{mM,n} \leq \sqrt{d}/2$, obtained from $X^*_{mM,1} = \{(1/2,\ldots,1/2)^\top\}$.)

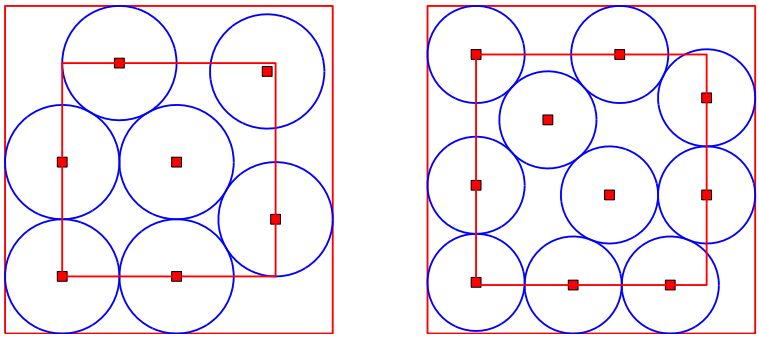

FIGURE 2. *Equivalence between maximin-optimal design and sphere packing when $\mathscr{X}$ is the square $[0,1]^2$, with $n = 7$ (left, the same design is presented in Figure 1-left) and $n = 10$ (right); (4) gives $\underline{R}^*_7 \simeq 0.21324$ and $\underline{R}^*_{10} \simeq 0.17841$; $n_* = 4$ for $d = 2$ and (8) gives $\overline{R}^*_7 \simeq 0.74364$ and $\overline{R}^*_{10} \simeq 0.55479$.*

Packing density, covering, and lattices. Any given design X_n provides an upper bound $\Phi_{mM}(X_n)$ on $\Phi^*_{mM,n}$ and a lower bound $\Phi_{Mm}(X_n)$ on $\Phi^*_{Mm,n}$. In particular, regular arrangements called lattices, form good candidates for obtaining accurate bounds.

When $\mathscr{X}$ can tile $\mathbb{R}^d$, the hypercube being a typical example, results on packing density can be used to improve (8): n non-overlapping balls with radius r contained in $\mathscr{X}$ cannot occupy a volume larger than $\delta_d V$, with $V = \mathrm{vol}(\mathscr{X})$ and δ_d the optimum (i.e., greatest) sphere-packing density in $\mathbb{R}^d$. Therefore, $r < [\delta_d V/(nV_d)]^{1/d}$. When $\mathscr{X}$ is the hypercube $[-1,1]^d$, so that $\mathscr{B}(\mathbf{0},1)$ is the largest ball included in $\mathscr{X}$, then $V = 2^d$ and $\Phi^*_{Mm,n} < 2r/(1-r)$, see (6), which gives $\Phi^*_{Mm,n} < 2\delta_d^{1/d}/[(nV_d)^{1/d} - 2\delta_d^{1/d}]$ for the hypercube $[0,1]^d$.

The determination of packings with greatest density is a subject that has a rich history. When $d = 2$, the fact that the hexagonal packing arrangement has highest density $\delta_2 = \pi\sqrt{3}/6$ among lattice packings was known to Lagrange (1773), but the first proof of its optimality among all possible arrangements is due to L. Fejes Tóth (1943). This gives $\Phi^*_{Mm,n} < 1/(3^{1/4}\sqrt{n/2} - 1)$ for the unit square. A further slight improvement is given by Oler (1961), with $\Phi^*_{Mm,n} < [1 + \sqrt{1 + 2(n-1)/\sqrt{3}}]/(n-1)$. For $d = 3$, Kepler conjectured in 1611 that the face-centred cubic lattice (commonly exhibited on fruit stands), with $\delta_3 = \pi\sqrt{2}/6$, has highest density among all possible arrangements. Gauss proved in 1831 that this was indeed the case among lattice packings, and the Kepler conjecture was finally proved by Hales and Ferguson in 2006; see Lagarias (2011) for details, including corrections that followed the original proof of the conjecture. This gives $\Phi^*_{Mm,n} < 1/[(n/\sqrt{2})^{1/3} - 1]$ for the unit cube.

 ISSN: 2102-6238

Little is known about optimal packings in dimension d larger than 3, except for lattices, see the major reference (Conway and Sloane, 1999). The densest lattice packings are known up to $d = 8$, the optimality of the E_8-lattice among all possible arrangements (including non-lattice ones) has only been proved recently (Viazovska, 2016).

Optimal arrangements in terms of covering minimise the average number of spheres containing a point of the space, which Conway and Sloane (1999, Chap. 2) call thickness. When $\mathscr{X}$ can tile $\mathbb{R}^d$, thinnest arrangements provide lower bounds on $\Phi^*_{mM,n}$, but are only known for $d = 1$ and 2. Thinnest lattices are known up to $d = 5$ (they differ from packing-optimal lattices, and are related with them through a duality property for $d = 2$ and 3); Conway and Sloane (1999, Chap. 3) give the best covering lattices known up to $d = 24$. Further results are provided for instance in (Vallentin, 2003).

Design constructions based on lattices are considered in (Niederreiter, 1992, Chap. 5) in connection with the numerical integration of periodic functions, see also Bates et al. (1996) and Riccomagno et al. (1997) for the relation with optimal experimental design for Fourier regression. The performance of lattices in terms of discrepancy measures is considered in (Niederreiter, 1992, Chap. 5) and (Hickernell, 1998b). One may refer, e.g., to Korobov (1960), Sloan and Walsh (1990), Sloan and Reztsov (2002) and Nuyens (2007) for the construction of good lattice rules.

Performance of random designs. Wahl et al. (2014) give the distribution of $\Phi_{Mm}(X_n)$ when the n points $\mathbf{x}_i$ in X_n are independently uniformly distributed in $[0,1]^d$, together with approximations of this distribution for large n. As suggested in their paper, this result can be used to evaluate the maximin performance of a given design, through the calculation of the probability that a random design will have a higher Φ_{Mm} value. Otherwise, one may refer to Janson (1986, 1987) for results on the asymptotic distribution of Φ_{mM} for random designs uniformly distributed in the hypercube, and to Aaron et al. (2014) for an extension to distributions with Lipschitz continuous density and bounded support.

2.3. The issue of dimension

Making simplifying assumptions on f. From the inequality (4), $\Phi_{mM}(X_n) < \varepsilon$ requires $n > \mathrm{vol}(\mathscr{X})/(V_d\,\varepsilon^d)$. Difficulties can thus be expected in high dimension with moderate n when the accuracy of the prediction of $f(\cdot)$ at a point $\mathbf{x}$ depends on the isotropic distance between $\mathbf{x}$ and the design X_n, as described in Section 2.1. Kernel-based prediction methods adapted to the assumption that $f(\cdot)$ only depends on a few factors and their low-level interactions may then allow one to get around the curse of dimensionality, see, e.g., Durrande et al. (2012) for additive models and Ginsbourger et al. (2014) for the Sobol'-Hoeffding decomposition used in sensitivity analysis. See also Rasmussen and Williams (2006, Chap. 4). Along the same vein, discrepancy (see Section 2.5.1) relies on volumes rather than distances and corresponds to kernels that favour functions that can be modelled well by sums of few factors interactions, see Hickernell (1998a,b). Low-discrepancy designs which are known to perform well for integration (see, e.g., Niederreiter, 1992) may thus also be good for function approximation in high dimension.

Without suitable assumptions of $f(\cdot)$, for large d we shall necessarily encounter the intrinsic difficulties of high dimension, as beautifully explained in (Blum et al., 2016, Chap. 2). We only mention two of them, considering situations where $\mathscr{X}$ is a ball or a cube.

 ISSN: 2102-6238

Towards non-uniform space-filling design. The first difficulty is common to the ball and the cube: the volumes of the unit ball and unit cube are concentrated along their borders. For $\varepsilon \in [0,1)$, denoting by $\mathscr{K}_\varepsilon$ the hypercube $[-(1-\varepsilon)/2,(1-\varepsilon)/2]^d$, we have

$$\frac{\mathsf{vol}(\mathscr{K}_0)-\mathsf{vol}(\mathscr{K}_\varepsilon)}{\mathsf{vol}(\mathscr{K}_1)} = \frac{\mathsf{vol}(\mathscr{B}(\mathbf{0},1))-\mathsf{vol}(\mathscr{B}(\mathbf{0},1-\varepsilon))}{\mathsf{vol}(\mathscr{B}(\mathbf{0},1))} = 1-(1-\varepsilon)^d > 1-\exp(-d\varepsilon)\,,$$

which quickly tends to one for any $\varepsilon > 0$ as d increases. When trying to spread n design points uniformly in a high dimensional $\mathscr{X}$, we can thus expect that a large majority of them will be close to the border of $\mathscr{X}$. To circumvent that difficulty, we may consider that prediction accuracy deep inside $\mathscr{X}$ is more important than along its border, and use a non-stationary kernel in a kernel-based method, or modify the metric used. For instance, when $\mathscr{X}$ is centered at $\mathbf{0}$, one may replace a stationary kernel $K(\mathbf{x},\mathbf{z}) = k(\|\mathbf{x}-\mathbf{z}\|)$ by $K'(\mathbf{x},\mathbf{z}) = k(\|\mathbf{x}-\mathbf{z}\|)/[s(\|\mathbf{x}\|)s(\|\mathbf{z}\|)]$, with $s(\cdot)$ a positive and increasing function on $\mathbb{R}^+$, see, e.g., Rasmussen and Williams (2006, Chap. 4). The change of metric can rely on a nonlinear mapping (warping) of $\mathbf{x}$: one may consider for instance the transformed space $\mathscr{X}' = \{\mathbf{x}' = \mathbf{x}/s(\|\mathbf{x}\|) : \mathbf{x} \in \mathscr{X}\}$, where $\mathscr{X}$ is centered at $\mathbf{0}$, $s(\cdot)$ is positive and increasing and such that $t \mapsto t/s(t)$ is strictly increasing for $t \in [0, \max_{\mathbf{x}\in\mathscr{X}} \|\mathbf{x}\|]$ (so that the mapping $\mathbf{x} \in \mathscr{X} \mapsto \mathbf{x}' = \mathbf{x}/s(\|\mathbf{x}\|)$ is one-to-one).

Rounding the cube. The second noticeable difficulty concerns the striking difference between the geometrical shapes of a ball and a cube in high dimension. As d increases, the volume of the unit cube is one but the maximum distance possible between two points is $\sqrt{d}$. In contrast, the volume V_d of the unit ball tends to zero but the maximum distance possible between two points equals two. A first consequence is that the squared distance between pairs of points uniformly distributed in the unit cube is concentrated around its mean value $d/6$: a direct application of Hoeffding's concentration inequality gives $\mathsf{Prob}\left\{\left|\|\mathbf{x}-\mathbf{z}\|^2 - d/6\right| > \beta\sqrt{d}\right\} < 2\exp(-2\beta^2)$ when $\mathbf{x}$ and $\mathbf{z}$ are independently uniformly distributed in the unit cube. A second consequence is that whereas the unit cube centered at the origin is contained in the unit ball for $d \le 4$, its 2^d vertices are at distance $\sqrt{d}/2$ from the origin and lie outside the ball for $d \ge 5$, and the volume of the complement of the ball with respect to the cube tends to one as d increases to infinity. On the other hand, the ball is never contained in the cube: the centre of each $(d-1)$-dimensional face of the cube is at distance 1/2 from the origin, well inside the ball. This indicates that when d is large, inferring the behaviour of an unknown $f(\cdot)$ seems much more difficult when $\mathscr{X}$ is a cube than when it is a ball. The usual motivation for assuming that $\mathbf{x} = (x_1,\ldots,x_d)$ lies in a cube is that it accounts for the situation where all input factors x_i take their extreme values at the same time. As an attempt to struggle with the curse of dimensionality, one may suppose that only q factors out of d can be extreme simultaneously. After suitable renormalisation, this can be achieved by taking $\mathscr{X}$ as the intersection between the unit cube centered at the origin and the ball $\mathscr{B}(\mathbf{0},\alpha\sqrt{d}/2)$: when $\sqrt{q/d} \le \alpha < \sqrt{(q+1)/d}$, $\mathscr{X}$ intersects all $(d-q)$-faces of the cube but none of its $(d-q-1)$-faces (a 0-face being a vertex, a 1-face an edge, and so on, a $(d-1)$-face is a face).

To summarize, the difficulties mentioned in this section point out that minimax and maximin optimal designs based on an isotropic notion of distance do not seem relevant for large d. A few suggestions on how to possibly circumvent some of those difficulties have been indicated,

 ISSN: 2102-6238

but deeper investigations are clearly required. This explains why we shall restrict our attention to situations where $d \lesssim 10$; when we mention large d in the rest of the paper, this should be interpreted as d around 10, or slightly larger than 10.

2.4. *Evaluation of* Φ_{mM}

Whereas the evaluation of $\Phi_{Mm}(X_n)$ is straightforward, that of $\Phi_{mM}(X_n)$ requires the maximisation of $d(\mathbf{x}, X_n)$ with respect to $\mathbf{x} \in \mathscr{X}$, see (1). The evaluation is of course much simpler when $\mathbf{x}$ is restricted to belong to a finite set, and a common approach is to (under-)estimate $\Phi_{mM}(X_n)$ by $\tilde{\Phi}_{mM}(X_n) = \max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n)$, with $\mathscr{X}_Q$ a finite set $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\} \in \mathscr{X}^Q$, typically a regular grid, or the first Q points of a low-discrepancy sequence. A large value of Q is required to obtain a precise approximation of $\Phi_{Mm}(X_n)$, and for a given level of accuracy Q should increase quickly with d. Indeed, the triangle inequality for Hausdorff distance gives

$$\begin{aligned}\max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n) \leq \Phi_{mM}(X_n) = d_H(X_n, \mathscr{X}) \quad &\leq \quad d_H(X_n, \mathscr{X}_Q) + d_H(\mathscr{X}_Q, \mathscr{X}) \\ &= d_H(X_n, \mathscr{X}_Q) + \Phi_{mM}(\mathscr{X}_Q) \\ &\leq \max\left\{\max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n), \Phi_{mM}(\mathscr{X}_Q)\right\} + \Phi_{mM}(\mathscr{X}_Q),\end{aligned}$$

and therefore $0 \leq \Phi_{mM}(X_n) - \max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n) \leq \Phi_{mM}(\mathscr{X}_Q)$ for any reasonable $\mathscr{X}_Q$ such that $\Phi_{mM}(\mathscr{X}_Q) \leq \max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n)$. On the other hand, one can easily construct examples with $\mathscr{X}_Q$ a regular grid such that $\Phi_{mM}(X_n) = \max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n) + \Phi_{mM}(\mathscr{X}_Q)$. The lower bound (4) then indicates that Q must grow as $O(1/\varepsilon^d)$ to ensure that $\Phi_{mM}(X_n) < \max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n) + \varepsilon$.

