HAL
open science

On the elimination of inessential points in the smallest enclosing ball problem

Luc Pronzato

To cite this version:

Luc Pronzato. On the elimination of inessential points in the smallest enclosing ball problem. Optimization Methods and Software, 2019, 34 (2), pp.225-247. 10.1080/10556788.2017.1359266 . hal-01863587

HAL Id: hal-01863587

https://hal.science/hal-01863587

Submitted on 28 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

To appear in *Optimization Methods & Software*
Vol. 00, No. 00, Month 20XX, 1–23

On the elimination of inessential points in the smallest enclosing ball problem

L. Pronzato[a*]

[a] *Université Côte d'Azur, CNRS, Laboratoire I3S, Sophia Antipolis*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

We consider the construction of the smallest ball $\mathscr{B}^*$ enclosing a set $\mathscr{X}_n$ formed by n points in $\mathbb{R}^d$. We show that any probability measure on $\mathscr{X}_n$, with mean c and variance matrix V, provides a lower bound b on the distance to c of any point on the boundary of $\mathscr{B}^*$, with b having a simple expression in terms of c and V. This inequality permits to remove inessential points from $\mathscr{X}_n$, which do not participate to the definition of $\mathscr{B}^*$, and can be used to accelerate algorithms for the construction of $\mathscr{B}^*$. We show that this inequality is, in some sense, the best possible. A series of numerical examples indicates that, when d is reasonably small ($d \leq 10$, say) and n is large (up to 10^5), the elimination of inessential points by a suitable two-point measure, followed by a direct (exact) solution by quadratic programming, outperforms iterative methods that compute an approximate solution by solving the dual problem.

Keywords: minimum enclosing ball; smallest ball; Chebyshev centre; core sets; optimal design of experiments

AMS Subject Classification: 90C25; 90C46; 62K05

1. Introduction

Given a set of n points $\mathscr{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$, $d \geq 2$, we consider the algorithmic construction of the minimum ball $\mathscr{B}^*(\mathscr{X}_n)$ enclosing $\mathscr{X}_n$. We are interested in particular in the situation where d is reasonably small but n can be large. For $c \in \mathbb{R}^d$ and $r \in \mathbb{R}^+$, we denote by $\mathscr{B}_d(c, r)$ the (closed) ball $\{X \in \mathbb{R}^d : \|X - c\| \leq r\}$, with $\|.\|$ the Euclidean norm. We shall write $\mathscr{B}^*(\mathscr{X}_n) = \mathscr{B}_d(c_n^*, r_n^*)$, where c_n^* (the Chebyshev centre of $\mathscr{X}_n$) minimises

$$f(c) = \max_{i=1,\ldots,n} \|X_i - c\|^2 \tag{1}$$

with respect to $c \in \mathbb{R}^d$ and $r_n^* = \max_{i=1,\ldots,n} \|X_i - c_n^*\|$. A ball $\mathscr{B}_d(c, r)$ is said to be a $(1+\epsilon)$-approximation to $\mathscr{B}^*(\mathscr{X}_n)$, $\epsilon > 0$, when $\mathscr{X}_n \subset \mathscr{B}_d(c, r)$ and $r \leq (1+\epsilon) r_n^*$; a subset $\mathscr{X}_q \subseteq \mathscr{X}_n$ is said to be an ϵ-core set of $\mathscr{X}_n$ if $\mathscr{B}^*(\mathscr{X}_q) = \mathscr{B}_d(c_q^*, r_q^*)$ is such that $r_q^* \leq r_n^* \leq (1+\epsilon) r_q^*$.

The construction of $\mathscr{B}^*(\mathscr{X}_n)$ is a classical optimisation problem, for which many algorithms have been proposed in the literature, see, e.g., the historical sketch in [7] and the references in [28]. A recent application concerns the construction of space-filling de-

*Corresponding author. Email: Luc.Pronzato@cnrs.fr

signs for compter experiments based on an extension of Lloyd's clustering algorithm [16]. Some methods are exact and rely on extensions of linear programming algorithms, see [5, 9, 26]; some use the dual formulation of the problem and construct a sequence of $(1+\epsilon_k)$-approximations of $\mathscr{B}^*(\mathscr{X}_n)$ with ϵ_k tending to zero, see [4, 28]. The former are exponential in d and are thus restricted to problems with moderate dimension ($d \lesssim 20$, say); the latter can also solve problems with large d and compute a $(1+\epsilon)$-approximation to $\mathscr{B}^*(\mathscr{X}_n)$ in $O(nd/\epsilon)$ arithmetic operations, returning an ϵ-core set of size $O(1/\epsilon)$, see [6, 28].

Both types of methods can strongly benefit from a reduction of the size of $\mathscr{X}_n$, in the same way as algorithms for the construction of the minimum-volume ellipsoid containing $\mathscr{X}_n$ can be accelerated when inessential points are eliminated by the inequality of [11], see [24, Sect. 3.6]. The objective of removing inessential points presents some similarities with that of obtaining small ϵ-core sets, with one capital difference though: a point X_i is called inessential when $\mathscr{B}^*(\mathscr{X}_n \setminus \{X_i\})$ *exactly* coincides with $\mathscr{B}^*(\mathscr{X}_n)$, which happens in particular when X_i lies in the interior of $\mathscr{B}^*(\mathscr{X}_n)$. By removing inessential points, we thus aim at constructing small 0-core sets. Although we know there always exists a 0-core set of size at most $d+1$, its construction requires the knowledge of $\mathscr{B}^*(\mathscr{X}_n)$. The objective of the paper is to derive a simple inequality that any point X_j on the boundary of $\mathscr{B}^*(\mathscr{X}_n)$ must satisfy, without knowing $\mathscr{B}^*(\mathscr{X}_n)$. More precisely, we show that for any probability measure ξ on $\mathscr{X}_n$, with $c(\xi)$ and $V(\xi)$ the corresponding mean and covariance matrix respectively, any point X_j on the boundary of $\mathscr{B}^*(\mathscr{X}_n)$ satisfies

$$\|X_j - c(\xi)\|^2 \geq \text{trace}[V(\xi)] + \gamma(\xi) - \sqrt{\gamma(\xi)\{2\text{trace}[V(\xi)] + \gamma(\xi)\}}\,, \tag{2}$$

where $\gamma(\xi) = \max_{i=1,\ldots,n} \|X_i - c(\xi)\|^2 - \text{trace}[V(\xi)]$. We also prove that this bound on $\|X_j - c(\xi)\|^2$ is, in some sense, the best possible, and a comparison with the bound previously proposed in [2] is provided. Since algorithms based on the dual formulation of the smallest enclosing ball problem generate a sequence of measures ξ^k, they provide for free a sequence of inequalities that can be used as sieves to eliminate inessential points from $\mathscr{X}_n$, and thereby generate a sequence of 0-core sets of decreasing size. When imbedded in the algorithm, these sieves yield an increasing simplification of iterations, and thus an acceleration of the algorithm, see [2]. Moreover, the 0-core set obtained after a few iterations may be small enough to allow the efficient use of an exact quadratic programming (QP) algorithm for the construction of $\mathscr{B}^*(\mathscr{X}_n)$

The paper is organised as follows. Section 2 introduces the notation and presents the QP and dual formulations of the minimum enclosing ball problem. The inequality (2) is proved in Section 3, where we also explain why it cannot be improved. Two iterative algorithms are presented in Section 4: a multiplicative algorithm inspired from experimental design theory and the vertex-direction algorithm of [28]. Some computational results are presented in Section 5 that illustrate the benefit of the elimination of inessential points, when using an iterative algorithm to solve the dual problem or before using QP for the direct approach. In particular, they indicate that for moderate d the application of an exact QP algorithm to the resulting 0-core set yields the exact minimum ball $\mathscr{B}^*(\mathscr{X}_n)$ at reasonable computational cost. Section 6 briefly concludes.

2. Quadratic programming and dual formulations

For any $X_i \in \mathscr{X}_n$ and c and c_0 in $\mathbb{R}^d$, we can write $\|X_i - c\|^2 = \|X_i - c_0\|^2 - 2(X_i - c_0)^\top(c - c_0) + \|c - c_0\|^2$. Therefore, $f(c)$ defined in (1) can be written as

$$f(c) = \max_{i=1,\ldots,n} \left\{ \|X_i - c_0\|^2 - 2(X_i - c_0)^\top(c - c_0) \right\} + \|c - c_0\|^2 , \tag{3}$$

and its minimisation is equivalent to the minimisation of $\|c - c_0\|^2 + t$ with respect to $(c, t) \in \mathbb{R}^{d+1}$, subject to the n linear constraints

$$\|X_i - c_0\|^2 - 2(X_i - c_0)^\top(c - c_0) \le t \,, \; i = 1, \ldots, n \,. \tag{4}$$

When d is small enough, simplex-type or projection methods can thus be used to obtain the *exact* solution in finite time (assuming calculations with infinite precision); see in particular the introduction of [10] and the references therein. In case the QP solver requires a strictly convex problem, one may add a regularisation term, quadratic in t, to the objective function and minimise $\|c - c_0\|^2 + t + \delta t^2$ with δ arbitrarily small (the solution obtained being however not exact in this case).

On the other hand, the dual formulation of the problem yields iterative methods that construct a sequence of $(1+\epsilon_k)$-approximations of $\mathscr{B}^*(\mathscr{X}_n)$ with ϵ_k tending to zero, which are of particular interest when d is large. Direct calculation, using Lagrangian duality, shows that the construction of $\mathscr{B}^*(\mathscr{X}_n)$ is equivalent to the determination of Lagrange coefficients that define weights $w = (w_1, \ldots, w_n)$ in the probability simplex

$$\mathscr{P}_n = \{ w \in \mathbb{R}^n : \sum_{i=1}^n w_i = 1 \text{ and } w_i \ge 0, \, i = 1, \ldots, n \} \tag{5}$$

and maximise

$$\phi(w) = \text{trace}[V(w)] = \sum_{i=1}^n w_i \|X_i - c(w)\|^2 \,, \tag{6}$$

where $c(w) = \sum_{i=1}^n w_i X_i$ and $V(w) = \sum_{i=1}^n w_i [X_i - c(w)][X_i - c(w)]^\top$; see, e.g., [7, 28]. The centre c_n^* of $\mathscr{B}^*(\mathscr{X}_n)$ corresponds to $c(w^*)$ for the optimal weights w^* maximising $\phi(w)$, and its radius r_n^* equals $\sqrt{\phi(w^*)}$. The weights w define a probability measure ξ on the X_i, and $c(w)$ and $V(w)$ respectively correspond to the mean and variance matrix for ξ (which, with a slight abuse of notation, we shall also denote by $c(\xi)$ and $V(\xi)$).