When $\mathscr{X}$ is a d-dimensional hypercube, or a convex bounded polyhedron (a polytope) of $\mathbb{R}^d$ with d small enough (say, $d \lesssim 5$), tools from algorithmic geometry provide the exact evaluation of $\Phi_{Mm}(X_n)$ (note that in most applications n is growing moderately with d, often linearly). The third method presented uses a Markov Chain Monte Carlo (MCMC) algorithm to construct an estimate of $\Phi_{Mm}(X_n)$ and can also be used with larger d.

Delaunay triangulation (Pronzato and Müller, 2012). Suppose that X_n is an n-point design in $\mathscr{X} = [0,1]^d$. Consider the set X'_m, with $m = (2d+1)n$ points, formed by X_n and its $2d$ reflections through the $(d-1)$-dimensional faces of $\mathscr{X}$. The Delaunay triangulation of X'_m yields a collection of d-dimensional simplices (each one having $d+1$ vertices), with the property that their circumscribed spheres $\mathscr{S}_j$ do not contain any point of X'_m in their interior, see Okabe et al. (1992), Boissonnat and Yvinec (1998). The value $\max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_n)$ is then attained when $\mathbf{x}$ is the centre of one of those spheres, and therefore, $\max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n)$ gives the exact value of $\Phi_{mM}(X_n)$ when $\mathscr{X}_Q$ is the finite set given by the centres of circumscribed spheres that belong to $\mathscr{X}$. One should notice, however, that the complexity of the construction grows exponentially fast with d, since $Q = |\mathscr{X}_Q| = O(m^{\lceil d/2 \rceil})$, which gives a computational time growing like $O(m^{1+\lceil d/2 \rceil})$, see Cignoni et al. (1998). The approach is therefore restricted to small values of d (even if removing points not necessary for the construction from X'_m). An example is shown in Figure 3-left.

The method requires the computation of the circumscribed sphere $\mathscr{S}$ enclosing $d+1$ points $\mathbf{u}_i$ in $\mathbb{R}^d$, which can be easily performed as follows. Let $\mathbf{U}$ denote the $d \times (d+1)$ matrix with columns given by the $\mathbf{u}_i$, and let $\mathbf{c}$ and r denote the centre and radius of $\mathscr{S}$, respectively. By

 ISSN: 2102-6238

definition, $\mathbf{c}$ and r satisfy $\|\mathbf{c}-\mathbf{u}_i\|^2 = r^2$, that is $2\mathbf{u}_i^\top \mathbf{c} + (r^2 - \mathbf{c}^\top \mathbf{c}) = \|\mathbf{u}_i\|^2$, for all i. Denoting $w = r^2 - \mathbf{c}^\top \mathbf{c}$ and $\mathbf{d} = \mathrm{diag}(\mathbf{U}^\top \mathbf{U})$, the vector with i-th component $\|\mathbf{u}_i\|^2$, we can rewrite this in matrix form as

$$[2\mathbf{U}^\top \mathbf{1}] \begin{pmatrix} \mathbf{c} \\ w \end{pmatrix} = \mathbf{d}\,,$$

and obtain the values of $\mathbf{c}$ and $r = \sqrt{w + \mathbf{c}^\top \mathbf{c}}$ from the solution of this linear system. Note that $|\det[\mathbf{U}^\top \mathbf{1}]|/d! = \mathsf{vol}[\mathbf{u}_1, \ldots, \mathbf{u}_{d+1}]$, the volume of the simplex formed by the $\mathbf{u}_i$. Therefore, $\det[2\mathbf{U}^\top \mathbf{1}] \neq 0$ when the points are in general position, and there is no circumscribed sphere (or one may consider that its centre is at infinity) when the system is inconsistent.

Voronoï tessellation (Cortés and Bullo, 2005, 2009). A collection of n sets $\mathscr{C}_i$ forms a tessellation of the set $\mathscr{A}$ if $\mathscr{C}_i \cap \mathscr{C}_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^n \mathscr{C}_i = \mathscr{A}$. The Voronoï tessellation based on the sites, or generators, in $X_n \in \mathscr{X}^n$, partitions $\mathbb{R}^d$ into n cells $\mathscr{C}_i$, with $\mathscr{C}_i$ consisting of points closer to $\mathbf{x}_i$ than to any other site in X_n. Each cell thus corresponds to a convex polyhedron in $\mathbb{R}^d$, some cells being open (and infinite). When $\mathscr{X}$ is a polytope of $\mathbb{R}^d$, the intersections of the $\mathscr{C}_i$ with $\mathscr{X}$ provide a tessellation of $\mathscr{X}$ into n bounded convex polyhedra. The value $\max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_n)$ is then attained when $\mathbf{x}$ is a vertex of one of these polyhedra, and one only need to compute $\max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n)$ with $\mathscr{X}_Q$ the collection of these vertices. Again, the construction is computationally expensive for large d, with Q growing like $O(n^{\lceil d/2 \rceil})$, see Klee (1980). Dealing with infinite cells can be easily avoided by adding a few generators $\mathbf{x}'_j$ out of $\mathscr{X}$ (at least $d+1$), far enough from $\mathscr{X}$ to ensure that the corresponding cells do not intersect $\mathscr{X}$. For instance, when $\mathscr{X} = [0,1]^d$, we can add to X_n the $2d$ generators given by $(1/2)\mathbf{1} \pm a\,\mathbf{e}_j$, $j = 1, \ldots, d$; taking $a > 1/2 + \sqrt{d}$ then implies $\max_{\mathbf{x}\in\mathscr{X}, \mathbf{x}_i \in X_n} \|\mathbf{x} - \mathbf{x}_i\| \leq \sqrt{d} < \min_{\mathbf{x}\in\mathscr{X}} \|\mathbf{x} - \mathbf{x}'_j\| = a - 1/2$. The n Voronoï cells corresponding to generators in X_n are then finite and their intersections with $\mathscr{X}$ are not perturbed by the presence of the extra generators $\mathbf{x}'_j$. These polyhedral intersections can be computed using for example the method in (Walter and Piet-Lahanier, 1989) — with a non-negligible contribution to the total computational cost of the method. An illustration is given in Figure 3-right.

Estimation via MCMC. Suppose that we have a sample of Q points $\mathbf{x}^{(j)}$ i.i.d. in $\mathscr{X}$. We can then compute the Q distances $d_j = d(\mathbf{x}^{(j)}, X_n)$, with associated order statistics $d_{1:Q} \geq d_{2:Q} \geq \cdots \geq d_{Q:Q}$. Let k be a fixed integer between 1 and Q (assuming that $Q \gg d$, we suggest $k = \max\{10, d\}$). Using results in (Zhigljavsky and Žilinskas, 2007, Chap. 2) and (Zhigljavsky and Hamilton, 2010) based on extreme-value theory, we can then estimate $\Phi_{mM}(X_n)$ by

$$\widehat{\Phi}_{mM}(X_n) = d_{1:Q} + C_k (d_{1:Q} - d_{k:Q})$$

where $C_k = b_1/(b_k - b_1)$ with $b_i = \Gamma(i + 1/d)/\Gamma(i)$. Moreover, the asymptotic confidence level of the interval

$$I_{k,\delta} = \left[d_{1:Q}, d_{1:Q} + \frac{d_{1:Q} - d_{k:Q}}{(1-\delta^{1/k})^{-1/d} - 1} \right]$$

tends to $1-\delta$ for $Q \to \infty$. When the points $\mathbf{x}^{(j)}$ are fixed, the precision of the estimation of $\Phi_{mM}(X_n)$ by $\widehat{\Phi}_{mM}(X_n)$ is poor and the interval $I_{k,\delta}$ is large, unless Q is very large. However, the order statistics $d_{j:Q}$ for large j carry little information about the value of $\Phi_{mM}(X_n)$. This suggests using a multilevel splitting algorithm, for which all $\mathbf{x}^{(j)}$ at distance d_j from X_n less than some

 ISSN: 2102-6238

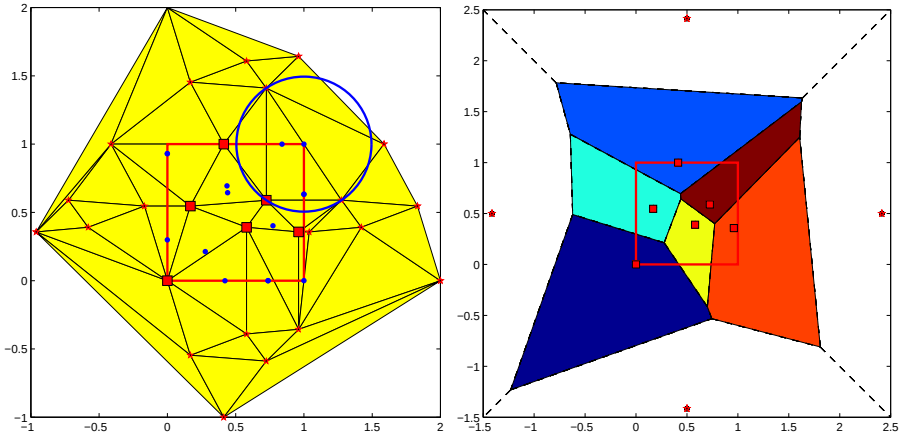

FIGURE 3. *Evaluation of $\Phi_{mM}(X_n)$ for X_n a 6-point design (squares) in $\mathscr{X} = [0,1]^2$. Delaunay triangulation (left): there are 45 triangles, 12 centres of circumscribed circles (dots), the circle with largest radius is plotted. Voronoï tessellation (right): the $2d = 4$ additional generators are indicated by stars, the six (finite) cells intersected with $\mathscr{X}$ have 14 distinct vertices (28 in total), the most distant from X_n corresponds to the upper right corner of $\mathscr{X}$.*

L_ℓ are replaced by points sampled independently (and uniformly) in the set $\mathscr{X}(L_\ell) = \{\mathbf{x} \in \mathscr{X} : d(\mathbf{x}, X_n) > L_\ell\}$, for an increasing sequence of levels L_ℓ. The method of Guyader et al. (2011) is particularly attractive in this context for generating uniform samples in a sequence of sets $\mathscr{X}(L_\ell)$, due to its efficiency and the straightforward level sequence that it proposes: at step ℓ, the next level is set at $L_{\ell+1} = \min_{j=1,\dots,Q} d_j$, and the point $\mathbf{x}_{j^*}$ (unique with probability one) such that $d_{j^*} = L_{\ell+1}$ is replaced by a new point sampled in $\mathscr{X}(L_{\ell+1})$. These splitting iterations are stopped when the width of the confidence interval $I_{k,\delta}$, with $\delta = 0.05$, say, is less than some prescribed precision ε.

One may use rejection sampling (the acceptance-rejection method) to sample uniformly in $\mathscr{X}(L)$ when L is small enough. However, sampling uniformly in $\mathscr{X}(L)$ is not straightforward when L approaches $\Phi_{mM}(X_n)$, and when the number of rejections becomes larger than some given threshold (say, 100) one may switch to the MCMC method with Metropolis-Hastings transitions proposed in (Guyader et al., 2011). Denote by $\mathsf{Proj}_{\mathscr{X}}(\mathbf{x})$ the projection of $\mathbf{x}$ onto $\mathscr{X}$ (a truncation when $\mathscr{X}$ is an hypercube). At step ℓ, first replace $\mathbf{x}_{j^*}$ by a $\mathbf{x}_{j^{**}}$ chosen at random among the other $\mathbf{x}_j$, second perform K successive steps of a random walk, where a movement from $\mathbf{x}$ to $\mathsf{Proj}_{\mathscr{X}}(\mathbf{x}+\mathbf{z})$, with $\mathbf{z}$ normally distributed $\mathscr{N}(\mathbf{0}, \sigma \mathbf{I}_d)$, is accepted if and only if $d(\mathbf{x}+\mathbf{z}, X_n) > L_{\ell+1} = d_{j^*}$. Our numerical experiments indicate that choosing $Q = \max\{nd, 100\}$, K between 10 and 20 and σ equal to the lower bound $\underline{R}_n^*$ given by (4) gives satisfactory results. An illustration is given in Figure 4.

Figure 5 presents boxplots of (the logarithms of) computing times for the three methods above ($K = 10$, $\delta = 0.05$, $\varepsilon = 10^{-3}$ for MCMC), for $n = 50$ and d varying from 2 to 5 (we performed 100 evaluations of Φ_{mM} for random Latin-hypercube designs X_n in $[0,1]^d$). Voronoi tessellation appears to be the fastest when $d = 2$ and 3, Delaunay triangulation is the fastest for $d = 4$ and MCMC is preferable for $d \geq 5$. The computational time grows roughly exponentially with d for the first two methods, but is only linear in d for MCMC.

 ISSN: 2102-6238

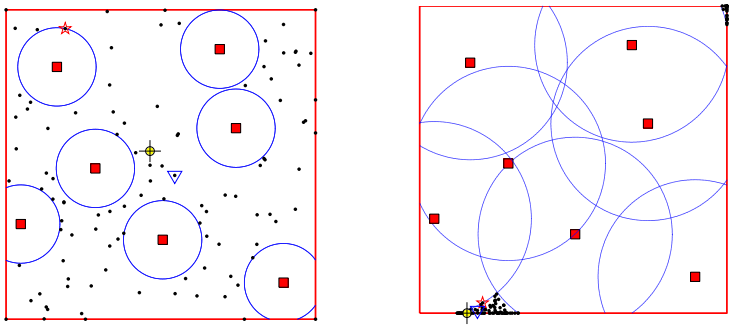

FIGURE 4. *Evaluation of* $\Phi_{mM}(X_n)$ *via MCMC for a 7-point design (squares) in* $\mathcal{X} = [0,1]^2$. *The initial* $\mathbf{x}^{(j)}$, $j = 1,\ldots,Q = 104$, *correspond to 100 random points in* $\mathcal{X}$ *(uniformly distributed) and the 4 vertices. The figure on the left (resp. on the right) shows the situation after 30 (resp. 400) splitting iterations: the 7 balls centered at the design points have radius* $\min_{j=1,\ldots,Q} d(\mathbf{x}^{(j)}, X_7)$ *for the current set of points* $\mathbf{x}^{(j)}$*; the point* $\mathbf{x}_{j^*}$ *(star) that lies exactly on a circle is replaced by* $\mathbf{x}_{j^{**}}$ *(triangle) which is then updated into the point indicated by a cross after 10 Metropolis-Hastings transitions.*

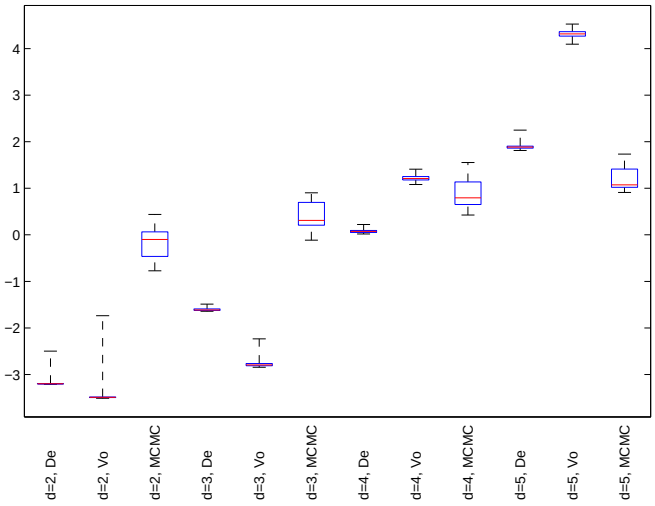

FIGURE 5. *Boxplot of* log*(computing times) (in seconds) for the evaluation of* $\Phi_{mM}(X_n)$ *with Delaunay triangulation (De), Voronoi tessellation (Vo) and a MCMC algorithm (MCMC). 100 repetitions with* X_n *a random Latin-hypercube design in* $[0,1]^d$, $n = 50$ *and* $d = 2,3,4,5$*; the median, 25% and 75% percentiles are indicated, together with most extreme values.*

2.5. *Relation with other criteria for optimal design*

2.5.1. *Model-free design criteria*

Discrepancy. A rather intuitive approach to achieve the objective of observing "everywhere" in $\mathcal{X}$, as formulated in Section 1, is to minimise a "distance" between the distribution of observations sites $\mathbf{x}_i$ and the uniform distribution. The notion of discrepancy captures this intuition well, see (Niederreiter, 1992, Chap. 2), and different versions can be considered which receive an analytical

 ISSN: 2102-6238

expression, see Hickernell (1998a,b), Fang and Ma (2001), and can therefore be minimised, see Fang et al. (2003), Fang et al. (2005) and (Fang et al., 2006, Chaps. 3&4). The use of low-discrepancy sequences forms a cheap alternative to algorithmic discrepancy minimisation (Niederreiter, 1992, Chaps. 3&4). Assume that $\mathscr{X} = [0,1]^d$. Interestingly enough, a bound on the dispersion $\Phi_{mM}(X_n)$ of an n-point design can be derived from its (extreme) discrepancy $D_n(X_n)$,

$$\Phi_{mM}(X_n) \leq \sqrt{d}\, D_n^{1/d}(X_n)\,,$$

see (Niederreiter, 1992, Th. 6.6) (this formalises the common sense that a sequence distributed uniformly in $\mathscr{X}$ is dense in $\mathscr{X}$). We therefore have bounds on Φ_{mM} for designs X_n formed by the first n points of a low-discrepancy sequence (for which $D_n = O[(\log n)^d/n]$), see for instance Mitchell (1990) for the Halton sequence. However, numerical evaluations indicate that these bounds are rather pessimistic.

Divergence measures and entropy. An alternative approach is to consider that the $\mathbf{x}_i$ are i.i.d. with a density φ in $\mathscr{X}$, and to minimise an estimate $\hat{D}_n$ of a divergence measure D between φ and the uniform density υ on $\mathscr{X}$, or to maximise an estimate $\hat{H}_n$ of the entropy H of φ, $\hat{D}_n$ and $\hat{H}_n$ being estimated from X_n considered as a sample of size n. Different divergences or entropies can be considered, see, e.g., Basseville (2013), minimising a divergence $D(\varphi,\upsilon)$ between φ and υ being generally equivalent to maximising a suitable entropy $H(\varphi)$. Jourdan and Franco (2009b, 2010) maximise a plug-in estimator of the Shannon entropy $H_1(\varphi) = -\int \varphi(\mathbf{x}) \log \varphi(\mathbf{x})\, \mathrm{d}\mathbf{x}$ and estimate φ with (Gaussian) kernel smoothing; i.e., their density estimator is

$$\widehat{\varphi}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \tilde{\varphi}_{\sigma_n^2}(\mathbf{x} - \mathbf{x}_i)\,, \tag{9}$$

where σ_n is a bandwidth parameter and $\tilde{\varphi}_{s^2}$ is the density of the d-dimensional normal distribution $\mathscr{N}(0, s^2\mathbf{I}_d)$. To avoid computationally demanding numerical integrations, they substitute the empirical estimator $\hat{H}_{1,n} = -(1/n)\sum_{i=1}^{n} \log \widehat{\varphi}_n(\mathbf{x}_i)$ for $H_1(\widehat{\varphi}_n)$. In (Jourdan and Franco, 2009a), the authors maximise the Nearest-Neighbour (NN) estimator of Shannon entropy derived by Kozachenko and Leonenko (1987), which amounts at maximising the product of the distances between each design point and its NN in X_n (see also the next paragraph on the maximisation of Euclidean functionals on graphs).

Other entropies can be considered too, such as Tsallis entropy of order α, defined by $H_\alpha(\varphi) = [1/(\alpha-1)]\,[1 - \int \varphi^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}]$ for $\alpha \neq 1$, with $H_\alpha(\varphi) \to H_1(\varphi)$ as $\alpha \to 1$. When $\alpha > 0$ and φ is constrained to be supported on $\mathscr{X}$, $H_\alpha(\varphi)$ is maximum for the uniform distribution. Notice that the second-order Tsallis entropy ($\alpha = 2$) yields a particularly simple construction, since $\int_{\mathbb{R}^d} \widehat{\varphi}_n^2(\mathbf{x})\,\mathrm{d}\mathbf{x} = (1/n^2)\sum_{i,j=1}^{n} \tilde{\varphi}_{2\sigma_n^2}(\mathbf{x}_i - \mathbf{x}_j)$ gives

$$H_2(\widehat{\varphi}_n) = 1 - \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{\varphi}_{2\sigma_n^2}(\mathbf{x}_i - \mathbf{x}_j)\,. \tag{10}$$

Also note that $\hat{H}_{1,n}$ and $H_2(\widehat{\varphi}_n)$ only depend on distances between pairs of points in X_n, as it will also be the case for the design criteria in the next two paragraphs.

 ISSN: 2102-6238

Euclidean functionals on graphs. In a seminal paper, Beardwood et al. (1959) have shown that when n points $\mathbf{x}_i$ are i.i.d. with the density φ in $\mathscr{X} = [0,1]^d$, the lengths $|e_i|$ of the edges e_i of the Traveling-Salesman (TS) graph $\mathscr{G}_{TS}(X_n)$ having the $\mathbf{x}_i$ as vertices satisfy

$$\frac{\sum_{e_i \in \mathscr{G}_{TS}(X_n)} |e_i|}{n^{(d-1)/d}} \xrightarrow{\text{a.s.}} C(d) \int \varphi^{(d-1)/d}(\mathbf{x})\, \mathrm{d}\mathbf{x} \text{ as } n \to \infty,$$

where $C(d)$ only depends on d. This result has then been extended by Steele (1981) to other Euclidean functionals, see also Redmond and Yukich (1994) for references and a unifying framework based on the notion of quasiadditivity, and has generated an abundant literature. One may refer for instance to Redmond and Yukich (1996), Yukich (1998), Penrose and Yukich (2003), Wade (2007), Penrose and Yukich (2011) for extensions to power-weighted edge lengths, considering various types of convergence and different conditions on φ. Basically, when φ is uniform on $\mathscr{X}$, for β in a suitable range we have

$$\Phi_{\mathscr{G},\beta}(X_n) = \frac{\sum_{e_i \in \mathscr{G}(X_n)} |e_i|^\beta}{n^{(d-\beta)/d}} \to C(\beta,d) \int \varphi^{(d-\beta)/d}(\mathbf{x})\, \mathrm{d}\mathbf{x} \text{ as } n \to \infty, \tag{11}$$

with $C(\beta,d)$ a constant depending only on β and d (and on the type of graph considered). Here $\mathscr{G}(X_n)$ may denote the Minimum Spanning Tree (MST), or the (k-th) Nearest-Neighbour (NN), TS, Voronoï, Delaunay, Gabriel or sphere of influence graph... and the type of convergence (L^q or a.s.) depends on the range of admissible β and on properties of the graph considered. For instance, Penrose and Yukich (2011) show that for the k-th NN graph, we have a.s. and L^2 convergence for $\beta > 0$ and L^q convergence ($q = 1,2$) for $\beta \in (-d/q, 0)$; see also Leonenko et al. (2008). An informal justification for the growth rate $n^{(d-\beta)/d}$ for $S_n = \sum_{e_i \in \mathscr{G}(X_n)} |e_i|^\beta$ is as follows: consider X'_n formed by 2^d identical copies of $X_n \in [0,1]^{d \times n}$, X'_n having $n' = 2^d n$ points in $[0,2]^d$. For a suitable graph $\mathscr{G}$, the effect of edges connecting points lying in two adjacent hypercubes can be neglected, and $\sum_{e'_i \in \mathscr{G}(X'_{n'})} |e'_i|^\beta \approx 2^d S_n$. Here $X'_{n'} \in [0,2]^{d \times n'}$, but renormalisation to $[0,1]^d$ yields a design $X_{n'} \in [0,1]^{d \times n'}$. The edges in $\mathscr{G}(X_{n'})$ are twice shorter than in $\mathscr{G}(X'_{n'})$, therefore $S_{n'} = \sum_{e_i \in \mathscr{G}(X_{n'})} |e_i|^\beta \approx 2^{d-\beta} S_n$, and $S_n = O(n^t)$ gives $2^{d-\beta} = 2^{dt}$, i.e., $t = 1 - \beta/d$.

When $\alpha > 1$, $H_\alpha(\varphi)$ is maximum for the uniform distribution, and maximising $H_\alpha(\varphi)$ is equivalent to minimising $\int \varphi^\alpha(\mathbf{x})\, \mathrm{d}\mathbf{x}$. Considering property (11), this can be achieved by minimising $\Phi_{\mathscr{G},\beta}(X_n)$ with $\beta < 0$. A classification of designs based on the values of the empirical mean and standard deviation of the edge lengths of the MST graph built from X_n is proposed in (Franco et al., 2009), but to the best of our knowledge the sum of power-weighted edge lengths $\Phi_{\mathscr{G},\beta}(X_n)$, for the MST or any other graph among those mentioned above, has scarcely been used as a criterion for space-filling design (NN, and to a lesser extent MST, being the most promising due to their easy construction; see, e.g., Jourdan and Franco (2009a) for the maximisation of the sum of logarithms of edge lengths in the NN graph).

L^q relaxation and energy. Denote by $d_*(\mathbf{x}_i) = \min_{j=1,\dots,n,\, j \neq i} \|\mathbf{x}_i - \mathbf{x}_j\|$ the NN distance to $\mathbf{x}_i$ in the design X_n. We can write $\Phi_{Mm}(X_n) = \min_{i=1,\dots,n} d_*(\mathbf{x}_i)$, so that, for any $q > 0$,

$$\frac{1}{n} \sum_{i=1}^n d_*^{-q}(\mathbf{x}_i) \leq \max_{i=1,\dots,n} d_*^{-q}(\mathbf{x}_i) = \Phi_{Mm}^{-q}(X_n) \leq \sum_{i=1}^n d_*^{-q}(\mathbf{x}_i)$$

 ISSN: 2102-6238

and

$$\left[\sum_{i=1}^{n} d_*^{-q}(\mathbf{x}_i)\right]^{-1/q} \leq \Phi_{Mm}(X_n) \leq \left[\frac{1}{n}\sum_{i=1}^{n} d_*^{-q}(\mathbf{x}_i)\right]^{-1/q}. \tag{12}$$

This indicates that a design $X^*_{n,NN,q}$ maximising $[\sum_{i=1}^n d_*^{-q}(\mathbf{x}_i)]^{-1/q}$ is nearly optimal for Φ_{Mm} when q is large enough. Indeed, (12) implies that $\Phi_{Mm}(X^*_{n,NN,q})/\Phi^*_{Mm,n} \geq n^{-1/q}$, which tends to 1 as $q \to \infty$. Note that $d_*(\mathbf{x}_i)$ is equal to the length $|e_i|$ of the edge with source at $\mathbf{x}_i$ in the NN graph $\mathscr{G}_{NN}(X_n)$ built on X_n, so that $X^*_{n,NN,q}$ minimises $\Phi_{\mathscr{G}_{NN},\beta}(X_n)$ defined by (11) with $\beta = -q < 0$.