There exist other geometrical problems for which the dual is known to correspond to an optimal design problem, i.e., to the construction of an optimal probability measure on $\mathscr{X}_n$. In particular, the determination of the minimum-volume ellipsoid, with fixed centre c, containing $\mathscr{X}_n$, is equivalent to the D-optimal design problem corresponding to the maximisation of $\det[M(w)]$ with respect to $w = (w_1, \ldots, w_n) \in \mathscr{P}_n$, with $M(w)$ the information matrix

$$M(w) = \sum_{i=1}^n w_i (X_i - c)(X_i - c)^\top$$

for the estimation of the d unknown parameters θ in the linear regression model $y_i = (X_i - c)^\top \theta + \varepsilon_i$, where the ε_i are i.i.d. observation errors; see [19]. The optimal ellipsoid is given by $\{x \in \mathbb{R}^d : (x-c)^\top M(w^*)(x-c) \leq 1/d\}$, with w^* an optimal vector of weights. When the centre of the ellipsoid is free, the determination of the minimum-volume enclosing ellipsoid forms a D-optimal design problem in $\mathbb{R}^{d+1}$ [21]: the optimal ellipsoid is given by the intersection between the minimum enclosing ellipsoid, centred at the origin, for the n points $(X_i, 1) \in \mathbb{R}^{d+1}$, and the hyperplane $\{z \in \mathbb{R}^{d+1} : z_{d+1} = 1\}$; see also [17, Sects. 5.6 & 9.1] and the references therein. The connection between the construction of the thinnest covering cylinder and a D_s-optimal design problem is established in [20] for cylinders with fixed centre and in [21] when the centre is free.

On the other hand, the maximisation of $\mathrm{trace}[V(w)]$ in (6) is not equivalent to an A-optimal design problem, for which one minimises $\mathrm{trace}[M^{-1}(w)]$ for some information matrix $M(w)$. As shown in the next section, the connection with an optimal design problem can nevertheless be used to derive the inequality (2), using an approach resembling that in [11].

3. An inequality to eliminate inessential points

Consider the more general situation where $\mathscr{X}$ denotes a compact subset of $\mathbb{R}^d$, with Ξ the set of probability measures on $\mathscr{X}$. For any $\xi \in \Xi$, denote

$$c(\xi) = \mathsf{E}_\xi(x) = \int_{\mathscr{X}} x\, \xi(\mathrm{d}x) \text{ and } \phi(\xi) = \mathrm{trace}[\mathsf{Var}(\xi)] = \int_{\mathscr{X}} \|x - c(\xi)\|^2\, \xi(\mathrm{d}x)\,, \tag{7}$$

so that $c(\xi) = c(w)$ and $\phi(\xi) = \phi(w)$ in the finite case where $\mathscr{X} = \mathscr{X}_n$ with $w_i = \xi(X_i)$, $i = 1, \ldots, n$. The dual problem to the determination of $\mathscr{B}^*(\mathscr{X})$ corresponds to the maximisation of $\phi(\xi)$ with respect to $\xi \in \Xi$: the centre c^* and radius r^* of $\mathscr{B}^*(\mathscr{X})$ satisfy $c^* = c(\xi^*)$ and $r^* = \sqrt{\phi(\xi^*)}$, where ξ^* maximises $\phi(\xi)$ with respect to $\xi \in \Xi$.

3.1 *A necessary and sufficient condition for optimality*

First note that Ξ is convex: for any $\xi, \nu \in \Xi$ and $\alpha \in [0,1]$, $(1-\alpha)\xi + \alpha\nu \in \Xi$. Denote $g(\alpha) = \phi[(1-\alpha)\xi + \alpha\nu]$, which is a quadratic function of α. The directional derivative of $\phi(\xi)$ at ξ in the direction $\nu \in \Xi$ is given by

$$F_\phi(\xi;\nu) = \left.\frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha}\right|_{\alpha=0} = \int \|x - c(\xi)\|^2\, \nu(\mathrm{d}x) - \phi(\xi)\,. \tag{8}$$

Note that $\mathrm{d}^2 g(\alpha)/\mathrm{d}\alpha^2 = -2\|c(\nu) - c(\xi)\|^2 \leq 0$, showing that $\phi(\cdot)$ is concave. It is not strictly concave[1], but any pair ξ_a^* and ξ_b^* of optimal measures necessarily satisfy $c(\xi_a^*) = c(\xi_b^*)$, implying that the optimal ball is unique. Concavity implies that $\xi^* \in \Xi$ is optimal if and only if $F_\phi(\xi^*;\nu) \leq 0$ for all $\nu \in \Xi$. This is equivalent to $F_\phi(\xi^*;\delta_x) \leq 0$ for all $x \in \mathscr{X}$, with δ_x the delta measure at x. Moreover, $F_\phi(\xi^*;\xi^*) = 0$ implies that $F_\phi(\xi^*;\delta_x) = 0$ for any x in the support of ξ; that is, $\xi^*\{x \in \mathbb{R}^d : F_\phi(\xi^*;\delta_x) = 0\} = 1$. We thus obtain the following property, usually called Equivalence Theorem in experimental design theory (see, e.g., [8, 12, 14, 18]). When $\mathscr{X}$ is finite, the conditions are equivalent to the Karush-Kuhn-Tucker optimality conditions in [28]; see also [7].

[1] There may exist $\xi \neq \nu$ such that $\phi(\xi) = \phi(\nu)$ and $c(\xi) = c(\nu)$, and then $g(\alpha)$ is constant for all $\alpha \in [0,1]$ (think for example of $\mathscr{X}_n$ given by the vertices of several regular simplices in $\mathbb{R}^d$ all having the same centre).

THEOREM 3.1 *The centre of $\mathscr{B}^*(\mathscr{X})$ is given by $c(\xi^*)$, where $\xi^* \in \Xi$ satisfies any of the three following equivalent conditions:*
(i) ξ^* maximises $\phi(\xi)$ with respect to $\xi \in \Xi$,
(ii) ξ^* minimises $\max_{x\in\mathscr{X}} \|x - c(\xi)\|^2$ with respect to $\xi \in \Xi$,
(iii)

$$\|x - c(\xi^*)\|^2 \leq \phi(\xi^*) \text{ for all } x \in \mathscr{X}\,. \tag{9}$$

Moreover, $\|x - c(\xi^*)\|^2 = \phi(\xi^*)$ for any x in the support of ξ^*.

3.2 $(1+\epsilon)$-approximations and ϵ-core sets

For any $\xi \in \Xi$, define

$$\gamma(\xi) = \max_{x\in\mathscr{X}} \|x - c(\xi)\|^2 - \phi(\xi)\,. \tag{10}$$

Since $\gamma(\xi) = \max_{x\in\mathscr{X}} F_\phi(\xi;\delta_x)$, Theorem 3.1 indicates that $\gamma(\xi) \geq 0$ for all $\xi \in \mathscr{X}$, with $\gamma(\xi^*) = 0$. In some sense, $\gamma(\xi)$ quantifies the (absolute) suboptimality of the measure ξ. In this section we show how it is related to the (relative) notions of $(1+\epsilon)$-approximation and ϵ-core set introduced in Section 1.

Consider the ball $\mathscr{B}(\xi) = \mathscr{B}_d(c(\xi), \sqrt{\gamma(\xi)+\phi(\xi)})$. It contains $\mathscr{X}$ by construction, and Theorem 3.1 indicates that the radius of $\mathscr{B}^*(\mathscr{X})$ equals $\sqrt{\phi(\xi^*)} \geq \sqrt{\phi(\xi)}$. Therefore, $\mathscr{B}(\xi)$ forms a $(1+\epsilon)$-approximation of $\mathscr{B}^*(\mathscr{X})$ for

$$\epsilon = \epsilon(\xi) = [1 + \gamma(\xi)/\phi(\xi)]^{1/2} - 1. \tag{11}$$

Let $\mathscr{S}(\xi)$ denote any compact subset of $\mathscr{X}$ such that $\xi[\mathscr{S}(\xi)] = 1$ (the support of ξ, say). From Theorem 3.1, the radius $r^*(\xi)$ of the smallest ball enclosing $\mathscr{S}(\xi)$ is not smaller than $\sqrt{\phi(\xi)}$, so that

$$\sqrt{\phi(\xi)} \leq r^*(\xi) \leq \sqrt{\phi(\xi^*)}\,, \tag{12}$$

where the second inequality follows from $\mathscr{S}(\xi) \subset \mathscr{X}$. On the other hand,

$$\phi(\xi^*) = \min_{c\in\mathbb{R}^d} \max_{x\in\mathscr{X}} \|x - c\|^2 \leq \max_{x\in\mathscr{X}} \|x - c(\xi)\|^2 = \gamma(\xi) + \phi(\xi) \tag{13}$$

(which is also a direct consequence of the concavity of $\phi(\cdot)$, which implies that, for any $\xi \in \Xi$, $\phi(\xi^*) \leq \phi(\xi) + F_\phi(\xi;\xi^*) \leq \phi(\xi) + \max_{x\in\mathscr{X}} F_\phi(\xi;\delta_x) = \gamma(\xi) + \phi(\xi)$). Therefore, the combination of (12) and (13) gives

$$r^*(\xi) \leq \sqrt{\phi(\xi^*)} \leq [1 + \gamma(\xi)/\phi(\xi)]^{1/2} r^*(\xi)\,,$$

indicating that $\mathscr{S}(\xi)$ is an ϵ-core set for ϵ given by (11).

These connections are used in particular in [28] to give a thorough characterisation of the convergence properties of two algorithms that generate sequences of measures ξ^k, in terms of their associated $(1+\epsilon_k)$-approximations and ϵ_k-core sets. See also Section 4.

3.3 *The inequality*

Following an approach similar to [11, 15], we now prove the main result of the paper.

THEOREM 3.2 *For any compact subset $\mathscr{X} \subset \mathbb{R}^d$ and any probability measure ξ on $\mathscr{X}$, any $y \in \mathscr{X}$ such that*

$$\|y - c(\xi)\|^2 < b[\phi(\xi), \gamma(\xi)] = \phi(\xi) + \gamma(\xi) - \sqrt{\gamma(\xi)[2\phi(\xi) + \gamma(\xi)]}\,, \tag{14}$$

where $\phi(\xi)$ and $\gamma(\xi)$ are respectively defined by (7) and (10), is in the interior of the smallest ball $\mathscr{B}^*(\mathscr{X})$ enclosing $\mathscr{X}$.

Proof. Take any ξ in Ξ and consider $\gamma(\xi)$ defined by (10). Then, $\|x - c(\xi)\|^2 \leq \phi(\xi) + \gamma(\xi)$ for all $x \in \mathscr{X}$, which implies that

$$\int_{\mathscr{X}} \|x - c(\xi)\|^2 \, \xi^*(\mathrm{d}x) = \phi(\xi^*) + \|c(\xi^*) - c(\xi)\|^2 \leq \phi(\xi) + \gamma(\xi) \tag{15}$$

for an optimal measure ξ^*. Also, (9) implies

$$\int_{\mathscr{X}} \|x - c(\xi^*)\|^2 \, \xi(\mathrm{d}x) = \phi(\xi) + \|c(\xi^*) - c(\xi)\|^2 \leq \phi(\xi^*)\,. \tag{16}$$

Consider now any y on the boundary of $\mathscr{B}^*(\mathscr{X})$. From Theorem 3.1 and the triangular inequality, it satisfies $\|y - c(\xi^*)\| = \sqrt{\phi(\xi^*)} \leq \|y - c(\xi)\| + \|c(\xi^*) - c(\xi)\|$, that is,

$$\|y - c(\xi)\| \geq \sqrt{\phi(\xi^*)} - \|c(\xi^*) - c(\xi)\|\,. \tag{17}$$

We do not know the values of $\phi(\xi^*)$ and $c(\xi^*)$, but we can compute a lower bound on the right-hand side of (17), using (15) and (16). Denote $u = \sqrt{\phi(\xi^*)}$ and $v = \|c(\xi^*) - c(\xi)\|$. The set $\{(u, v) \in \mathbb{R}^2 : u^2 + v^2 \leq \phi(\xi) + \gamma(\xi) \text{ and } u^2 - v^2 \geq \phi(\xi)\}$ is convex, and the minimum of $u - v$ is obtained for $u = \sqrt{\phi(\xi) + \gamma(\xi)/2}$ and $v = \sqrt{\gamma(\xi)/2}$; Figure 1 gives an illustration. Therefore, (17) implies that $\|y - c(\xi)\|^2 \geq b[\phi(\xi), \gamma(\xi)]$. ■

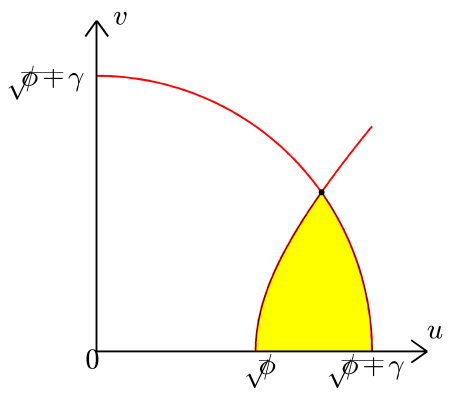

Figure 1. Determination of the lower bound (14) in the proof of Theorem 3.2: admissible set for (u, v) (coloured) and optimum point minimising $u - v$ (dot).

Note that $b(\phi,\gamma) = \phi + \gamma - \sqrt{\gamma[2\phi+\gamma]}$ is decreasing in γ, with $b(\phi,0) = \phi$ and $\lim_{\gamma\to\infty} b(\phi,\gamma) = 0$. The right-hand side of (14) gives the tightest lower bound on $\|y - c(\xi)\|^2$ for a y on the boundary of $\mathscr{B}^*(\mathscr{X})$, in the sense of the following theorem.

THEOREM 3.3 *For any integer $d \geq 2$ any $\gamma > 0$ and $\delta > 0$, there exist a compact subset $\mathscr{X}$ of $\mathbb{R}^d$, a probability measure ξ on $\mathscr{X}$, and a point y on the boundary of $\mathscr{B}^*(\mathscr{X})$ such that $\gamma = \max_{x\in\mathscr{X}} \|x - c(\xi)\|^2 - \phi(\xi)$ and $\|y - c(\xi)\|^2 < b[\phi(\xi),\gamma] + \delta$, with $b(\phi,\gamma)$ as in Theorem 3.2.*

Proof. The proof relies on the construction of an example. The dimension d is irrelevant, and we only need to consider a finite set $\mathscr{X}_3$ with three points X_1, X_2 and X_3 whose first two coordinates are respectively $(0,-1)$, $(0,1)$ and $(1+a,0)$, $a > 0$, with ξ the measure that allocates weights α, α, and $1-2\alpha$ to X_1, X_2 and X_3, $\alpha \in (0,1/2)$. Then, the first two coordinates of $c(\xi)$ are $((1-2\alpha)(1+a),0)$, and $\phi(\xi) = 2\alpha[1+(1+a)^2(1-2\alpha)]$. Also, $\|X_1 - c(\xi)\|^2 - \phi(\xi) = \|X_2 - c(\xi)\|^2 - \phi(\xi) = (1-2\alpha)[(1+a)^2(1-4\alpha)+1]$ and $\|X_3 - c(\xi)\|^2 - \phi(\xi) = -2\alpha[(1+a)^2(1-4\alpha)+1]$, so that $\gamma = \|X_1 - c(\xi)\|^2 - \phi(\xi) = \|X_2 - c(\xi)\|^2 - \phi(\xi)$ for any $a \geq 0$ when $\alpha < 1/4$. For any $\alpha < 1/4$ and $\delta > 0$, we can then choose a smaller than some $h(\alpha,\delta)$ to obtain $\|X_3 - c(\xi)\|^2 < \phi(\xi) + \gamma - \sqrt{\gamma[2\phi(\xi)+\gamma]} + \delta$. For instance, when $\alpha = 1/6$, we can take $a < h(1/6,\delta) = \sqrt{9\,\delta - 1 + 2\sqrt{27\,\delta^2 + 9\,\delta + 1}} - 1$. On the other hand, the smallest ball containing $\{X_1, X_2\}$ is $\mathscr{B}_d(0,1)$, which shows that X_3 is on the boundary of $\mathscr{B}^*(\mathscr{X}_3)$ since $\|X_3\| > 1$. ■

It is instructive to compare the bound $b[\phi(\xi),\gamma(\xi)]$ in (14) with that derived in [2]. One may first note that (15) and (16) imply that

$$\text{for any } \xi \in \Xi\,,\ \|c(\xi^*) - c(\xi)\|^2 \leq \frac{\gamma(\xi)}{2} = \phi(\xi)\,\frac{(2\epsilon+\epsilon^2)}{2}\,, \tag{18}$$

with ϵ given by (11), whereas the simple geometric arguments used in [2] only give $\|c(\xi^*) - c(\xi)\|^2 \leq \phi(\xi)\,(2\epsilon+\epsilon^2)$. In the same paper, the authors combine this inequality with (17) and obtain that any point y on the boundary of $\mathscr{B}^*(\mathscr{X})$ satisfies

$$\|y - c(\xi)\| \geq \sqrt{\phi(\xi)}\,[1 - (2\epsilon+\epsilon^2)^{1/2}] = \sqrt{\phi(\xi)}\,[1 - \sqrt{\gamma(\xi)}/\sqrt{\phi(\xi)}]\,.$$

Note that $\gamma(\xi)$ must be smaller than $\phi(\xi)$ (i.e., $\epsilon < \sqrt{2}-1$) in order to get a positive bound able to eliminate points. To compare this result with Theorem 3.2, denote

$$b_{AY}[\phi(\xi),\gamma(\xi)] = \phi(\xi)[\max\{1 - \sqrt{\gamma(\xi)}/\sqrt{\phi(\xi)},0\}]^2\,; \tag{19}$$

$b_{AY}(\phi,\gamma)$ is decreasing in γ, with $b_{AY}(\phi,0) = \phi$ and $b_{AY}(\phi,\gamma) = 0$ for $\gamma \geq \phi$, and $b_{AY}(\phi,\gamma) < b(\phi,\gamma)$ given by (14) for any $\phi > 0$ and $\gamma > 0$. We can also write $b_{AY}(\phi,\gamma) = \phi(\xi)[\max\{1-(2\epsilon+\epsilon^2)^{1/2},0\}]^2$ and $b(\phi,\gamma) = \phi[(1+\epsilon)^2 - \{\epsilon(2+\epsilon)[1+(1+\epsilon)^2]\}^{1/2}]$, with ϵ the approximation level $\epsilon = (1+\gamma/\phi)^{1/2} - 1$, see (11). Figure 2-left presents $b(\phi,\gamma)/\phi$ (solid line) and $b_{AY}(\phi,\gamma)/\phi$ (dashed line) as functions of $\epsilon \in [0,1]$; the difference between the two curves is shown on the right part. The superiority of $b(\phi,\gamma)$ compared to $b_{AY}(\phi,\gamma)$ is also significant for small ϵ, so that when approaching the optimum with an iterative algorithm, the elimination of inessential points is likely to be more efficient with (14) than when using the bound in [2]. Note that the computational costs of the two bounds are roughly equivalent.