A similar relaxation can be applied directly to distances between all pairs of points, with

$$\left[\sum_{i,j=1,i\neq j}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^{-q}\right]^{-1/q} \leq \Phi_{Mm}(X_n) \leq \left[\frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^{-q}\right]^{-1/q}.$$

Maximising $[\sum_{i,j=1,i\neq j}^n \|\mathbf{x}_i - \mathbf{x}_j\|^{-q}]^{-1/q}$ for $q > 0$ is equivalent to minimising the energy criterion

$$E_q(X_n) = \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^{-q}, \tag{13}$$

and a design X^*_{n,E_q} minimising $E_q(X_n)$ satisfies $\Phi_{Mm}(X^*_{n,E_q})/\Phi^*_{Mm,n} \geq \binom{n}{2}^{-1/q} \to 1$ as $q \to \infty$. The continuous extension of the criterion E_q is the q-energy $E_q(\mu) = \int\int \|\mathbf{x} - \mathbf{x}'\|^{-q} \mathrm{d}\mu(\mathbf{x})\mathrm{d}\mu(\mathbf{x}')$, with μ a probability measure on $\mathscr{X}$. The determination of measures with minimum q-energy is one of the main subjects in potential theory, see, e.g., Landkof (1972), Saff (2010). A design X^*_{n,E_q} minimising $E_q(X_n)$ is called a set of n Fekete points. When $q \geq d$, a sequence of Fekete points in $\mathscr{X}$ is asymptotically uniformly distributed, with $\sum_{i,j=1,i\neq j}^n \|\mathbf{x}_i - \mathbf{x}_j\|^{-q}$ growing as $n^{1+q/d}$ for $q > d$ and as $n^2 \log n$ when $q = d$. When $0 < q < d$, a minimum energy measure exists, it is called the q-equilibrium energy measure and is the weak limit of a sequence of empirical measures associated with Fekete points, see Hardin and Saff (2004). Minimum-energy measures in a design context are considered for instance in (Zhigljavsky et al., 2010). Optimal design based on the minimisation of $E_q(X_n)$ has been proposed by Audze and Eglais (1977) for $q = 2$. From the results just mentioned, values of q larger than d are preferable in a space-filling perspective, see (Pronzato et al., 2016b). Numerical experiments indicate that in general the minimisations of $E_{10d}(X_n)$ and $\Phi_{\mathscr{G}_{NN},-d}(X_n)$ yield designs resembling $X^*_{Mm,n}$.

2.5.2. *Model-based optimal design*

Consider the random-field modeling framework of Section 2.1. Suppose that the covariance $\mathsf{E}\{Z_{\mathbf{x}}Z_{\mathbf{x}'}\} = \sigma^2 K(\mathbf{x},\mathbf{x}')$ is isotropic, with $K(\mathbf{x},\mathbf{x}') = K^{(s)}(\mathbf{x},\mathbf{x}') = C^s(\|\mathbf{x}-\mathbf{x}'\|)$ for all $\mathbf{x},\mathbf{x}'$, where $s > 0$, $C(0) = 1$ and $C(t)$ is a decreasing function of t. Johnson et al. (1990) show that a minimax-optimal design $X^*_{mM,n}$ with highest index minimises the maximum over $\mathscr{X}$ of the kriging variance $\rho_n^2(\mathbf{x})$, asymptotically, as $s \to \infty$.

Maximum Entropy Sampling (MES) supposes that $Z_{\mathbf{x}}$ is Gaussian and selects a design X_n such that the entropy of the distribution of $\mathbf{z}_n = (Z_{\mathbf{x}_1}, \ldots, Z_{\mathbf{x}_n})^\top$ is maximum, and this amounts to

 ISSN: 2102-6238

maximising $\det \mathbf{K}_n$, see Shewry and Wynn (1987). For $K(\mathbf{x},\mathbf{x}') = K^{(s)}(\mathbf{x},\mathbf{x}') = C^s(\|\mathbf{x}-\mathbf{x}'\|)$, Johnson et al. (1990) show that a maximin-optimal design $X^*_{Mm,n}$ with lowest index is asymptotically optimum in the sense of MES when $s \to \infty$.

Inspired from results in (Joseph et al., 2015), we can also make a connection between MES and the maximisation of (an estimator of) the entropy of the distribution of design points. Consider the Gaussian process model with $K(\mathbf{x},\mathbf{x}') = K_\theta(\mathbf{x},\mathbf{x}') = \exp(-\theta\|\mathbf{x}-\mathbf{x}'\|^2)$, $\theta > 0$. For any design $X_n \in \mathcal{X}^n$, $\det \mathbf{K}_n < 1$ from Hadamard's inequality. On the other hand, for θ large enough $\mathbf{K}_n$ is strictly diagonally dominant, and Gershgorin's theorem implies that each interval $[1-\Delta_j, 1+\Delta_j]$, with $\Delta_j = \sum_{i=1, i\neq j}^n \{\mathbf{K}_n\}_{ij} < 1$, $j = 1,\ldots,n$, contains exactly one eigenvalue of $\mathbf{K}_n$. Therefore, for large θ,

$$0 > \log\det \mathbf{K}_n > \sum_{i=1}^n \log(1-\Delta_i) \simeq -\sum_{i\neq j} K_\theta(\mathbf{x}_i - \mathbf{x}_j) = n + n^2(\pi/\theta)^{d/2}[H_2(\hat{\varphi}_n) - 1]\,,$$

where $H_2(\hat{\varphi}_n)$ is the second-order Tsallis entropy (10) and $\hat{\varphi}_n$ is given by (9) with $\sigma_n^2 = 1/(4\theta)$. When the random-field model actually represents some prior knowledge on the function $f(\cdot)$ to be approximated/interpolated, a large θ urges that we take the trend behaviour of the model into account. However, when it is only used as a convenient tool to design a space-filling experiment, we may maximise $H_2(\hat{\varphi}_n)$, or $\log\det \mathbf{K}_n$, to obtain designs with good maximin performance.

3. General global and local methods for maximin and minimax optimal design

The criterion Φ_{Mm} and its extensions $\Phi_{\mathcal{G}_{NN},-q}(X_n)$ and E_q, respectively given by (11) and (13) with $q > 0$, are generally multimodal, and their global optimisation forms a difficult task. In fact, there is little hope to construct a certified maximin-optimal design already for $d = 2$ and n larger than a few dozen, and the situation gets worse for larger d. The determination of minimax-optimal designs is even more difficult. For suitable convex sets, maximin-optimal design is equivalent to sphere-packing, see Section 2.2, and the case $d = 2$ has attracted substantial attention, in geometry first (see for instance Croft et al., 1991, Chap. D, and the references therein), and then in the optimisation community for the challenges it raises. It is worthwhile to mention a few results that give an idea of the difficulty of the problem. When $\mathcal{X}$ is a square, exact solutions were known for some time for $n = 1,\ldots,9,10$ and then $n = p^2$ with $p = 4,5,6$, see Maranas et al. (1995). Solutions are provided in the same paper for all $n \leq 30$, using nonlinear programming with multistart (repetitions of runs with different initial designs). Graham and Lubachevsky (1996) investigate disk packings that exhibit repeated patterns as n increases. For $p \leq 6$, the optimal packing with p^2 points is made of p rows of p disks, but this is no longer true for $p \geq 7$ as shown by Nurmela and Östergård (1997) who provide packings up to $n = 50$. Their optimisation method relies on the minimisation of an energy criterion E_q with large q (up to 10^{50}), using a continuous steepest-descent algorithm with multistart; the design variables $\mathbf{x}_i$, which lie in the square $[-1,1]^2$, are first reparameterised as $(\sin(\theta_{1,i}), \sin(\theta_{2,i}))^\top$ where $\theta_{1,i}, \theta_{2,i}$ are unconstrained. Proving optimality of a given packing is also a difficult task, and Nurmela and Östergård (1999) extend the range of known certified optimal packings to $n \leq 27$. Extension to $n = 28$ and 29 is provided in (Markót and Csendes, 2005) using branch and bound and interval analysis (with a computational time exceeding 50h for one optimality certification). Results for disk packing in

 ISSN: 2102-6238

a circle are provided for $n \leq 65$ by Graham et al. (1998), who minimise E_q with large q, as in (Nurmela and Östergård, 1997), and also use a billiard algorithm, see Section 4.2.1. Best known packings ($d = 2$ and 3) are collected at the web site `http://www.packomania.com/`, see also `https://spacefillingdesigns.nl/` for other metrics than Euclidean and results in higher dimension for Latin hypercube designs.

The objective of this section is to present rather simple algorithms that can be used to generate acceptable solutions in reasonable computational time. Global optimisation is considered first; the solution that it provides can then be used to initialise one of the local-search algorithms presented in Section 3.2.

3.1. Global optimisation via simulated annealing

Global optimisation methods based on simulated annealing are easy to implement and provide reasonably good solutions provided that enough iterations are performed. Although convergence to a global optimum is guaranteed for some suitable versions of the algorithm, it is recommended to repeat several optimisations initialised at different designs (e.g., randomly generated). One may refer to Bertsimas and Tsitsiklis (1993) for a short illuminating exposition of simulated annealing applied to optimisation over a finite set, and to Auffray et al. (2012) for a detailed presentation in the case of maximin-optimal design over a compact set $\mathscr{X}$. See also Fang et al. (2006, Chap. 4) for a description of design optimisation algorithms in the finite setting (the construction of optimal Latin-hypercube designs for instance). Below we give the principles of a simple simulated-annealing algorithm which can be straightforwardly applied to maximin and minimax optimal design.

Let Φ denote the design criterion considered; we suppose that $\Phi > 0$ must be minimised and denote $\Phi^* = \min_{X_n \in \mathscr{X}^n} \Phi(X_n)$. The algorithm relies on the construction of a Markov chain $X_n^{(k)} \to X_n^{(k+1)}$ (initialised at some $X_n^{(1)}$) with stationary regime such that its invariant measure is the Gibbs measure with density proportional to $\exp[-C\Phi(X_n)]$ for some $C > 0$. Then, when $C = C_k$ increases slowly enough with k (typically, $C_k = c\log(k+1)$ for some sufficiently large c), one can ensure that $\lim_{k\to\infty} \mathsf{Prob}[\Phi(X_n^{(k)}) < \Phi^* + \varepsilon] = 1$ for any $\varepsilon > 0$. A faster increase of C_k, such as $C_k = ck$, is often used; it usually yields faster convergence to a local minimiser of Φ, without the previous convergence guarantee, however. The usual practice is to construct transitions with a Metropolis-Hastings algorithm, see Tierney (1998), generally incorporating a Gibbs sampler, see Roberts and Rosenthal (2006). Then, if $X_n^{(k+)}$ denotes the proposal generated from $X_n^{(k)}$, with density $Q[X_n^{(k)}, \cdot]$, $X_n^{(k+)}$ is accepted (i.e., $X_n^{(k+1)} = X_n^{(k+)}$) with probability $P_{acc}(k, X_n^{(k)}, X_n^{(k+)}) = \min\left\{\exp(C_k\,[\Phi(X_n^{(k)}) - \Phi(X_n^{(k+)})])\,Q[X_n^{(k+)}, X_n^{(k)}]/Q[X_n^{(k)}, X_n^{(k+)}], 1\right\}$ and $X_n^{(k+1)} = X_n^{(k)}$ otherwise. There are two hidden difficulties here: first, the ratio of densities $Q[X_n^{(k+)}, X_n^{(k)}]/Q[X_n^{(k)}, X_n^{(k+)}]$ may be difficult to compute, especially when $\mathscr{X}$ has a complicated shape or is not known explicitly, see Auffray et al. (2012) for a thorough discussion; second, the choice of the constant c in $C_k = c\log(k+1)$, or $C_k = ck$, is often critical. We thus suggest a simplified kernel, based on the acceptance probability

$$P_{acc}(k, X_n^{(k)}, X_n^{(k+)}) = \min\left\{\exp\left(C_k\,\frac{\Phi(X_n^{(k)}) - \Phi(X_n^{(k+)})}{\Phi(X_n^{(k)})}\right), 1\right\}. \tag{14}$$

 ISSN: 2102-6238

The transition kernel obtained for $C_k = C$ constant does not leave invariant the associated Gibbs measure, and we cannot guarantee the convergence property above. On the other hand, defining $\tau_k = \arg\min_{i=1,\dots,k} \Phi(X_n^{(i)})$, we have $\lim_{k\to\infty} \mathsf{Prob}[\Phi(X_n^{(\tau_k)}) < \Phi^* + \varepsilon] = 1$ for any $\varepsilon > 0$ (note that this property, although comforting, is rather weak and is also satisfied for a pure random search, or when $C_k = C$ constant in (14)).

Consider the two criteria $\Phi_1(X_n) = (\sum_{e_i \in \mathcal{G}_{NN}(X_n)} |e_i|^{-q})^{1/q}$ and $\Phi_2(X_n) = [E_q(X_n)]^{1/q}$, $q > 0$, see (11), (13). For q large enough, their minimisation is almost equivalent to the minimisation of $\Phi_3(X_n) = 1/\Phi_{Mm}(X_n)$. These criteria are positively homogeneous of degree -1, that is, $\Phi_i(\gamma X_n) = (1/\gamma)\,\Phi_i(X_n)$, $i = 1,2,3$. One may then choose C_k in (14) in order to set $P_{acc}(1, X_n, \gamma X_n)$ to some prescribed value when γ is close to 1. For the three Φ_i above, $P_{acc}(1, X_n, \gamma X_n) = 1$ for $\gamma \geq 1$ and can be approximated by $1 + C_1(\gamma - 1)$ for $\gamma \lesssim 1$, which equals 0 for $\gamma = \gamma_0 = 1 - 1/C_1$. We may then consider that $P_{acc}(1, X_n, \gamma X_n)$ is negligible for $\gamma < \gamma_0$, and take $C_1 = 1/(1-\gamma_0)$ obtained for a reasonable value of γ_0, for instance $\gamma_0 = 0.99$ when $C_k = c\log(k+1)$ with $c = 1/[(1-\gamma_0)\log 2]$, and $\gamma_0 = 0.75$ when $C_k = ck$ with $c = 1/(1-\gamma_0)$ (a smaller γ_0 is preferable here since the acceptance probability decreases faster). Similar reasoning with the minimisation of $\Phi_4(X_n) = \Phi_{mM}(X_n)$ yields the approximation $1 + C_k(1-\gamma)$ for $P_{acc}(1, X_n, \gamma X_n)$ when $\gamma \gtrsim 1$ (and $P_{acc}(1, X_n, \gamma X_n) = 1$ when $\gamma \leq 1$), and then $c = 1/[(\gamma_0 - 1)\log 2]$ and $c = 1/(\gamma_0 - 1)$ for the two cases $C_k = c\log(k+1)$ and $C_k = ck$, respectively.

To summarize, the proposed algorithm is as follows.

0) Select an initial design $X_n^{(1)}$, choose some large K, set $k = 1$.
1) Construct a random permutation $\mathscr{I}_k$ of the index set $\{1,\dots,n\}$, set $X'_n = X_n^{(k)}$.
   For $i = 1,\dots,n$, do
   — compute $\mathbf{x}'_j = \mathsf{Proj}_{\mathscr{X}}(\mathbf{x}_j + \mathbf{z}_i)$, where j denotes the i-th element of $\mathscr{I}_k$, $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_d)$ and $\mathsf{Proj}_{\mathscr{X}}$ is the projection on $\mathscr{X}$ (a simple truncation when $\mathscr{X}$ is an hypercube)
   — replace X'_n by $X''_n = X'_n \setminus \{\mathbf{x}_j\} \cup \{\mathbf{x}'_j\}$ with probability $P_{acc}(k, X'_n, X''_n)$ given by (14).
2) Set $X_n^{(k+1)} = X'_n$, $k \leftarrow k+1$; if $k = K$, stop, otherwise return to step 1.

The systematic scan of step 1, where j takes successively all values in $\mathscr{I}_k$, can be replaced by a random scan where j is sampled from $\mathscr{I}_k$ with some distribution. It may be uniform over $\mathscr{I}_k$, or biased in favour of the $\mathbf{x}_j$ that influence most the current value of Φ (for instance, considering pairs of closest points for Φ_{Mm}, or points $\mathbf{x}_j$ such that $\Phi_{mM}(X_n) = \max_{\mathbf{x}\in\mathscr{X}} \|\mathbf{x} - \mathbf{x}_j\|$ for Φ_{mM}), seemingly without crucial influence on performance. The choice $\sigma = \underline{R}_n^*$ given by (4) gives satisfactory results, letting σ decrease with k may improve performance close to the optimum.

3.2. *Local optimisation via gradient-type algorithms*

3.2.1. *Maximin optimal design*

Energy minimisation. A rather straightforward approach for the local maximisation of Φ_{Mm} consists in using a differentiable surrogate, such as E_q given by (13) for a large $q > 0$. Indeed, E_q can be minimised with a standard nonlinear programming method (e.g., steepest descent or conjugate gradient), without constraints when a suitable reparametrisation adapted to the particular form of $\mathscr{X}$ can be used, such as $\{\mathbf{x}_i\}_j = \sin(\theta_{ij})$, $i = 1,\dots,n$, $j = 1,\dots,d$ when $\mathscr{X} = [-1,1]^d$, see, e.g., Nurmela and Östergård (1997).

 ISSN: 2102-6238

Generalised gradient. The direct maximisation of Φ_{Mm} is more difficult, since the criterion is non-differentiable and not concave. However, Φ_{Mm} is globally Lipschitz (with constant $\sqrt{2}$) and is thus differentiable almost everywhere; see Cortés and Bullo (2005, 2009), who consider the equivalent disk-packing problem in the case $d = 2$. Considering X_n as a $d \times n$ matrix, for any pair $i \neq j \in \{1,\dots,n\}^2$ the gradient at X_n of the function $d_{ij} : X_n \in \mathscr{X}^n \mapsto \|\mathbf{x}_i - \mathbf{x}_j\|$ is the $d \times n$ matrix $\nabla_{d_{ij}}(X_n)$ with all columns equal to $\mathbf{0}$, except the i-th and j-th ones, respectively equal to $(\mathbf{x}_i - \mathbf{x}_j)/\|\mathbf{x}_i - \mathbf{x}_j\|$ and $-(\mathbf{x}_i - \mathbf{x}_j)/\|\mathbf{x}_i - \mathbf{x}_j\|$. The generalised gradient of Φ_{Mm} at X_n is then the set $\partial_{\Phi_{Mm}}(X_n)$ given by the convex hull of the $\nabla_{d_{ij}}(X_n)$ such that $\|\mathbf{x}_i - \mathbf{x}_j\| = \Phi_{Mm}(X_n)$, and when $\mathscr{X}$ is convex one may use a subgradient-type algorithm based on iterations

$$X_n^{(k+1)} = \mathsf{Proj}_{\mathscr{X}}\left[X_n^{(k)} + \gamma_k \tilde{\nabla}_{\Phi_{Mm}}(X_n^{(k)})\right], \tag{15}$$

with $\tilde{\nabla}_{\Phi_{Mm}}(X_n^{(k)})$ any element of $\partial_{\Phi_{Mm}}(X_n^{(k)})$ and (γ_k) a positive sequence of step-sizes satisfying $\lim_{k\to\infty} \gamma_k = 0$ and $\sum_k \gamma_k = \infty$. An abundant documentation exists on the choice of a suitable sequence (γ_k). The choice $\gamma_k = c/k$ for some $c > 0$ is typical (c of the same order of magnitude as $\Phi_{Mm}(X_n^{(1)})$ is reasonable). One may take $\tilde{\nabla}_{\Phi_{Mm}}(X_n^{(k)}) = \nabla_{d_{ij}}(X_n^{(k)})$ corresponding to a pair of points $(\mathbf{x}_i, \mathbf{x}_j)$ in $X_n^{(k)}$ at minimal distance. The iteration (15) then leaves all points in X_n unchanged but $\mathbf{x}_i$ and $\mathbf{x}_j$, which are respectively updated into $\mathsf{Proj}_{\mathscr{X}}[\mathbf{x}_i + \gamma_k \mathbf{u}_{ij}]$ and $\mathsf{Proj}_{\mathscr{X}}[\mathbf{x}_j - \gamma_k \mathbf{u}_{ij}]$, with $\mathbf{u}_{ij} = (\mathbf{x}_i - \mathbf{x}_j)/\|\mathbf{x}_i - \mathbf{x}_j\|$ having norm 1. One may also individualise the step-sizes and use a different sequence $(\gamma_k^{(i)})$ for each $\mathbf{x}_i$ in X_n. Following the idea of Kesten (1958), the decrease of step-sizes can be individualised too, with $\gamma_{k_i}^{(i)} = c/k_i$ for $\mathbf{x}_i$, where k_i is updated into $k_i + 1$ only when two successive directions of move for $\mathbf{x}_i$ make an angle larger than $\pi/2$.

The algorithm converges to a local maximum of Φ_{Mm}, but convergence is non-monotonic and rather slow — very slow if c is too small or too large. The extension to sphere packing in a polytope is straightforward: we maximise $\Phi_{SP}(X_n) = \min\{\Phi_{Mm}(X_n), 2\min_{i=1,\dots,n} d(\mathbf{x}_i, B(\mathscr{X}))\}$, with $B(\mathscr{X})$ the border of $\mathscr{X}$ (which consists of $(d-1)$-dimensional hyperplanes), and the generalised gradient of Φ_{SP} can be constructed easily, see Cortés and Bullo (2005, 2009). An advantage of sphere packing (which is equivalent to maximin optimal design for suitable $\mathscr{X}$, see Section 2.2), is that the generalised gradient directly accounts for the constraints $X_n \in \mathscr{X}^n$.

Steepest ascent. A naive implementation of the (projected) steepest-ascent method, which consists in determining the direction of fastest increase of Φ_{Mm} at $X_n^{(k)}$, and then maximising Φ_{Mm} in that direction (an unidimensional problem), may not converge to a local maximum, see, e.g., Example 9.1 in (Bonnans et al., 2006). One may use instead the method of successive approximations, where at iteration k all pairs of points such that $\|\mathbf{x}_i - \mathbf{x}_j\| \geq \Phi_{Mm}(X_n) - \varepsilon_k$ are taken into account for the construction of a generalised gradient, for a decreasing sequence of positive constants ε_k, see Sections III-7 and IV-6 of (Dem'yanov and Malozemov, 1974). The method ensures monotonic convergence towards a local maximum of Φ_{Mm} and is easy to implement when $\mathscr{X}$ has a simple shape. However, its slow convergence confines its use to problems with small n and d, and the billiard simulations of Section 4.2.1 seem more efficient.

 ISSN: 2102-6238

3.2.2. *Minimax optimal design*

Similar ideas can be used for the minimisation of Φ_{mM}, which is non-differentiable and non-convex but is globally Lipschitz with constant 1. When $\mathscr{X}$ is convex, one may use a subgradient-type algorithm based on

$$X_n^{(k+1)} = X_n^{(k)} - \gamma_k \tilde{\nabla}_{\Phi_{mM}}(X_n^{(k)}) \tag{16}$$

(note that no projection on $\mathscr{X}$ is required, compare with (15)). One may take $\tilde{\nabla}_{\Phi_{mM}}(X_n^{(k)})$ having all its columns equal to $\mathbf{0}$ but the i-th one, which is equal to $(\mathbf{x}_i - \mathbf{x}^*)/\|\mathbf{x}_i - \mathbf{x}^*\|$, where $\mathbf{x}_i \in X_n$ and $\mathbf{x}^* \in \mathscr{X}$ are such that $\|\mathbf{x}_i - \mathbf{x}^*\| = \Phi_{mM}(X_n)$. The sequence (γ_k) is as in Section 3.2.1, $\mathbf{x}_i$ and $\mathbf{x}^*$ can be obtained by the methods used in Section 2.4 to evaluate Φ_{mM}. As discussed in that section, this confines the use of the algorithm to small dimension d.

When a Voronoï tessellation is used, one may apply a subgradient iteration similar to (16) to each $\mathbf{x}_i$ individually, considering the furthest point from $\mathbf{x}_i$ in its own Voronoï cell. This gives

$$\mathbf{x}_i^{(k+1)} = \mathbf{x}_i^{(k)} - \gamma_{k,i}(\mathbf{x}_i^{(k)} - \mathbf{x}^{*,i})/\|\mathbf{x}_i^{(k)} - \mathbf{x}^{*,i}\|\,,$$

where $\mathbf{x}^{*,i} = \arg\max_{\mathbf{x}\in\mathscr{C}_i^{(k)}} \|\mathbf{x} - \mathbf{x}_i^{(k)}\|$ and $\mathscr{C}_i^{(k)}$ is the Voronoï cell with generator $\mathbf{x}_i^{(k)}$, see Cortés and Bullo (2005, 2009); see also the `kmeans` method in Section 4.3.3. A discrete version can also be used, based on a Voronoï tessellation of a discretisation $\mathscr{X}_Q$ of $\mathscr{X}$.

4. Specific methods for maximin and minimax optimal design

The objective of this section is to present a few optimisation methods that rely upon the particular form of the design problem. The first one is a simple greedy (one-step-ahead) algorithm, called "coffee-house design" in (Müller, 2001), (Müller, 2007, Chap. 4). See also Kennard and Stone (1969). We show that it generates designs X_n having at least 50% minimax and maximin efficiency.

4.1. A greedy algorithm with 50% efficiency

The algorithm is as follows:

0) Select $\mathbf{x}_1 \in \mathscr{X}$, set $X_1 = \{\mathbf{x}_1\}$ and $k = 1$.
1) For $k = 1, 2 \ldots$ do
   find $\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_k)$, set $X_{k+1} = X_k \cup \{\mathbf{x}^*\}$.

The point $\mathbf{x}^*$ can be obtained by a Voronoï tessellation of $\mathscr{X}$ (when d is small enough) or a MCMC method, see Section 2.4. Of course, its determination is much facilitated when a finite set $\mathscr{X}_Q$ is substituted for $\mathscr{X}$. An illustration is presented in Figure 6-left. Besides its simplicity, an important merit of this construction is that it guarantees the following efficiency bounds for X_k (efficiencies belong to $[0,1]$ by definition):

$$\frac{\Phi^*_{mM,k}}{\Phi_{mM}(X_k)} \geq \frac{1}{2}\ (k \geq 1) \quad \text{and} \quad \frac{\Phi_{Mm}(X_k)}{\Phi^*_{Mm,k}} \geq \frac{1}{2}\ (k \geq 2)\,. \tag{17}$$

Indeed, by construction $\Phi_{Mm}(X_{k+1}) = d(\mathbf{x}_{k+1}, X_k) = \Phi_{mM}(X_k)$ for all $k \geq 1$. Therefore, (2) implies that $\Phi^*_{mM,k} \geq (1/2)\,\Phi_{mM}(X_k)$ and that $\Phi^*_{Mm,k+1} \leq 2\Phi^*_{mM,k} \leq 2\,\Phi_{mM}(X_k) = 2\,\Phi_{Mm}(X_{k+1})$,

 ISSN: 2102-6238

see Gonzalez (1985). The design X_n generated by this algorithm can be used to initialise those presented in Sections 4.2.2 and 4.3.3 (which preserve the 50% efficiency bound for the criterion that is optimised).

4.2. Maximin optimal design

4.2.1. Billiards

For small d, following Lubachevsky and Stillinger (1990) and Lubachevsky (1991), one can efficiently generate tight (jammed) packings in $\mathscr{X}$ by simulating a billiard with n rigid balls $\mathscr{B}(\mathbf{x}_i, r)$ in $\mathscr{X}$, having identical radius r and initial velocities $\mathbf{v}_i$, $i = 1, \ldots, n$. In absence of collision, the balls move along straight lines, and the model assumes that all collisions, against boundaries or between balls, are elastic. When a ball encounters a boundary, the velocity component parallel to the boundary is unchanged, but the sign of normal component is reversed. When two balls have a collision, their normal velocity components are unchanged, but the balls exchange their parallel velocity components. If r is fixed, the behaviour is typical of a chaotic system, see, e.g., Sinai and Chernov (1987). To ensure that jamming occurs at a finite time, we let $r = r(t) = r(0) + st$ increase with the simulation time t at constant speed s. The normal velocity components of the balls must then be increased by some factor $h \geq 2s$ after each of their collision, to avoid that they overlap or stick to each other or to a border (this steady increase of velocities can be compensated by a periodic renormalisation of the $\mathbf{v}_i$ to avoid numerical difficulties). In its simplest form, the billard algorithm advances from collision to collision, the calculation of next collision time being straightforward (as the minimum solution of a set of quadratic equations). Collision times t_i then correspond to time steps for the algorithm, with i indicating the iteration number. The implementation is rather straightforward when $\mathscr{X}$ is an hypercube, and the extension to general polytopes of $\mathbb{R}^d$ does not raise any particular difficulty. A more advanced and more efficient version is proposed in (Lubachevsky and Stillinger, 1990; Lubachevsky, 1991). The choice of $r(0)$ is not critical, one may take $r(0) = 0$ (there are few collisions when r is small, and therefore r grows quickly during the initial iterations). The initial velocities can be taken i.i.d. with the uniform distribution in $[0,1]^d$, the initial design $X_n(0)$ is arbitrary (one must simply ensure that all initial balls $\mathscr{B}(\mathbf{x}_i(0), r(0))$ do not overlap and lie in the interior of $\mathscr{X}$). Large values of s give fast convergence to a local (jamming) solution, but small s favour convergence to a better solution (with large final radius $r(\infty)$). Different stopping rules may be used. For instance, one may stop the algorithm at iteration k such that

$$k > K \text{ and } \max_{k-K \leq j \leq k-1} \max_i \|\mathbf{x}_i(t_{j+1}) - \mathbf{x}_i(t_j)\| < \varepsilon\,, \tag{18}$$

for some $\varepsilon > 0$ and a large enough window K ($K = 10n$, say). The method is quite efficient for generating fairly good designs for small d, also when n is large (several thousands in Lubachevsky and Stillinger, 1990; Lubachevsky, 1991), and eventually optimal packings when combined with a multistart procedure.