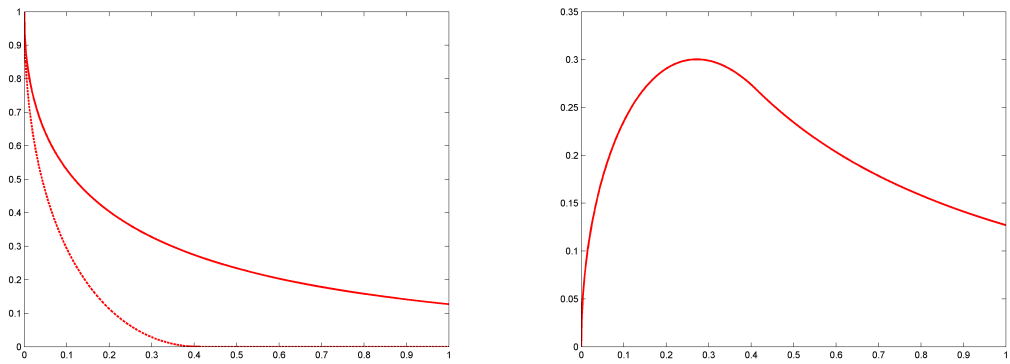

Figure 2. $b(\phi,\gamma)/\phi$ (solid line, left), $b_{AY}(\phi,\gamma)/\phi$ (dashed line, left) and $[b(\phi,\gamma)-b_{AY}(\phi,\gamma)]/\phi$ (right) as functions of $\epsilon = (1+\gamma/\phi)^{1/2}-1$.

3.4 *Effectiveness of the elimination*

Take any probability measure ξ on $\mathscr{X}$ and consider a point y eliminated by (14), that is, such that $\|y-c(\xi)\|^2 < b[\phi(\xi),\gamma(\xi)]$. By construction of the bound (14), it satisfies $\|y-c(\xi^*)\| \leq \sqrt{\phi(\xi^*)}$ (this can be directly checked, using the triangular inequality $\|y-c(\xi^*)\| \leq \|y-c(\xi)\| + \|c(\xi^*)-c(\xi)\|$ and the inequalities (16) and (18)). Therefore, y belongs to $\mathscr{B}^*(\mathscr{X}) = \mathscr{B}_d(c(\xi^*),\sqrt{\phi(\xi^*)})$. Let $\mathscr{I}(\xi)$ denote the set of inessential points eliminated by (14) and μ denote the Lebesgue measure on $\mathscr{X}$. We thus have

$$\omega(\xi) = \frac{\mu[\mathscr{I}(\xi)\cap\mathscr{B}^*(\mathscr{X})]}{\mu[\mathscr{B}^*(\mathscr{X})]} = \frac{\mu[\mathscr{I}(\xi)]}{\mu[\mathscr{B}^*(\mathscr{X})]} \leq \left(\frac{b[\phi(\xi),\gamma(\xi)]}{\phi(\xi^*)}\right)^{d/2} .$$

Denote $\delta(\xi) = \gamma(\xi)/\phi(\xi)$, and suppose that $\delta(\xi)=\delta>0$. Then, $b[\phi(\xi),\gamma(\xi)] = \phi(\xi)(1+\delta-\sqrt{\delta(2+\delta)})$ and Lemma 3.2 of [28] implies that $\phi(\xi^*) > \phi(\xi)(1+\delta^2/[4(1+\delta)])$. Therefore,

$$\omega(\xi) < h^{d/2}(\delta)\,, \tag{20}$$

with $h(\delta) = 4(1+\delta)(1+\delta-\sqrt{\delta(2+\delta)})/[4(1+\delta)+\delta^2] < 1$, implying that $\mu[\mathscr{I}(\xi)]/\mu[\mathscr{B}^*(\mathscr{X})] \to 0$ as $d\to\infty$. We can thus expect that in general, for points X_i approximately uniformly distributed in a compact set, the effectiveness of the sieve formed by (14) will decrease as the dimension d increases. This can be investigated more precisely in some simple situations. Define

$$\alpha(\xi) = \frac{\mu[\mathscr{I}(\xi)]}{\mu(\mathscr{X})}\,,$$

the proportion of points eliminated by (14), and let ξ_u denote the uniform probability measure on $\mathscr{X}$.

$\mathscr{X}$ is the d-dimensional ball $\mathscr{B}_d(0,1)$. In that case, $\mathscr{X} = \mathscr{B}^*(\mathscr{X})$ and $\alpha(\xi)=\omega(\xi)$ for any ξ. When $x\sim\xi_u$, then $\|x\|$ has the density $\varphi(r) = dr^{d-1}$, $r\in[0,1]$, and $\phi(\xi_u) = d/(d+2)$, $\gamma(\xi_u) = 1-\phi(\xi_u) = 2/(d+2)$. This gives $b[\phi(\xi_u),\gamma(\xi_u)] = 1-2\sqrt{d+1}/(d+2)$

and therefore

$$\alpha(\xi_u) = b^{d/2}[\phi(\xi_u), \gamma(\xi_u)] = \left(1 - \frac{2\sqrt{d+1}}{d+2}\right)^{d/2},$$

which is a decreasing function of d, the values of $\alpha(\xi_u)$ being already moderate for small d, with $\alpha(\xi_u|d=2) = 1 - \sqrt{3}/2 \simeq 0.1340$ and $\alpha(\xi_u|d=3) = \sqrt{5}/25 \simeq 0.0894$. Similarly, for the bound (19) of [2] we obtain $b_{AY}[\phi(\xi_u), \gamma(\xi_u)] = d/(d+2)\,(1 - \sqrt{2/d})^2$ for $d > 2$ (and 0 for $d = 1, 2$). The values of $b[\phi(\xi_u), \gamma(\xi_u)]$ and $b_{AY}[\phi(\xi_u), \gamma(\xi_u)]$ are plotted against d in Figure 3-left; the corresponding proportions $\alpha(\xi_u)$ are presented in Figure 3-right.

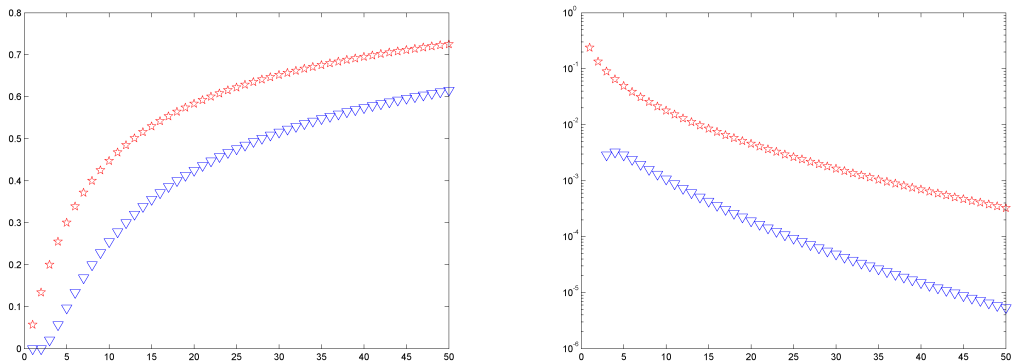

Figure 3. $b[\phi(\xi_u), \gamma(\xi_u)]$ (stars, left), $b_{AY}[\phi(\xi_u), \gamma(\xi_u)]$ (triangles, left) and corresponding proportions $\alpha(\xi_u)$ of eliminated points (right, log-scale) as functions of d.

$\mathscr{X}$ is the hypercube $[-1/2, 1/2]^d$. Direct calculation gives $\phi(\xi_u) = d/12$ and $\gamma(\xi_u) = d/4 - \phi(\xi_u) = d/6$, so that $b[\phi(\xi_u), \gamma(\xi_u)] = d(1/4 - \sqrt{2}/6)$. For $d \leq 17$, $\mathscr{B}_d(c(\xi_u), b^{1/2}[\phi(\xi_u), \gamma(\xi_u)]) \subset \mathscr{X}$, and $\alpha(\xi_u) = b^{d/2}[\phi(\xi_u), \gamma(\xi_u)]\, V_d$, with $V_d = \mathrm{vol}[\mathscr{B}_d(0,1)] = \pi^{d/2}/\Gamma(d/2+1)$ the volume of the d-dimensional unit ball $\mathscr{B}_d(0,1)$. Again $\alpha(\xi_u)$ is a decreasing function of d, with $\alpha(\xi_u|d=2) = (3 - 2\sqrt{2})\,\pi/6 \simeq 0.0898$ and $\alpha(\xi_u|d=3) = (\sqrt{2}-1)^3\,\pi/6 \simeq 0.0372$. Note that (19) does not permit to eliminate any point since $\gamma(\xi_u) > \phi(\xi_u)$.

Although (20) indicates that the effectiveness of the elimination of inessential points decreases with d for a fixed δ (that is, for a fixed level of approximation $1 + \epsilon = \sqrt{1+\delta}$, see Section 3.2), the proportion $\alpha(\xi)$ can be significant when ξ approaches optimality (so that $\delta = \delta(\xi)$ is small enough in (20)). In particular, algorithms for the solution of the dual formulation of the smallest enclosing ball problem generate sequences of measures ξ^k that can be used as sieves to progressively eliminate points. Two such methods are presented in the next section.