4.2.2. Computational geometry and exchange algorithm

When d is small ($d = 2$ or 3, say), design constructions can exploit powerful tools from computational geometry, such as Delaunay triangulation and Voronoï tessellation (see Section 2.4). For

 ISSN: 2102-6238

instance, Persson and Strang (2004) present a mesh generator based on Delaunay triangulation, with repulsion forces between mesh points (exercised along the edges of Delaunay simplices) and repulsion forces acting normally to the boundary of $\mathscr{X}$ (just as large to prevent the nodes from moving outside $\mathscr{X}$). The simulation of the corresponding dynamical system typically generates meshes of high quality, with nodes that present the characteristics of a good space-filling design. The method can be used for sets $\mathscr{X}$ or arbitrary shape (not necessarily convex, possibly with holes).

The arguments used in Section 2.2 to prove (3) motivate the following algorithm, of the exchange-type (see Fedorov 1972).

0) Select an initial design $X_n^{(1)}$ (e.g., generate $X_n^{(1)}$ with the algorithm of Section 4.1), set $k=1$.
1) Construct a random permutation $\mathscr{I}_k$ of the index set $\{1,\dots,n\}$, set $X'_n = X_n^{(k)}$.
   For $i=1,\dots,n$, do
   — find a point $\mathbf{x}^* \in \operatorname{Arg\,max}_{\mathbf{x}\in\mathscr{X}} \min_{\mathbf{x}_\ell\in X'_n,\ \ell\neq j} \|\mathbf{x}-\mathbf{x}_\ell\|$, with j the i-th element of $\mathscr{I}_k$,
   — replace $\mathbf{x}_j$ by $\mathbf{x}^*$ in X'_n.
2) If $\Phi_{Mm}(X'_n) = \Phi_{Mm}(X_n^{(k)})$ stop; otherwise set $X_n^{(k+1)} = X'_n$, $k \leftarrow k+1$, return to step 1.

When $d=2$ or 3, the fastest determination of a $\mathbf{x}^*$ at step 1 is via Voronoï tessellation, see Section 2.4. The MCMC method presented in the same section permits to use the algorithm for larger d. As usual for such exchange algorithms, convergence to a global optimum is not guaranteed. An illustration is presented in Figure 6-right. The same algorithm can also be used when $\mathscr{X}$ is replaced by $\mathscr{X}_Q$ finite, see Marengo and Todeschini (1992). Instead of replacing each $\mathbf{x}_j$ at step 1, one may perform the best replacement only among the n considered, which makes the method more similar to Fedorov's exchange algorithm (Fedorov, 1972).

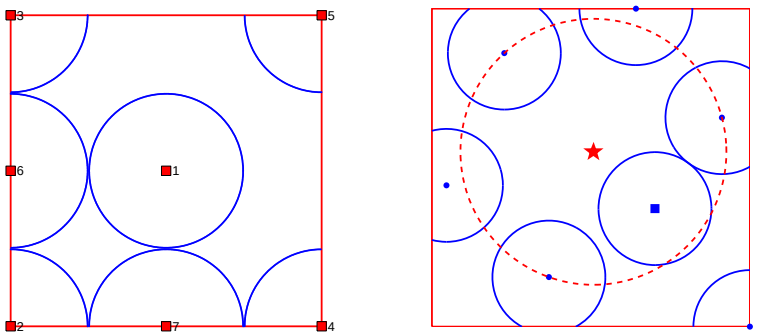

FIGURE 6. *Left: greedy construction of Section 4.1 when $\mathscr{X}$ is a square centered at $\mathbf{x}_1$: the design points are arranged regularly, in the order indicated; the design is clearly sub-optimal for Φ_{mM} and Φ_{Mm} (the circles have radius $\Phi_{Mm}(X_7)/2$) but it nevertheless satisfies (17). Right: illustration of the algorithm of Section 4.2.2 when $\mathscr{X}$ is a square: at iteration k the design point $\mathbf{x}_j$ (square) is replaced by $\mathbf{x}^*$ (star), equidistant from the three design points on the circle in dashed line (and maximising $d(\mathbf{x}, X_n^{(k)} \setminus \{\mathbf{x}_j\})$, $\mathbf{x}\in\mathscr{X}$).*

 ISSN: 2102-6238

4.3. *Minimax optimal design*

4.3.1. *Initialisation by shrinkage of a maximin-optimal design*

The criterion Φ_{Mm} is easier to optimise (and to evaluate) than Φ_{mM}. When a maximin-optimal (or almost optimal) design X_n is available, one may try to improve its minimax performance (while, on the other hand, suffering a reduction of $\Phi_{Mm}(X_n)$) by shrinking X_n towards the centre of $\mathscr{X}$. For instance, when $\mathscr{X} = [0,1]^d$, we can consider the shrunk design $X_n(\delta) = \{\mathbf{x}_1(\delta), \ldots, \mathbf{x}_n(\delta)\}$, with $\mathbf{x}_i(\delta) = \mathbf{c} + (\mathbf{x}_i - \mathbf{c})/(1+\delta)$ for all i and $\mathbf{c} = \mathbf{1}/2$ being the centre of $\mathscr{X}$, and choose a design $X_n(\delta^*)$ by minimising $\phi(\delta) = \Phi_{mM}[X_n(\delta)]$ with respect to $\delta \geq 0$. Note that the function $\delta \mapsto \phi(\delta)$ may have local minima. The design $X_n(\delta^*)$ can be used in particular to initialise the algorithms of Sections 3.2.2 and 4.3.3.

4.3.2. *Exchange algorithm for grid-restricted relaxed minimax optimal design*

When $\mathscr{X}$ is replaced by a finite set $\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$, a regular grid for instance, for $X_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ one may approximate $\tilde{\Phi}_{mM}(X_n) = \max_{\mathscr{X} \in \mathscr{X}_Q} d(\mathbf{x}, X_n)$ by

$$\Phi_p(X_n) = \left[\frac{1}{Q}\sum_{k=1}^{Q} d^p(\mathbf{x}^{(k)}, X_n)\right]^{1/p} \quad \text{or} \quad \Phi_{p,q}(X_n) = \left[\frac{1}{Q}\sum_{k=1}^{Q}\left(\sum_{i=1}^{n}\|\mathbf{x}^{(k)} - \mathbf{x}_i\|^{-q}\right)^{-p/q}\right]^{1/p},$$

with $p, q > 0$, which satisfy

$$\frac{1}{Q^{1/p}}\tilde{\Phi}_{mM}(X_n) \leq \Phi_p(X_n) \leq \tilde{\Phi}_{mM}(X_n) \quad \text{and} \quad \frac{1}{n^{1/q}\,Q^{1/p}}\tilde{\Phi}_{mM}(X_n) \leq \Phi_{p,q}(X_n) \leq \tilde{\Phi}_{mM}(X_n)$$

and therefore $\lim_{p\to\infty}\Phi_p(X_n) = \tilde{\Phi}_{mM}(X_n)$ and $\lim_{p,q\to\infty}\Phi_{p,q}(X_n) = \tilde{\Phi}_{mM}(X_n)$. One may then apply an exchange algorithm, similar to that in Section 4.2.2, to the minimisation of $\Phi_p(\cdot)$ or $\Phi_{p,q}(\cdot)$, see Royle and Nychka (1998).

4.3.3. `kmeans` *and centroids*

Let $\mathscr{T}_n = \{\mathscr{C}_i, i = 1, \ldots, n\}$ form a tessellation of $\mathscr{X}$ and let X_n be an arbitrary set of n points in $\mathscr{X}$. The functional $(\mathscr{T}_n, X_n) \mapsto \mathscr{E}_2(\mathscr{T}_n, X_n) = \sum_{i=1}^{n}\int_{\mathscr{C}_i}\|\mathbf{x} - \mathbf{x}_i\|^2\,\mathrm{d}\mathbf{x}$ is minimised when the $\mathscr{C}_i$ are the Voronoï regions defined by the generators $\mathbf{x}_i$, and simultaneously the $\mathbf{x}_i$ are the centroids of the corresponding $\mathscr{C}_i$, that is, $\mathbf{x}_i = (\int_{\mathscr{C}_i}\mathbf{x}\,\mathrm{d}\mathbf{x})/\mathrm{vol}(\mathscr{C}_i)$, the condition being necessary but not sufficient, see Flury (1993); Du et al. (1999). These centroids, called principal points in (Flury, 1993), form a design X_n that should thus perform reasonably in terms of space-filling, see Lekivetz and Jones (2015) for another type of clustering-based design that uses Ward's minimum-variance criterion (Ward Jr., 1963) for a discretised version of $\mathscr{X}$. The randomised `kmeans` method presented below does not require the explicit construction of a Voronoï tessellation, see MacQueen (1967).

0) Select $X_n^{(1)}$, set $k = 1$ and $n_i = 1$ for all $i = 1, \ldots, n$.
1) Select $\mathbf{x}$ at random (uniformly) in $\mathscr{X}$.

 ISSN: 2102-6238

2) Find $i^* = \arg\min_{i=1,\dots,n} \|\mathbf{x}_i - \mathbf{x}\|$, $\mathbf{x}_i \in X_n^{(k)}$; substitute $(n_{i^*}\mathbf{x}_{i^*} + \mathbf{x})/(n_{i^*}+1)$ for $\mathbf{x}_i$ in $X_n^{(k)}$, $n_{i^*} \leftarrow n_{i^*}+1$.
3) If some convergence criterion is met, such as (18) for some $\varepsilon > 0$ and $K > 1$, then stop; otherwise $k \leftarrow k+1$, and return to step 1.

A more popular version is Lloyd's method (1982), which corresponds to fixed-point iterations for the mapping $T(\cdot)$ defined by $X_n^{(k)} \mapsto X_n^{(k+1)} = T(X_n^{(k)}) = \{T_1(\mathbf{x}_1^{(k)}), \dots, T_n(\mathbf{x}_n^{(k)})\}$, with $T_i(\mathbf{x}_i) = (\int_{\mathscr{C}_i(X_n)} \mathbf{x}\,\mathrm{d}\mathbf{x})/\mathsf{vol}[\mathscr{C}_i(X_n)]$, where the $\mathscr{C}_i(X_n)$ form a Voronoï tessellation of $\mathscr{X}$ with the $\mathbf{x}_i$ in X_n as generators.

The $\mathscr{C}_i(X_n^{(k)})$ can be computed exactly when d is small enough, see Section 2.4. Constructing Voronoï tessellations for a discretised version $\mathscr{X}_Q$ of $\mathscr{X}$ can be considered too, and the corresponding `kmeans` algorithm is available in standard statistical toolboxes. Note that it converges to what Flury (1993) calls self-consistent points, but not necessarily to a design X_n (and a Voronoï tessellation $\mathscr{T}_n$) minimising $\mathscr{E}_2$. Since minimax-optimal design is related to the construction of a centroidal tessellation for the energy functional $\mathscr{E}_q(\mathscr{T}_n, X_n) = \sum_{i=1}^n \int_{\mathscr{C}_i} \|\mathbf{x} - \mathbf{x}_i\|^q \,\mathrm{d}\mathbf{x}$ when $q \to \infty$, we propose the following modified version of Lloyd's algorithm; see also Cortés and Bullo (2005, 2009, Sect. 5.3).

0) Select $X_n^{(1)}$, set $k = 1$.
1) Compute the Voronoï tessellation $\{\mathscr{C}_i, i = 1, \dots, n\}$ of $\mathscr{X}$ (or $\mathscr{X}_Q$) corresponding to the generators in $X_n^{(k)}$.
2) For $i = 1, \dots, n$ determine the smallest ball $\mathscr{B}(\mathbf{c}_i, r_i)$ enclosing $\mathscr{C}_i$, replace $\mathbf{x}_i$ by $\mathbf{c}_i$ in $X_n^{(k)}$.
3) If some convergence criterion is met, then stop; otherwise $k \leftarrow k+1$, return to step 1.

Note that $\Phi_{mM}(X_n^{(k)})$ (or $\tilde{\Phi}_{mM}(X_n) = \max_{\mathbf{x} \in \mathscr{X}_Q} d(\mathbf{x}, X_n)$) decreases monotonically along iterations, so that one can use the stopping rule $\Phi_{mM}(X_n^{(k)}) - \Phi_{mM}(X_n^{(k+1)}) < \varepsilon \ll 1$ at step 3.

When the Voronoï tessellation is computed for a finite set $\mathscr{X}_Q$, the $\mathscr{B}(\mathbf{c}_i, r_i)$ of step 2 are the smallest balls containing a finite set of points. This is also the case when $\mathscr{X}$ is a polytope of $\mathbb{R}^d$, since the $\mathscr{B}(\mathbf{c}_i, r_i)$ should simply contain the vertices of the polytopic cells $\mathscr{C}_i$. The determination of the smallest ball containing a given set of points of $\mathbb{R}^d$ is a classical optimisation problem (its centre is called the Chebyshev centre of the set), for which many algorithms have been proposed in the literature, see, e.g., Botkin and Turova-Botkina (1994), Xu et al. (2003), Yildirim (2008). In fact, exploiting the equivalence with a convex Quadratic-Programming (QP) problem in $\mathbb{R}^{d+1}$, any standard QP solver gives the solution (exactly in finite time for some solvers). Indeed, denote by $\mathbf{z}_i \in \mathbb{R}^d$ the m points considered, the smallest enclosing ball $\mathscr{B}^* = \mathscr{B}(\mathbf{c}^*, r^*)$ is obtained by minimising $f(\mathbf{c}) = \max_{i=1,\dots,m} \|\mathbf{z}_i - \mathbf{c}\|^2$ with respect to $\mathbf{c} \in \mathbb{R}^d$. For any $\mathbf{c}_0$ in $\mathbb{R}^d$, we can write $\|\mathbf{z}_i - \mathbf{c}\|^2 = \|\mathbf{z}_i - \mathbf{c}_0\|^2 - 2(\mathbf{z}_i - \mathbf{c}_0)^\top(\mathbf{c} - \mathbf{c}_0) + \|\mathbf{c} - \mathbf{c}_0\|^2$. Therefore, $f(\mathbf{c}) = \max_{i=1,\dots,m} \left\{\|\mathbf{z}_i - \mathbf{c}_0\|^2 - 2(\mathbf{z}_i - \mathbf{c}_0)^\top(\mathbf{c} - \mathbf{c}_0)\right\} + \|\mathbf{c} - \mathbf{c}_0\|^2$, and the minimisation of $f(\mathbf{c})$ is equivalent to the minimisation of $\|\mathbf{c} - \mathbf{c}_0\|^2 + t$, subject to the m linear constraints $\|\mathbf{z}_i - \mathbf{c}_0\|^2 - 2(\mathbf{z}_i - \mathbf{c}_0)^\top(\mathbf{c} - \mathbf{c}_0)\} \le t$, $i = 1, \dots, m$, with respect to $(\mathbf{c}, t) \in \mathbb{R}^{d+1}$ (when the QP solver requires a strictly convex problem, one may add a small regularisation term, quadratic in t, to the objective function, $10^{-12}t^2$ say).

When m is large, it is convenient to remove the $\mathbf{z}_i$ that will lie for sure in the interior of $\mathscr{B}^*$. From Lagrangian theory, the dual of the smallest-ball problem is the following: maximise $\phi(\mathbf{w}) = \sum_{i=1}^m w_i \|\mathbf{z}_i - \mathbf{c}(\mathbf{w})\|^2$, where $\mathbf{c}(\mathbf{w}) = \sum_{i=1}^m \mathbf{z}_i$ and $\mathbf{w} = (w_1, \dots, w_m)$ with $w_i > 0$ and $\sum_{i=1}^m w_i = 1$. This dual problem presents some ressemblance with an experimental design problem (approximate

 ISSN: 2102-6238

theory), see Pronzato et al. (2016a). For any weights $\mathbf{w}$, denote $\varepsilon = \varepsilon(\mathbf{w}) = \max_{i=1,\ldots,m} \|\mathbf{z}_i - \mathbf{c}(\mathbf{w})\|^2 - \phi(\mathbf{w})$. Following an approach similar to Harman and Pronzato (2007), we can show that $\|\mathbf{c}(\mathbf{w}) - \mathbf{c}^*\| \leq \sqrt{\varepsilon/2}$ and that any $\mathbf{z}_i$ such that $\|\mathbf{z}_i - \mathbf{c}(\mathbf{w})\|^2 < \phi(\mathbf{w}) + \varepsilon - \sqrt{\varepsilon(2\phi(\mathbf{w}) + \varepsilon)}$ necessarily lie in the interior of $\mathscr{B}^*$, and therefore do not need to be considered further. A natural choice for filtering the $\mathbf{z}_i$ before using the QP solver is $w_i = 1/m$ for all i, but the initial $\mathbf{w}$ of (Yildirim, 2008), defined by $w_{i_1} = w_{i_2} = 1/2$ and $w_i = 0$ for all $i \neq i_1, i_2$, with $i_1 = \arg\max_{i=1,\ldots,m} \|z_i - z_1\|$ and $i_2 = \arg\max_{i=1,\ldots,m} \|z_i - z_{i_1}\|$, is generally more efficient. Also, choosing $\mathbf{c}_0$ out of the convex hull of the $\mathbf{z}_i$ generally yields a faster solution (one may take for instance $\mathbf{c}_0 = 2\mathbf{z}_{i_a} - \mathbf{z}_{i_b}$ with $i_a = \arg\max_{i=1,\ldots,m}\{\mathbf{z}_i\}_1$ and $i_b = \arg\min_{i=1,\ldots,m}\{\mathbf{z}_i\}_1$).

Although there is no guaranteed convergence to a minimax-optimal design, for small enough d this centroid method, possibly combined with a multistart procedure, is rather efficient for generating designs with good minimax performance at reasonable computational cost.

4.3.4. Bayesian optimal design and Karhunen-Loève decomposition of a random-field model

The construction presented here is based on the developments in (Gauthier and Pronzato, 2014, 2016), following the original ideas of Fedorov (1996), and is motivated by the connection between minimax-optimal design and minimisation of the maximum kriging variance, see Section 2.5.2. As in Section 2.1, we consider a kernel K defining a RKHS $\mathscr{H}$, and suppose that f corresponds to the realisation of a centred random field $Z_{\mathbf{x}}$ with covariance $\mathsf{E}\{Z_{\mathbf{x}} Z_{\mathbf{x}'}\} = K(\mathbf{x}, \mathbf{x}')$. We suppose that K is continuous on $\mathscr{X} \times \mathscr{X}$. Let μ denote a measure on $\mathscr{X}$, typically a measure of interest for the definition of an Integrated Mean Squared Error $\int_{\mathscr{X}} \rho_n^2(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$ on $\mathscr{X}$. The computations are facilitated when μ is supported on a discretised version $\mathscr{X}_Q$ of $\mathscr{X}$, see Gauthier and Pronzato (2014); in particular μ may be uniform on $\mathscr{X}_Q$. Consider the linear operator T_μ on $L^2(\mathscr{X}, \mu)$ defined by $T_\mu[f](\mathbf{x}) = \int_{\mathscr{X}} f(\mathbf{y}) K(\mathbf{x}, \mathbf{y}) \mathrm{d}\mu(\mathbf{y})$, $f \in L^2(\mathscr{X}, \mu)$, $\mathbf{x} \in \mathscr{X}$, and respectively denote by λ_j and φ_j the eigenvalues and eigenfunctions of its Mercer decomposition. We assume that the λ_j are ordered by decreasing values and that the φ_j are orthonormal in $L^2(\mathscr{X}, \mu)$. The Karhunen-Loève decomposition of $Z_{\mathbf{x}}$ corresponds to $\sum_{k:\lambda_k > 0} \sqrt{\lambda_k} \zeta_k \varphi_k(\mathbf{x})$, where the random variables ζ_k are orthogonal, centred, with variance 1. A truncation to the m largest λ_j yields a linear regression model, $Z_{\mathbf{x}} = \sum_{j=1}^m \beta_j \varphi_j(\mathbf{x}) + \varepsilon(\mathbf{x}) = \phi^\top(\mathbf{x})\beta + \varepsilon(\mathbf{x})$, with correlated errors ε satisfying $\mathsf{E}\{\varepsilon(\mathbf{x})\} = 0$, $\mathsf{E}\{\varepsilon(\mathbf{x})\varepsilon(\mathbf{x}')\} = K(\mathbf{x}, \mathbf{x}') - \sum_{j=1}^m \lambda_j \varphi_j(\mathbf{x}) \varphi_j(\mathbf{x}')$, and parameters β_j having a prior distribution with zero mean and covariance $\Lambda_m = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$. To facilitate the construction of optimal designs, we shall neglect the effect of correlations, and consider an approximate model $Z'_{\mathbf{x}} = \sum_{j=1}^m \beta_j \varphi_j(\mathbf{x}) + \varepsilon'(\mathbf{x})$, where now $\mathsf{E}\{\varepsilon'(\mathbf{x})\varepsilon'(\mathbf{x}')\} = 0$ for $\mathbf{x} \neq \mathbf{x}'$. An heteroscedastic model with $\mathsf{E}\{\varepsilon'^2(\mathbf{x})\} = K(\mathbf{x}, \mathbf{x}) - \sum_{j=1}^m \lambda_j \varphi_j^2(\mathbf{x})$ is used in (Gauthier and Pronzato, 2016), but we consider here a model with constant variance $\mathsf{E}\{\varepsilon'^2(\mathbf{x})\} = \sigma^2$ for all $\mathbf{x}$ (for instance $\sigma^2 = [1/\mu(\mathscr{X})] \int_{\mathscr{X}} [K(\mathbf{x}, \mathbf{x}) - \sum_{j=1}^m \lambda_j \varphi_j^2(\mathbf{x})] \mathrm{d}\mu(\mathbf{x})$). We can then easily compute an optimal design for this approximate model. Indeed, the regularised LS estimator $\hat{\beta}$ that minimises $\sigma^{-2}(\mathbf{z}_n - \mathbf{P}_m\beta)^\top \mathbf{I}_n (\mathbf{z}_n - \mathbf{P}_m\beta) + \beta^\top \Lambda_m^{-1} \beta$, with $\{\mathbf{P}_m\}_{ij} = \varphi_j(\mathbf{x}_i)$, has covariance $\mathbf{C}_m = (\mathbf{M}_n + \Lambda_m^{-1})^{-1}$, where $\mathbf{M}_n = (1/\sigma^2) \sum_{i=1}^n \phi(\mathbf{x}_i) \phi^\top(\mathbf{x}_i)$, and an IMSE-optimal design minimises $\int_{\mathscr{X}} \phi^\top(\mathbf{x}) \mathbf{C}_m \phi(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) = \mathrm{trace}[(\mathbf{M}_n + \Lambda_m^{-1})^{-1}]$. One may then use all the machinery of approximate design theory, i.e., consider information matrices $\mathbf{M}(\xi) = (1/\sigma^2) \int_{\mathscr{X}} \phi(\mathbf{x}) \phi^\top(\mathbf{x}) \mathrm{d}\xi(\mathbf{x})$ defined for probability measures ξ on $\mathscr{X}$ (called design measures), and determine an opti-

 ISSN: 2102-6238

mal design measure that minimises the (convex) Bayesian A-optimality criterion $\xi \mapsto \Phi(\xi) = \text{trace}[(\mathbf{M}(\xi) + \Lambda_m^{-1}/n)^{-1}]$, see Pilz (1983), Fedorov (1996), Spöck and Pilz (2010), Gauthier and Pronzato (2016). Similarly, a Bayesian D-optimal design measure ξ^* maximises the concave function $\xi \mapsto \det(\mathbf{M}(\xi) + \Lambda_m^{-1}/n)$, an optimisation problem for which many algorithms are available, see, e.g., (Pronzato and Pázman, 2013, Chap. 9). Concavity implies that the directional derivative at ξ^* in the direction of any other probability measure ν on $\mathscr{X}$ is non-positive, from which we obtain

$$\max_{\mathbf{x}\in\mathscr{X}} \boldsymbol{\phi}^\top(\mathbf{x})(\mathbf{M}(\xi^*) + \Lambda_m^{-1}/n)^{-1}\boldsymbol{\phi}^\top(\mathbf{x}) \le \sigma^2 \left\{m - \text{trace}[\Lambda_m^{-1}(n\mathbf{M}(\xi^*) + \Lambda_m^{-1})^{-1}]\right\},$$

which sets an upper bound on the maximum mean-squared prediction error for the optimal design ξ^*. Numerical experiments show that optimal design measures with $m \approx n$ often concentrate their mass around $n' \approx n$ points, and that the designs $X_{n'}$ corresponding to those n' support points are well spread over $\mathscr{X}$, with good minimax performance when $K(\mathbf{x},\mathbf{x}') = C(\|\mathbf{x}-\mathbf{x}'\|)$ with $C(t)$ decreasing fast enough with t. Further investigations are required to quantify this performance more precisely. The procedure described in (Gauthier and Pronzato, 2016) can be used to construct a design of given size n.

5. Conclusions and perspectives

We have listed a few properties of minimax and maximin optimal designs, including connections with other space-filling approaches, and indicated several methods which, in principle, can be used for the construction of optimal designs. The high complexity of the optimisation problems involved means that, in general, one should be satisfied with a suboptimal design, or even with a design that simply performs "not too badly" in terms of Φ_{Mm} or Φ_{mM}. In this respect, simulated annealing (Section 3.1) or the specific methods of Section 4.2 perform reasonably well. However, the dimension of the optimisation problem is $n \times d$, which may quickly become prohibitive. A reasonable solution can usually be reached when n is large with d small, but the case when d is large (say, $d \approx 10$) is much more problematic, especially for minimax-optimal design. Indeed, methods based on computational geometry are then inefficient, and `kmeans` and centroids suppose a discretisation of $\mathscr{X}$ which may provide inaccurate results for large d. Combining the MCMC method proposed in Section 2.4 for the evaluation of Φ_{mM} with an optimisation algorithm is a challenging and motivating issue. Finally, as explained in Section 2.3, minimax and maximin optimal designs are not adapted to problems with very large d, and a few suggestions have been made in the same section on possible extensions and topics for further investigation.

Acknowledgements

This work is based on lectures given at the spring schools ASPEN'2014 ("Analyse de Sensibilité, Propagation d'incertitudes et Exploration Numérique de modèles en sciences de l'environnement", École de Physique, Les Houches, France, May 4-9 2014) and PECNUM'2015 ("Plans d'Expériences Classiques et NUMériques", École Centrale de Lyon, France, May 18-22 2015). I warmly thank the organisers of the two events (Clémentine Prieur, Hervé Monod, Robert Faivre and Laurence Viry for the former, and Céline Helbert, François Wahl, Robert Faivre and Alexandre Janon for

 ISSN: 2102-6238

the latter) who gave me the opportunity for preparing this material. The paper benefited from several fruitful discussions with Maria-João Rendas (Université Côte d'Azur, CNRS, Laboratoire I3S) and Anatoly Zhigljavsky (Cardiff University). I also thank the referees and invited editors for their constructive comments that helped to improve the presentation.

References

Aaron, C., Cholaquidis, A., and Fraiman, R. (2014). On the maximal multivariate spacing extension and convexity tests. *arXiv preprint arXiv:1411.2482*.

Atkinson, A., Donev, A., and Tobias, R. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press.

Audze, P. and Eglais, V. (1977). New approach for planning out of experiments. *Problems of Dynamics and Strengths*, 35:104–107.

Auffray, Y., Barbillon, P., and Marin, J.-M. (2012). Maximin design on non hypercube domains and kernel interpolation. *Statistics and Computing*, 22(3):703–712.

Basseville, M. (2013). Divergence measures for statistical data processing — An annotated bibliography. *Signal Processing*, 93(4):621–633.

Bates, R., Buck, R., Riccomagno, E., and Wynn, H. (1996). Experimental design and observation for large systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 77–94.

Beardwood, J., Halton, J., and Hammersley, J. (1959). The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55(4):299–327.

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, Boston.

Bertsimas, D. and Tsitsiklis, J. (1993). Simulated annealing. *Statistical Science*, 8(1):10–15.