4. Algorithms for the dual

4.1 *A multiplicative algorithm*

We return to the case of a finite set $\mathscr{X}_n$, with $w_i = \xi(X_i)$ the weight allocated by the measure ξ to the point $X_i \in \mathscr{X}_n$. Starting with weights $w_i^0 > 0$ for all i, for instance

$w_i^0 = 1/n$, consider the application of successive iterations of the form

$$w_i^{k+1} = \widehat{w}_i^{k+1} = w_i^k \, \frac{\|X_i - c(w^k)\|^2}{\sum_{j=1}^n w_j^k \, \|X_j - c(w^k)\|^2} \, , \; i = 1, \ldots, n \, . \tag{21}$$

This type of algorithm is called multiplicative in the literature on optimal experimental design: the weights w_i^k of the measure ξ^k at iteration k are simply multiplied by positive factors $f_i(w^k)/\sum_{j=1}^k w_j^k f_j(w^k)$, with here $f_i(w^k) = \|X_i - c(w^k)\|^2 = \mathrm{d}\phi(w)/\mathrm{d}w_i\big|_{w=w^k}$. In the case of D-optimal design, similar iterations ensure monotonic convergence to the minimum-volume ellipsoid containing $\mathscr{X}_n$, see [22, 23, 29]. Here the iteration (21) does not guarantee that $\phi(w^{k+1}) > \phi(w^k)$ for all non-optimal w^k, and, following [27], we consider iterations of the (more general) form

$$\begin{aligned} w_i^{k+1} = \widetilde{w}_i^{k+1}(\beta_k) &= w_i^k \, [1 + \beta_k \, F_\phi(\xi^k; \delta_{X_i})] \\ &= w_i^k \, \{1 + \beta_k \, [\|X_i - c(w^k)\|^2 - \phi(w^k)]\} \, , \end{aligned} \tag{22}$$

where $\beta_k \geq 0$, $F_\phi(\xi; \nu)$ is the directional derivative defined in (8), and where ξ^k allocates weight w_i^k to X_i, $i = 1, \ldots, n$. Note that $\sum_{i=1}^n \widetilde{w}_i^{k+1}(\beta_k) = 1$ and that all $\widetilde{w}_i^{k+1}(\beta_k)$ remain non-negative if β_k is small enough. Also note that $\widetilde{w}_i^{k+1}[1/\phi(w^k)] = \widehat{w}_i^{k+1}$ given by (21). The iteration (22) corresponds to a projected second-order method for the maximisation of $\phi(w)$, see [27] and [17, Sect. 9.1], and there always exists a step-size $\beta_k > 0$ such that $\phi(w^{k+1}) > \phi(w^k)$ when w^* is not optimal. Since here $\phi[\widetilde{w}_i^{k+1}(\beta_k)]$ is quadratic in β_k, the maximising value β_k^* can be calculated explicitly and is given by

$$\beta_k^* = \frac{\sum_{i=1}^k \widehat{w}_i^{k+1} [\|X_i - c(w^k)\|^2 - \phi(w^k)]}{2\phi(w^k) \, \|c(\widehat{w}^{k+1}) - c(w^k)\|^2} \, , \tag{23}$$

where the components of $\widehat{w}^{k+1}$ are given by (21). Since the iteration (21) is simpler than (22)-(23), it is advisable to always try the former first, and switch to the latter only if (21) does not yield an increase of $\phi(\cdot)$ (numerical experimentation indicates that this is rather exceptional). To ensure that all components of $\widetilde{w}_i^{k+1}(\beta_k)$ remain non-negative, we should normally take $\beta_k = \min\{\beta_k^*, \beta_{k,\max}\}$, where $\beta_{k,\max} = [\phi(w^k) - \min_{j=1,\ldots,n} \|X_j - c(w^k)\|^2]^{-1} \geq 1/\phi(w^k)$, see (22). However, from the quadratic dependence of $\phi[\widetilde{w}_i^{k+1}(\beta_k)]$ in β_k, $\phi(\widehat{w}_i^{k+1}) \leq \phi(w^k)$ is equivalent to $1/\phi(w^k) \geq 2\,\widetilde{\beta}_k^*$ and thus implies $\beta_{k,\max} \geq 2\beta_k^*$. The construction is summarised in Algorithm 1.

Algorithm 1 stops when a $(1 + \epsilon_k)$-approximation of $\mathscr{B}^*(\mathscr{X}_n)$ is obtained, with $\epsilon_k = \sqrt{1 + \gamma(w^k)/\phi(w^k)} - 1 < \epsilon$. The sequence $\{\phi(w^k)\}$ is monotonically increasing, but the investigation of its convergence properties as $k \to \infty$ is out of the scope of this paper and will be considered elsewhere. The complexity of each iteration is roughly proportional to n, and the algorithm may benefit from the elimination of inessential points using the results of Section 3.3. This is considered in the next section.

4.2 *Elimination of inessential points by the multiplicative algorithm*

The uniform measure, with $w_i^0 = 1/n$ for all i, used to initialise Algorithm 1 can be used to eliminate inessential points from $\mathscr{X}_n$. For a given n, the proportion $\alpha(w^0)$ of points that can be eliminated depends on the precise location of the X_i, but we can consider the limiting situation where n tends to infinity and the X_i are uniformly distributed in a

Algorithm 1 Multiplicative algorithm for the smallest enclosing ball problem

Require: $\mathscr{X}_n$ a set of n points in $\mathbb{R}^d$ and $\epsilon > 0$.
Set $w_i^0 = 1/n$ for $i = 1, \ldots, n$; $k \leftarrow 0$;
compute $c(w^0)$, $\phi(w^0)$ and $\gamma(w^0)$.
while $\gamma(w^k)/\phi(w^k) > (1+\epsilon)^2 - 1$ **do**
  compute $\widehat{w}_i^{k+1}$ given by (21), compute $c(\widehat{w}^{k+1})$ and $\phi(\widehat{w}^{k+1})$;
  if $\phi(\widehat{w}_i^{k+1}) > \phi(w^k)$ **then** set $w^{k+1} = \widehat{w}^{k+1}$;
  else compute $w_i^{k+1} = \widetilde{w}_i^{k+1}(\beta_k^*)$ given by (22)-(23), compute $c(w^{k+1})$ and $\phi(w^{k+1})$;
  end if
  compute $\gamma(w^{k+1})$, $k \leftarrow k+1$;
end while
return w^k, $c(w^k)$, $\epsilon_k = \sqrt{1+\gamma(w^k)/\phi(w^k)} - 1$

compact set $\mathscr{X} \subset \mathbb{R}^d$ with strictly positive d-dimensional Lebesgue measure μ and equal to the closure of its interior. The X_i may be independently identically distributed in $\mathscr{X}$ with the probability measure $\xi_u = \mu/\mathrm{vol}(\mathscr{X})$, with $\mathrm{vol}(\mathscr{X})$ the volume of $\mathscr{X}$, or they may correspond to the first n points of a low-discrepancy sequence on $\mathscr{X}$, see, e.g., [13, Chap. 3]. In both situations,

$$\lim_{n\to\infty} \alpha(w^0) = \alpha(\xi_u) = \xi_u \left\{ \mathscr{B}_d \left(c(\xi_u), b^{1/2}[\phi(\xi_u), \gamma(\xi_u)] \right) \cap \mathscr{X} \right\} ,$$

where $b(\phi, \gamma)$ is given by (14) and the convergence is almost sure when the X_i are i.i.d.

The values of $\alpha(\xi_u)$ obtained in Section 3.4 for the case where $\mathscr{X}$ is a d-dimensional ball or hypercube suggest that the elimination of inessential points via (14) will be generally not very effective when using ξ_u only. Below we investigate how the situation improves when applying several iterations (21).

In terms of probability measure, the iteration (21) can be written as

$$\xi^{k+1}(\mathrm{d}x) = \frac{\|x - c(\xi^k)\|^2 \, \xi^k(\mathrm{d}x)}{\int_{y\in\mathscr{X}} \|y - c(\xi^k)\|^2 \, \xi^k(\mathrm{d}y)} \, , \; x \in \mathscr{X} \, .$$

When initialised at the uniform measure ξ_u on $\mathscr{X}$, it corresponds to the limiting behaviour of (21) as $n \to \infty$ for points X_i uniformly distributed in $\mathscr{X}$. When 0 is a centre of symmetry for $\mathscr{X}$, $\phi(\xi^{k+1}) > \phi(\xi^k)$, $c(\xi^k) = 0$ and $\max_{x\in\mathscr{X}} \|x - c(\xi^k)\|^2 = M$ for all k, with $M = 1$ when $\mathscr{X} = \mathscr{B}_d(0,1)$ and $M = d/4$ when $\mathscr{X} = [-1/2, 1/2]^d$. Direct calculation gives $b[\phi, M-\phi] = M - \sqrt{M^2 - \phi^2}$, which is increasing in ϕ, so that $\alpha(\xi^{k+1}) > \alpha(\xi^k)$.

Consider the case $\mathscr{X} = \mathscr{B}_d(0,1)$. After k iterations, $\phi(\xi^k) = \int_0^1 r^2 \, \varphi_k(r)\mathrm{d}r$, with $\varphi_k(r) = (d+2k)\, r^{d-1+2k}$, which gives $\phi(\xi^k) = (d+2k)/(d+2k+2)$. The proportion of points eliminated by (14) *after those k iterations* is

$$\alpha(\xi^k) = \left\{ 1 - [1 - \phi^2(\xi^k)]^{1/2} \right\}^{d/2} = \left(1 - 2\,\frac{\sqrt{d+1+2k}}{d+2+2k} \right)^{d/2} , \tag{24}$$

which is decreasing in d for fixed k, but increases in k for fixed d, with $\lim_{k\to\infty} \alpha(\xi^k) = 1$.

The value of α^k slightly improves when inessential points are removed *after each iteration*, provided the mass of eliminated points is suitably distributed on the remain-

ing ones. Suppose for instance that we simply renormalise the total mass of remaining points. Then, at iteration $k \geq 1$, $\phi(\xi^k) = \int_{A^{1/2}(\xi^{k-1})}^1 r^2\, \varphi_k(r)\mathrm{d}r$, where $\varphi_k(r) = (d+2k)\,[1-A^{(d+2k)/2}(\xi^{k-1})]^{-1}\, r^{d-1+2k}$, $r \in [A^{1/2}(\xi^{k-1}), 1]$, with $A(\xi) = 1-\sqrt{1-\phi^2(\xi)}$. This gives

$$\phi(\xi^k) = \frac{d+2k}{d+2(k+1)}\,\frac{1-A^{d/2+k+1}(\xi^{k-1})}{1-A^{d/2+k}(\xi^{k-1})} \text{ and } \alpha(\xi^k) = A^{d/2}(\xi^k)\,,\ k \geq 1\,. \tag{25}$$

Numerical evaluations for different d and k indicate that $\alpha(\xi^k)$ is only marginally larger than the value in (24), with the consequence that trying to remove inessential points at each iteration of Algorithm 1 is generally not very efficient.

4.3 *A vertex-direction algorithm*

Algorithm 4.1 of [28] is similar to the algorithm of [25] for the construction of the minimum ellipsoid containing $\mathscr{X}_n$ and to the algorithm proposed in [3] for the construction of a D-optimal design measure. The detailed analysis in [28] indicates in particular that the algorithm asymptotically presents linear convergence; see also [1]. An initialisation at a two-point measure is proposed,

$$\xi_2 = (1/2)(\delta_{X_{i_1}} + \delta_{X_{i_2}})\,,\ \text{with } i_1 = \arg\max_{i=1,\dots,n} \|X_i - X_1\| \text{ and } i_2 = \arg\max_{i=1,\dots,n} \|X_i - X_{i_1}\|\,, \tag{26}$$

so that $w_{i_1} = w_{i_2} = 1/2$ and $w_i = 0$ for all $i \neq i_1, i_2$ (when the order of indices is randomised, X_1 can be considered as randomly drawn among the X_i). This construction ensures that X_{i_1} and X_{i_2} will be far apart, without requiring the computation of all $n(n-1)/2$ pair distances. It is a key argument in the complexity analysis of the algorithm. Direct calculation gives $\phi(\xi_2) = \|X_{i_1} - X_{i_2}\|^2/4$.

The method is summarised in Algorithm 2 below, with two small modifications compared with the original version in [28]: (i) the choice between a plus-iteration (displacement in the direction of the furthest point X_{i^+} to the current center $c(w^k)$) or a minus-iteration (reduction of the weight allocated to the closest point X_{i^-} to $c(w^k)$ among the current support $\mathscr{J}(w^k)$) is based on the comparison between the values of $\phi(w^{k+1})$ corresponding to these two options, whereas [28] simply compares $\gamma(w^k)$ with $\gamma^-(w^k)$; (ii) the algorithm is stopped when $\gamma(w^k)/\phi(w^k) \leq (1+\epsilon)^2 - 1$, whereas the condition is $\max\{\gamma(w^k), \gamma^-(w^k)\}/\phi(w^k) \leq (1+\epsilon)^2 - 1$ in [28]. These minor differences do not modify the complexity analysis in the same paper, and the algorithm returns a $(1+\epsilon)$-approximation in $18 + 50/\epsilon$ iterations at most.

The two-point measure ξ_2 defined by (26) can also be used to eliminate inessential points. Let X_{i^*} denote the furthest point in $\mathscr{X}_n$ from $c(\xi_2) = (X_{i_1} + X_{i_2})/2$. Then, $\|X_{i^*} - c(\xi_2)\| \leq \sigma\, \|X_{i_2} - X_{i_1}\|$ for some $\sigma > 0$ implies that $\gamma(\xi_2)/\phi(\xi_2) \leq 4\sigma^2 - 1$ and thus

$$\frac{b[\phi(\xi_2), \gamma(\xi_2)]}{\phi(\xi_2)} \geq \tau^2 = 4\sigma^2 - \sqrt{16\sigma^4 - 1}\,.$$

Any point X_i such that $\|X_i - c(\xi_2)\| < (\tau/2)\, \|X_{i_2} - X_{i_1}\|$ is thus in the interior of $\mathscr{B}^*(\mathscr{X}_n)$. On the other hand, note that the bound $b_{AY}[\phi(\xi_2), \gamma(\xi_2)]$ given by (19) is informative only when $\sigma < \sqrt{2}/2$ (to ensure that $\gamma(\xi_2) < \phi(\xi_2)$). Since τ is decreasing in σ, the smaller σ is, the more efficient the elimination of inessential points. For instance, when

Algorithm 2 Vertex-direction algorithm for the smallest enclosing ball problem

Require: $\mathscr{X}_n$ a set of n points in $\mathbb{R}^d$ and $\epsilon > 0$.
Set $w^0_{i_1} = w^0_{i_2} = 1/2$ and $w^0_i = 0$ for all $i \neq i_1, i_2$, where i_1 and i_2 are given by (26);
$k \leftarrow 0$;
Set $c(w^0) = (X_{i_1} + X_{i_2})/2$, $\phi(w^0) = \|X_{i_1} - X_{i_2}\|^2/4$, $\mathscr{J}(w^0) = \{i_1, i_2\}$, $\gamma^-(w^0) = 0$, $i^- = 1$, compute $\gamma(w^0)$ and $i^+ = \arg\max_{i=1,\dots,n} \|X_i - c(w^0)\|$.
while $\gamma(w^k)/\phi(w^k) > (1+\epsilon)^2 - 1$ **do**
  if $\gamma(w^k) > \gamma^-(w^k)/[1 - \gamma^-(w^k)/\phi(w^k)]$ **then**
    compute $\alpha_k = \gamma(w^k)/\{2[\phi(w^k) + \gamma(w^k)]\}$,
    set $w^{k+1}_{i^+} = (1-\alpha_k)w^k_{i^+} + \alpha_k$ and $w^{k+1}_i = (1-\alpha_k)w^k_i$ for all $i \neq i^+$,
    compute $c(w^{k+1}) = (1-\alpha_k)c(w^k) + \alpha_k X_{i^+}$;
  else
    compute $\alpha_k = \min\left\{\gamma^-(w^k)/\{2[\phi(w^k) - \gamma^-(w^k)]\}, w^k_{i^-}/(1 - w^k_{i^-})\right\}$,
    set $w^{k+1}_{i^-} = (1+\alpha_k)w^k_{i^-} - \alpha_k$ and $w^{k+1}_i = (1+\alpha_k)w^k_i$ for all $i \neq i^-$,
    compute $c(w^{k+1}) = (1+\alpha_k)c(w^k) - \alpha_k X_{i^-}$;
    if $\alpha_k = w^k_{i^-}/(1 - w^k_{i^-})$ **then**
      $\mathscr{J}(w^{k+1}) = \mathscr{J}(w^k) \setminus \{i^-\}$
    else
      $\mathscr{J}(w^{k+1}) = \mathscr{J}(w^k)$
    end if
  end if
  compute $\phi(w^{k+1})$, $\gamma(w^{k+1})$ and $i^+ = \arg\max_{i=1,\dots,n} \|X_i - c(w^{k+1})\|$,
  $i^- = \arg\min_{i=1 \in \mathscr{J}(w^{k+1})} \|X_i - c(w^{k+1})\|$ and $\gamma^-(w^{k+1}) = \phi(w^{k+1}) - \|X_{i^-} - c(w^{k+1})\|$;
  $k \leftarrow k+1$;
end while
return w^k, $c(w^k)$, $\epsilon_k = \sqrt{1 + \gamma(w^k)/\phi(w^k)} - 1$

$\mathscr{X} = \mathscr{B}_d(0,1)$ or $\mathscr{X} = [-1/2, 1/2]^d$ we can take $\sigma = 1/2$, which gives $\tau = 1$: all points in the interior of $\mathscr{B}_d(c(\xi_2), \|X_{i_2} - X_{i_1}\|/2)$ are eliminated (and ξ_2 is optimal whatever the choice of X_1 in $\mathscr{X}$). More generally, Lemma 3.1 in [28] gives $\sigma = 3/2$ for any $\mathscr{X}_n$, since

$$\|X_{i^*} - c(\xi_2)\| \leq \|X_{i^*} - X_{i_1}\| + \|X_{i_1} - c(\xi_2)\| \leq \|X_{i_2} - X_{i_1}\| + \frac{1}{2}\|X_{i_1} - X_{i_2}\| = \frac{3}{2}\|X_{i_1} - X_{i_2}\|\,.$$

This bound is not tight, however: equality can only be achieved when X_{i^*}, X_{i_1} and X_{i_2} are aligned, with X_{i_1} between X_{i^*} and X_{i_2}, which contradicts the fact that X_{i_1} is the furthest point in $\mathscr{X}_n$ from some X_1. A more precise analysis, see Appendix A, yields $\sigma = \sqrt{7}/2$, and the corresponding bound is tight. This indicates that, for any set $\mathscr{X}_n$ and for any point $X_1 \in \mathscr{X}_n$ used for the construction of ξ_2, any X_i such that

$$\|X_i - c(\xi_2)\| < 0.133974\,\|X_{i_2} - X_{i_1}\| < \sqrt{7 - 4\sqrt{3}}\,\|X_{i_2} - X_{i_1}\|/2 \tag{27}$$

can always be eliminated[2]. In practice, $\|X_{i^*} - c(\xi_2)\|$ is often much smaller than $\sqrt{7}\,\|X_{i_1} - X_{i_2}\|/2$, and ξ_2 proves generally more efficient than the uniform measure ξ_u for eliminating inessential points. This is illustrated in the next section.

[2]Although the value $\sigma = \sqrt{7}/2$ gives a tight bound, one may notice that the inequality (27) is suboptimal since the worst-case situations in Theorem 3.3 and Lemma A.1 correspond to different measures.

5. Computational results

Methods to be compared. In this section, we report the results of computational experiments comparing different methods for the construction of $\mathscr{B}^*(\mathscr{X}_n)$. The first one (henceforth QP) corresponds to the direct application of the QP solver of Matlab (the function `qp.m`) to the minimisation of (3), see Section 1. In the method QP_0, we first eliminate inessential points using the sieve (14) for ξ_2 given by (26) and then apply the same QP solver.

The choice of c_0 in (3) is arbitrary, and $c_0 = c(\xi_u) = (1/n)\sum_{i=1}^n X_i$ seems natural. However, we found that c_0 has a significant influence on the computational time, and that taking c_0 out of the convex hull $\mathrm{Conv}(\mathscr{X}_n)$ of $\mathscr{X}_n$ generally yields a faster computation of the optimal solution. Note that, when $c_0 \notin \mathrm{Conv}(\mathscr{X}_n)$, for any $t \in \mathbb{R}$ there exists a $c \in \mathbb{R}^d$ satisfying the constraints (4) (and the set of such feasible c is unbounded). On the other hand, no feasible c exists for small enough t when $c_0 \in \mathrm{Conv}(\mathscr{X}_n)$. In our computations we take $c_0 = 2\,X_{i_a} - X_{i_b}$, where $i_a = \arg\max_{i=1,\dots,n} u^\top X_i$ and $i_b = \arg\min_{i=1,\dots,n} u^\top X_i$, with $u^\top = (1, 0, \dots, 0)$ (the choice of u does not seem important). The QP solver is initialised at $(c(\xi_u), 0)$ (which is not necessarily feasible for (4)).