Biedermann, S. and Dette, H. (2001). Minimax optimal designs for nonparametric regression — a further optimality property of the uniform distribution. In A.C. Atkinson, P. H. and Müller, W., editors, *mODa'6 – Advances in Model–Oriented Design and Analysis, Proceedings of the 76th Int. Workshop, Puchberg/Schneeberg (Austria)*, pages 13–20, Heidelberg. Physica Verlag.

Blum, A., Hopcroft, J., and Kannan, R. (2016). *Foundations of Data Science*. http://www.cs.cornell.edu/jeh/bookJan25_2016.pdf.

Boissonnat, J.-D. and Yvinec, M. (1998). *Algorithmic Geometry*. Cambridge University Press.

Bonnans, J., Gilbert, J., Lemaréchal, C., and Sagastizábal, C. (2006). *Numerical Optimization. Theoretical and Practical Aspects*. Springer, Heidelberg. [2nd ed.].

Botkin, N. and Turova-Botkina, V. (1994). An algorithm for finding the Chebyshev center of a convex polyhedron. *Applied Mathematics and Optimization*, 29:211–222.

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

Conway, J. and Sloane, N. (1999). *Sphere Packings, Lattices and Groups*. Springer, New York. [3rd ed.].

Cortés, J. and Bullo, F. (2005). Coordination and geometric optimization via distributed dynamical systems. *SIAM Journal on Control and Optimization*, 44(5):1543–1574.

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

Croft, H., Falconer, K., and Guy, R. (1991). *Unsolved Problems in Geometry*. Springer, New York.

Dem'yanov, V. and Malozemov, V. (1974). *Introduction to Minimax*. Dover, New York.

Du, Q., Faber, V., and Gunzburger, M. (1999). Centroidal Voronoi tessellations: applications and algorithms. *SIAM Review*, 41(4):637–676.

Durrande, N., Ginsbourger, D., and Roustant, O. (2012). Additive covariance kernels for high-dimensional Gaussian process modeling. *Annales de la Faculté de Sciences de Toulouse*, 21(3):481–499.

Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, Boca Raton.

Fang, K.-T., Lu, X., and Winker, P. (2003). Lower bounds for centered and wrap-around L_2-discrepancies and construction of uniform designs by threshold accepting. *Journal of Complexity*, 19(5):692–711.

Fang, K.-T. and Ma, C.-X. (2001). Wrap-around L_2-discrepancy of random sampling, Latin hypercube and uniform designs. *Journal of Complexity*, 17(4):608–624.

 ISSN: 2102-6238

Fang, K.-T., Tang, Y., and Yin, J. (2005). Lower bounds for wrap-around L_2-discrepancy and constructions of symmetrical uniform designs. *Journal of Complexity*, 21(5):757–771.

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

Fedorov, V. (1996). Design of spatial experiments: model fitting and prediction. In Gosh, S. and Rao, C., editors, *Handbook of Statistics, vol. 13*, chapter 16, pages 515–553. Elsevier, Amsterdam.

Fedorov, V. and Leonov, S. (2014). *Optimal Design for Nonlinear Response Models*. CRC Press, Boca Raton.

Flury, B. (1993). Estimation of principal points. *Applied Statistics*, 42(1):139–151.

Franco, J., Vasseur, O., Corre, B., and Sergent, M. (2009). Minimum Spanning Tree: a new approach to assess the quality of the design of computer experiments. *Chemometrics and Intelligent Laboratory Systems*, 97:164–169.

Gauthier, B. and Pronzato, L. (2014). Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models. *SIAM/ASA J. Uncertainty Quantification*, 2:805–825. DOI 10.1137/130928534.

Gauthier, B. and Pronzato, L. (2016). Convex relaxation for IMSE optimal design in random field models. *Computational Statistics and Data Analysis (to appear)*. hal-01246483.

Ginsbourger, D., Roustant, O., Schuhmacher, D., Durrande, N., and Lenz, N. (2014). On ANOVA decompositions of kernels and Gaussian random field paths. *preprint arXiv:1409.6008*.

Gonzalez, T. (1985). Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306.

Graham, R. and Lubachevsky, B. (1996). Repeated patterns of dense packings of equal disks in a square. *The Electronic Journal of Combinatorics*, 3(1):R16.

Graham, R., Lubachevsky, B., Nurmela, K., and Östergård, P. (1998). Dense packings of congruent circles in a circle. *Discrete Mathematics*, 181(1):139–154.

Guyader, A., Hengartner, N., and Matzner-Løber, E. (2011). Simulation and estimation of extreme quantiles and extreme probabilities. *Applied Mathematics & Optimization*, 64(2):171–196.

Hardin, D. and Saff, E. (2004). Discretizing manifolds via minimum energy points. *Notices of the AMS*, 51(10):1186–1194.

Harman, R. and Pronzato, L. (2007). Improvements on removing non-optimal support points in *D*-optimum design algorithms. *Statistics & Probability Letters*, 77:90–94.

Hickernell, F. (1998a). A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 67(221):299–322.

Hickernell, F. (1998b). Lattice rules: how well do they measure up? In Helekalek, P. and Larcher, G., editors, *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statist.*, pages 109–166. Springer, New York.

Janson, S. (1986). Random coverings in several dimensions. *Acta Mathematica*, 156(1):83–118.

Janson, S. (1987). Maximal spacings in several dimensions. *The Annals of Probability*, 15(1):274–280.

Johnson, M., Moore, L., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26:131–148.

Joseph, V., Gul, E., and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, 102(2):371–380.

Jourdan, A. and Franco, J. (2009a). A new criterion based on Kullback-Leibler information for space filling designs. preprint arXiv:0904.2456, hal-00375820.

Jourdan, A. and Franco, J. (2009b). Plans d'expériences numériques d'information de Kullback-Leibler minimale. *Journal de la Société Française de Statistique*, 150(2):52–64.

Jourdan, A. and Franco, J. (2010). Optimal Latin hypercube designs for the Kullback–Leibler criterion. *AStA Advances in Statistical Analysis*, 94(4):341–351.

Kennard, R. and Stone, L. (1969). Computer aided design of experiments. *Technometrics*, 11(1):137–148.

Kesten, H. (1958). Accelerated stochastic approximation. *The Annals of Mathematical Statistics*, 29(1):41–59.

Klee, V. (1980). On the complexity of *d*-dimensional Voronoi diagrams. *Archiv der Mathematik*, 34(1):75–80.

Korobov, N. (1960). Properties and calculation of optimal coefficients. *Doklady Akademii Nauk SSSR*, 132(5):1009–1012.

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

Lagarias, J., editor (2011). *The Kepler Conjecture: The Hales-Ferguson Proof*. Springer, New York.

Landkof, N. (1972). *Foundations of Modern Potential Theory*. Springer, Berlin.

Lekivetz, R. and Jones, B. (2015). Fast flexible space-filling designs for nonrectangular regions. *Quality and Reliability*

 ISSN: 2102-6238

Engineering International, 31(5):829–837.
Leonenko, N., Pronzato, L., and Savani, V. (2008). A class of Rényi information estimators for multidimensional densities. *Annals of Statistics*, 36(5):2153–2182 (Correction in AS, 38(6):3837–3838, 2010).
Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.
Lubachevsky, B. (1991). How to simulate billiards and similar systems. *Journal of Computational Physics*, 94(2):255–283.
Lubachevsky, B. and Stillinger, F. (1990). Geometric properties of random disk packings. *Journal of Statistical Physics*, 60(5-6):561–583.
MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In LeCam, L. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297.
Maranas, C., Floudas, C., and Pardalos, P. (1995). New results in the packing of equal circles in a square. *Discrete Mathematics*, 142(1):287–293.
Marengo, E. and Todeschini, R. (1992). A new algorithm for optimal, distance-based, experimental design. *Chemometrics and Intelligent Laboratory Systems*, 16:37–44.
Markót, M. and Csendes, T. (2005). A new verified optimization technique for the "packing circles in a unit square" problems. *SIAM Journal on Optimization*, 16(1):193–219.
Mitchell, R. (1990). Error estimates arising from certain pseudorandom sequences in a quasirandom search method. *Mathematics of Computation*, 55(191):289–297.
Müller, W. (2001). Coffee-house designs. In Atkinson, A., Bogacka, B., and Zhigljavsky, A., editors, *Optimum Design 2000*, chapter 21, pages 241–248. Kluwer, Dordrecht.
Müller, W. (2007). *Collecting Spatial Data*. Springer, Berlin. [3rd ed.].
Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia.
Nurmela, K. and Östergård, P. (1997). Packing up to 50 equal circles in a square. *Discrete & Computational Geometry*, 18(1):111–120.
Nurmela, K. and Östergård, P. (1999). More optimal packings of equal circles in a square. *Discrete & Computational Geometry*, 22(3):439–457.
Nuyens, D. (2007). Fast construction of good lattice rules. Ph.D. Thesis, Katholieke Univ. Leuven.
Okabe, A., Books, B., Sugihama, K., and Chiu, S. (1992). *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Wiley, New York.
Oler, N. (1961). A finite packing problem. *Canadian Mathematical Bulletin*, 4:153–155.
Penrose, M. and Yukich, J. (2003). Weak laws of large numbers in geometric probability. *The Annals of Applied Probability*, 13(1):277–303.
Penrose, M. and Yukich, J. (2011). Laws of large numbers and nearest neighbor distances. In Wells, M. and Sengupta, A., editors, *Advances in Directional and Linear Statistics. A Festschrift for Sreenivasa Rao Jammalamadaka*, pages 189–199. Springer.
Persson, P.-O. and Strang, G. (2004). A simple mesh generator in MATLAB. *SIAM review*, 46(2):329–345.
Pilz, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models*, volume 55. Teubner-Texte zur Mathematik, Leipzig. (also Wiley, New York, 1991).
Pronzato, L. and Müller, W. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22:681–701.
Pronzato, L. and Pázman, A. (2013). *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Springer, LNS 212, New York.
Pronzato, L., Wynn, H., and Zhigljavsky, A. (2016a). Extended generalised variances, with applications. *Bernoulli (to appear)*. arXiv preprint arXiv:1411.6428, hal-01308092.
Pronzato, L., Wynn, H., and Zhigljavsky, A. (2016b). Extremal measures maximizing functionals based on simplicial volumes. *Statistical Papers (to appear)*. hal-01308116.
Pukelsheim, F. (1993). *Optimal Experimental Design*. Wiley, New York.
Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge (MA), USA.
Redmond, C. and Yukich, J. (1994). Limit theorems and rates of convergence for euclidean functionals. *The Annals of Applied Probability*, 4(4):1057–1073.
Redmond, C. and Yukich, J. (1996). Asymptotics for Euclidian functionals with power-weighted edges. *Stochastic Processes and their Applications*, 61:289–304.

 ISSN: 2102-6238

Riccomagno, E., Schwabe, R., and Wynn, H. (1997). Lattice-based D-optimum design for Fourier regression. *Annals of Statistics*, 25(6):2313–2327.
Roberts, G. and Rosenthal, J. (2006). Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability*, 6(4):2123–2139.
Royle, J. and Nychka, D. (1998). An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers and Geosciences*, 24(5):479–488.
Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435.
Saff, E. (2010). Logarithmic potential theory with applications to approximation theory. *Surveys in Approximation Theory*, 5(14):165–200.
Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, Heidelberg.
Schaback, R. (1995). Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3(3):251–264.
Shewry, M. and Wynn, H. (1987). Maximum entropy sampling. *Applied Statistics*, 14:165–170.
Silvey, S. (1980). *Optimal Design*. Chapman & Hall, London.
Sinai, Y. and Chernov, N. (1987). Ergodic properties of certain systems of two-dimensional discs and three-dimensional balls. *Russian Mathematical Surveys*, 42(3):181–207.
Sloan, I. and Reztsov, A. (2002). Component-by-component construction of good lattice rules. *Mathematics of Computation*, 71(237):263–273.
Sloan, I. and Walsh, L. (1990). A computer search of rank-2 lattice rules for multidimensional quadrature. *Mathematics of Computation*, 54(189):281–302.
Spöck, G. and Pilz, J. (2010). Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment*, 24(3):463–482.
Steele, J. (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *The Annals of Probability*, pages 365–376.
Stein, M. (1999). *Interpolation of Spatial Data. Some Theory for Kriging*. Springer, Heidelberg.
Sukharev, A. (1992). *Minimax Models in the Theory of Numerical Methods*. Springer (originally, Kluwer Academic Publishers), Dordrecht.
Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, 8(1):1–9.
Vallentin, F. (2003). Sphere coverings, lattices and tilings. Ph.D. Thesis, Technischen Universität München, Germany.
Vazquez, E. and Bect, J. (2011). Sequential search based on kriging: convergence analysis of some algorithms. Proc. 58th World Statistics Congress of the ISI, August 21-26, Dublin, Ireland, arXiv preprint arXiv:1111.3866v1.
Viazovska, M. (2016). The sphere packing problem in dimension 8. *arXiv preprint arXiv:1603.04246*.
Wade, A. (2007). Explicit laws of large numbers for random nearest-neighbour-type graphs. *Advances in Applied Probability*, 39(2):326–342.
Wahl, F., Mercadier, C., and Helbert, C. (2014). Measuring the quality of maximin space-filling designs. hal-00955294.
Walter, E. and Piet-Lahanier, H. (1989). Exact recursive polyhedral description of the feasible parameter set for bounded error models. *IEEE Transactions on Automatic Control*, 34:911–915.
Ward Jr., J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.
Xu, S., Freund, R., and Sun, J. (2003). Solution methodologies for the smallest enclosing circle problem. *Computational Optimization and Applications*, 25(1-3):283–292.
Yildirim, E. (2008). Two algorithms for the minimum enclosing ball problem. *SIAM Journal on Optimization*, 19(3):1368–1391.
Yukich, J. (1998). *Probability Theory of Classical Euclidean Optimization Problems*. Springer, Berlin.
Zhigljavsky, A., Dette, H., and Pepelyshev, A. (2010). A new approach to optimal design for linear models with correlated observations. *Journal of the American Statistical Association*, 105(491):1093–1103.
Zhigljavsky, A. and Hamilton, E. (2010). Stopping rules in k-adaptive global random search algorithms. *Journal of Global Optimization*, 48(1):87–97.
Zhigljavsky, A. and Žilinskas, A. (2007). *Stochastic Global Optimization*. Springer, New York.

 ISSN: 2102-6238