We also consider the iterative construction of an $(1+\epsilon)$-approximation of $\mathscr{B}^*(\mathscr{X}_n)$, using Algorithms 1 and 2 (henceforth A1 and A2), both with $\epsilon = 10^{-3}$. A1 and A2 do not eliminate any point. As noticed in Section 4.2, it is not very efficient to try to eliminate inessential points at each iteration of A1. Our experiments indicate that a suitable compromise between the computational cost of the elimination test and the benefit of reducing the dimension of w is obtained when the sieve (14) is used about every 5 iterations of A1 or A2; the corresponding methods are denoted by $\mathrm{A1}_5$ and $\mathrm{A2}_5$, respectively. For each of them, inessential points are also eliminated at the initialisation, using (14) with ξ_2. $\mathrm{A1}'_5$ and $\mathrm{A2}'_5$ differ from $\mathrm{A1}_5$ and $\mathrm{A2}_5$ by the stopping rule only: they are stopped when an $(1+\epsilon)$-approximation is obtained or earlier if $n-2d$ inessential points have already been eliminated. In case of early stopping, QP applied to the resulting 0-core set will thus have to deal with $2d$ constraints only (the value $2d$ is somewhat arbitrary, but seems reasonable for most situations since $\mathscr{B}^*(\mathscr{X}_n)$ has $d+1$ points at most on its boundary when the n points in $\mathscr{X}_n$ are in general position). In $\mathrm{A1}'_5$-QP and $\mathrm{A2}'_5$-QP we apply QP to the 0-core sets returned by $\mathrm{A1}'_5$ and $\mathrm{A2}'_5$, respectively. Finally, $\mathrm{A2}^*_5$ is similar to $\mathrm{A2}_5$ but uses $\epsilon = 10^{-6}$, and thus returns an $(1+\epsilon)$-approximation very close to the exact $\mathscr{B}^*(\mathscr{X}_n)$ given by QP, QP_0, $\mathrm{A1}'_5$-QP and $\mathrm{A2}'_5$-QP. We shall call these methods (including $\mathrm{A2}^*_5$) *exact* in what follows.

When using A1 or A2, points that are eliminated by (14) for the current measure ξ^k may carry a positive weight w_i^k, and the weights of remaining points then need to be renormalised. Denote by $\mathscr{I}_k$ the set of indices of those remaining points; following [11], we replace w_i^k by $z_i^k/(\sum_{j=1}^n z_j^k)$, where $z_i^k = 0$ for $i \notin \mathscr{I}_k$, $z_i^k = 1.1\,w_i^k$ if $\|X_i - c(w^k)\|^2 \geq \phi(w^k)$ and $z_i^k = w_i^k$ otherwise ($i \in \mathscr{I}_k$ and $\|X_i - c(w^k)\|^2 < \phi(w^k)$).

Measures of performance. The experiments were carried out on a PC with a clock speed of 2.50 GHz and 32 Go RAM.

We first compare (Tables 1, 4 and 7) the effectiveness of the sieve (14) for the uniform measure ξ_u used to initialise A1 and for ξ_2 given by (26): $\pi(\xi) = 1 - \alpha(\xi)$ gives the proportion of points that are not eliminated by ξ. To compare the efficiency of (14) with that of the bound (19) proposed in [2], we also give the value $\pi_{AY}(\xi_2)$ obtained when $b_{AY}[\phi(\xi_2), \gamma(\xi_2)]$ is used instead of $b[\phi(\xi_2), \gamma(\xi_2)]$. We also indicate the number κ of remaining points after using $\mathrm{A1}_5$, $\mathrm{A2}_5$ or $\mathrm{A2}^*_5$ that incorporate an elimination of

inessential points.

In Tables 2, 5 and 8, N_ϵ gives the number of iterations performed to reach the required precision ϵ for A1, A1$_5$, A2, A2$_5$ ($\epsilon = 10^{-3}$) and A2$_5^*$ ($\epsilon = 10^{-6}$), or to eliminate at least $n - 2d$ points for A1$_5'$ and A2$_5'$.

Finally, in Tables 3, 6 and 9 we compare the computational times of the different methods considered, with t(QP), the computational time of QP, taken as a reference: for each method M other than QP, with computational time t(M), we indicate the ratio ρ(M)=t(M)/t(QP).

n consecutive points of Sobol' low-discrepancy sequence in $[0,1]^d$. Table 1 indicates that ξ_2 is much more effective than the uniform measure ξ_u for eliminating points with (14) when d is not too large, $d \lesssim 10$ say; one may note the good agreement between $\pi(\xi_u)$ and the theoretical value $\pi^* = 1 - [\pi\, d(1/4 - \sqrt{2}/6)]^{d/2}/\Gamma(d/2+1)$ ($d \leq 17$) derived in Section 3.4. For d between 3 and 10, $\pi_{AY}(\xi_2)$ is most often significatively larger than $\pi(\xi_2)$, which illustrates the superiority of the bound (14) over (19). The number of remaining points after running A1$_5$ or A2$_5$ are very close in most cases. Exceptions, like $n = 10^3$ and $n = 10^4$ for $d = 3$ and $n = 10^5$ for $d = 4$, correspond to situations where A2$_5$ is used for less than 5 iterations, so that inessential points are only eliminated once (at the initialisation) whereas A1$_5$ makes much more iterations, see Table 2 (when less than 5 iterations are done, then $\kappa = n\,\pi(\xi_2)$). As expected, κ(A2$_5^*$) is smaller than κ(A2$_5$) in all cases, and Table 1 indicates that A2$_5^*$ is able to provide small 0-core sets for the sets $\mathscr{X}_n$ considered.

Table 2 shows that the elimination of inessential points does not directly influence the number of iterations required to reach a given precision: N_ϵ(A1$_5$) is often smaller than N_ϵ(A1), but not always; the effect on $A2$ is limited. A1$_5$ requires systematically more (sometimes much more) iterations than A2$_5$ to reach the required precision ϵ, which can be related to the general observation that multiplicative algorithms tend to be slow close to the optimum. This is consistent with the observations that sometimes A1$_5'$ requires significantly less iterations than A1$_5$, whereas N_ϵ(A2$_5'$) is close to N_ϵ(A2$_5$) in all circumstances: A1$_5$ may have reached an $(1+\epsilon')$-approximation, $\epsilon' > \epsilon$, close enough to the optimum to be able to eliminate many points, but may still require many iterations to reach an $(1+\epsilon)$-approximation. The number of iterations of A2$_5^*$ ($\epsilon = 10^{-6}$) shows a great variability among the cases considered, and the large values obtained for $d = 2$, $n = 10^3$ and $n = 10^4$ may look surprising. However, they do not contradict the complexity bound $N_\epsilon(A2) < 18 + 50/\epsilon$ of [28] and can be explained by the potential slow convergence of first-order methods close to the optimum. A simple example with $d = 2$ and $n = 4$ gives an illustration.

Take $\mathscr{X}_n = \{X_1, X_2, X_3, X_4\}$ with $X_1 = (1-a, a)^\top$, $X_2 = (a, 1-a)^\top$, $X_3 = (0,0)^\top$ and $X_4 = (1,1)^\top$, $a < 1/2$. When $a < 1/2 - \sqrt{3}/6$, then $\|X_1 - X_2\| > \|X_1 - X_3\|$, so that $i_1 = 2$ and $i_2 = 1$ in (26). The initial w^0 of $A2$ is thus $(1/2, 1/2, 0, 0)$, and $A2$ may require many iterations to reach precision ϵ depending on the value of a. For instance, for $\epsilon = 10^{-5}$, N_ϵ(A2)=6252 when $a = 10^{-3}$ and N_ϵ(A2)=62502 when $a = 10^{-4}$ (whereas N_ϵ(A1)=7361 and N_ϵ(A1)=1 for $a = 10^{-3}$ and $a = 10^{-4}$, respectively).

A noticeable observation from Table 3 is that a standard QP solver gives the solution in reasonable time if n is not too big, even for rather large d. A1$_5'$ (respectively, A2$_5'$) is slightly faster than A1$_5$ (respectively, A2$_5$) since it is stopped earlier; the comparison with A1 (respectively, A2) shows that the elimination of points significantly accelerates convergence[3]. Since A1$_5$ and A2$_5$ only provide $(1+\epsilon)$-approximations with $\epsilon = 10^{-3}$,

[3]The same observation can be made on Tables 6 and 9.

Table 1. Sobol' sequence in $[0,1]^d$: proportion π (in %) of points not eliminated and number κ(M) of remaining points after applying method M.

d	n	π^*	$\pi(\xi_u)$	$\pi(\xi_2)$	$\pi_{AY}(\xi_2)$	$\kappa(\mathrm{A1}_5)$	$\kappa(\mathrm{A2}_5)$	$\kappa(\mathrm{A2}_5^*)$
2	10^3	91.02	91.0	0.4	0.4	4	4	4
	10^4	91.02	91.04	0.04	0.04	4	4	4
	10^5	91.02	91.03	0.004	0.004	4	4	4
3	10^3	96.28	96.66	1.3	2.40	5	13	4
	10^4	96.28	96.34	1.23	3.96	11	123	4
	10^5	96.28	96.30	0.060	0.136	45	60	3
4	10^3	98.39	98.30	17.2	44.8	8	7	3
	10^4	98.39	98.41	0.02	0.02	2	2	2
	10^5	98.39	98.39	2.318	10.359	32	2318	5
5	10^3	99.28	99.30	34.1	70.5	10	8	5
	10^4	99.28	99.28	19.16	63.45	20	16	6
	10^5	99.28	99.28	5.976	29.445	27	16	5
10	10^3	99.98	99.8	75.8	99.4	13	13	8
	10^4	99.98	99.95	85.34	99.96	28	30	10
	10^5	99.98	99.98	45.128	95.730	40	48	11
20	10^3		99.9	99.7	100.00	34	34	13
	10^4		99.99	98.56	100.00	52	57	14
	10^5		99.999	95.217	99.999	53	40	11
30	10^3		99.9	99.9	100.00	28	28	12
	10^4		99.99	99.98	100.00	42	48	14
	10^5		99.999	98.897	100.00	98	108	16
40	10^3		99.9	100.00	100.00	46	33	13
	10^4		99.99	100.00	100.00	60	71	14
	10^4		99.999	99.989	100.00	162	121	19
50	10^3		99.9	100.00	100.00	43	51	15
	10^4		99.99	100.00	100.00	77	113	17
	10^5		99.999	100.00	100.00	185	155	27

comparing their computational time with that of QP is unfair. $\mathrm{A1}_5'$-QP is sometimes faster than QP, but is always slower than $\mathrm{A2}_5'$-QP, which is often faster than QP and sometimes the fastest among the exact methods considered. $\mathrm{A2}_5^*$ is seldom the fastest among exact methods and is often much slower than QP. In this example, QP_0 is faster than QP for $n \leq 10$ and slightly slower when $n \geq 20$ (i.e., when few points are eliminated by ξ_2); it is frequently the fastest exact method when $n \leq 5$.

n points i.i.d. $\mathscr{N}(0, I_d)$. Table 4 indicates that the elimination of inessential points is more efficient with $\mathrm{A1}_5$ than $\mathrm{A2}_5$, and that both methods are able to provide small 0-core sets. For $d \lesssim 10$, $\pi_{AY}(\xi_2)$ is significatively larger than $\pi(\xi_2)$, confirming the superiority of the bound (14) over (19). Table 5 gives the same indications as Table 2: sometimes $\mathrm{A1}_5'$ requires significantly less iterations than $\mathrm{A1}_5$, an indication of the slow convergence of the multiplicative algorithm near the optimum. Also, $N_\epsilon(\mathrm{A1}_5) > N_\epsilon(\mathrm{A2}_5)$ and $N_\epsilon(\mathrm{A1}_5') > N_\epsilon(\mathrm{A2}_5')$ in all cases. One may notice the large values of $N_\epsilon(\mathrm{A2}_5^*)$. Table 6 shows that QP_0 and $\mathrm{A2}_5$-QP are often the fastest among exact methods, which is never the case for $\mathrm{A2}_5^*$. QP_0 shows remarkably stable performance and is significantly faster than QP when $n \leq 5$ (i.e., when the elimination of inessential points by ξ_2 is effective, see Table 4) and is only slightly slower than QP for $n \geq 10$. QP is the fastest exact method for n small enough ($n \leq 10^3$) when $d \geq 10$ and for all $n \geq 10^3$ when d is large ($d \geq 40$).

n points i.i.d. uniformly in $\mathscr{B}_d(0,1)$. This corresponds to a difficult situation for algorithms 1 and 2, and due to the larger computational times required compared to previous examples we only consider $d \leq 40$ (and $n \leq 10^4$ for $d = 40$). Table 7 shows that the proportion of points eliminated by ξ_u or ξ_2 is very small already for $d = 10$. Now

Table 2. Sobol' sequence in $[0,1]^d$: number N_ϵ of iterations performed to reach precision $\epsilon = 10^{-3}$ ($\epsilon = 10^{-6}$ for $A2_5^*$).

d	n	A1	$A1_5$	$A1'_5$	A2	$A2_5$	$A2'_5$	$A2_5^*$
2	10^3	44	1	0	0	0	0	32263
	10^4	173	1	0	0	0	0	32263
	10^5	266	0	0	0	0	0	0
3	10^3	80	270	70	3	3	3	12
	10^4	253	169	169	3	3	3	34
	10^5	242	219	219	1	1	1	7
4	10^3	91	84	75	5	6	5	14
	10^4	94	0	0	0	0	0	0
	10^5	229	123	123	4	4	4	818
5	10^3	93	92	75	29	22	20	76
	10^4	212	88	88	63	81	81	178
	10^5	179	107	107	50	55	55	465
10	10^3	89	139	40	62	56	35	457
	10^4	175	97	97	69	79	79	446
	10^5	200	137	137	66	74	74	930
20	10^3	241	139	115	89	89	85	714
	10^4	166	152	152	44	37	37	348
	10^5	244	142	142	61	36	35	301
30	10^3	286	204	50	37	28	25	373
	10^4	336	237	210	87	63	55	1007
	10^5	342	222	222	76	66	66	959
40	10^3	206	117	80	28	26	15	132
	10^4	115	99	90	60	56	50	311
	10^5	359	188	188	76	45	45	744
50	10^3	153	103	60	56	44	30	336
	10^4	191	154	125	56	54	54	617
	10^5	266	143	143	93	79	79	1726

Table 3. Sobol' sequence in $[0,1]^d$: computational time t(QP) (in s) and ratios ρ(M)=t(M)/t(QP) — averaged over 10 repetitions. Italicized figures correspond to the fastest exact method.

d	n	t(QP)	QP_0	A1	$A1_5$	$A1'_5$	A2	$A2_5$	$A2'_5$	$A1'_5$-QP	$A2'_5$-QP	$A2_5^*$
2	10^3	0.006	0.40	2.13	0.27	0.22	0.16	0.14	0.14	0.37	*0.29*	864.1
	10^4	0.030	*0.08*	3.27	0.08	0.07	0.04	0.05	0.05	0.10	*0.08*	164.1
	10^5	0.27	*0.06*	4.69	0.07	0.07	0.04	0.06	0.06	0.08	*0.06*	0.06
3	10^3	0.004	*0.49*	4.11	11.05	3.20	0.32	0.30	0.30	3.43	0.59	0.70
	10^4	0.029	*0.10*	5.19	1.08	1.07	0.10	0.08	0.08	1.10	0.12	0.27
	10^5	0.28	*0.06*	4.43	0.20	0.20	0.06	0.07	0.07	0.20	0.07	0.07
4	10^3	0.004	*0.56*	4.75	4.23	3.99	0.44	0.49	0.45	4.25	0.68	0.83
	10^4	0.029	0.08	1.96	0.09	0.08	0.05	0.06	0.06	0.11	0.09	*0.06*
	10^5	0.32	*0.10*	4.32	0.17	0.17	0.11	0.07	0.07	0.18	*0.10*	0.45
5	10^3	0.005	*0.92*	4.06	3.80	3.20	1.30	1.03	0.97	3.48	1.25	2.81
	10^4	0.031	*0.28*	4.22	0.67	0.66	1.20	0.52	0.53	0.70	0.57	1.04
	10^5	0.32	0.14	4.08	0.20	0.20	1.05	0.12	0.12	0.21	*0.13*	0.31
10	10^3	0.007	*0.97*	3.18	4.12	1.64	2.28	1.79	1.24	2.05	1.63	11.6
	10^4	0.040	0.95	3.61	0.93	0.93	1.36	0.49	0.49	1.01	*0.57*	1.94
	10^5	0.40	0.57	7.03	0.64	0.64	2.16	0.26	0.26	0.65	*0.27*	0.58
20	10^3	*0.011*	1.06	6.18	3.02	2.62	2.21	1.73	1.65	3.21	2.20	10.9
	10^4	0.059	1.14	6.17	1.76	1.77	1.70	0.47	0.48	1.88	*0.58*	1.33
	10^5	0.55	1.15	12.71	1.99	2.01	3.00	0.50	0.50	2.02	*0.51*	0.58
30	10^3	*0.015*	1.08	5.66	2.78	1.05	0.76	0.52	0.49	1.65	1.06	3.96
	10^4	0.075	1.18	15.65	2.30	2.26	4.02	0.73	0.72	2.38	*0.84*	2.70
	10^5	0.74	1.22	23.86	2.80	2.80	5.05	0.93	0.93	2.82	*0.94*	1.12
40	10^3	0.021	1.05	3.38	1.56	1.23	0.52	0.42	0.31	1.93	*0.94*	1.19
	10^4	*0.092*	1.16	6.52	2.58	2.58	3.28	0.99	0.98	2.78	1.16	1.46
	10^5	*0.92*	1.28	27.89	4.06	4.05	5.68	1.04	1.03	4.07	1.06	1.14
50	10^3	*0.030*	1.04	2.02	1.04	0.77	0.71	0.45	0.36	1.49	1.06	1.98
	10^4	*0.12*	1.19	11.88	3.21	3.18	3.44	1.13	1.14	3.42	1.36	1.93
	10^5	*1.19*	1.24	21.67	4.24	4.23	7.31	1.14	1.14	4.26	1.17	1.36

Table 4. X_i i.i.d. $\mathcal{N}(0, I_d)$: proportion π (in %) of points not eliminated and number $\kappa(\mathrm{M})$ of remaining points after applying method M — averaged values over 100 repetitions, rounded to the nearest integer.

d	n	$\pi(\xi_u)$	$\pi(\xi_2)$	$\pi_{AY}(\xi_2)$	$\kappa(\mathrm{A1}_5)$	$\kappa(\mathrm{A2}_5)$	$\kappa(\mathrm{A2}_5^*)$
2	10^3	93.40	12.81	40.23	4	29	10
	10^4	94.86	9.00	37.00	5	240	65
	10^5	95.92	3.61	25.69	6	495	21
3	10^3	96.70	32.94	75.37	5	52	9
	10^4	97.68	18.59	66.15	6	112	17
	10^5	98.25	9.19	50.79	8	129	4
4	10^3	98.21	46.39	84.81	7	20	5
	10^4	98.77	34.30	82.96	8	127	5
	10^5	99.16	24.00	78.00	10	507	320
5	10^3	98.89	62.18	92.52	8	22	5
	10^4	99.31	47.69	94.04	10	81	5
	10^5	99.54	33.27	84.58	12	17	5
10	10^3	99.78	93.52	99.93	13	15	8
	10^4	99.91	89.03	99.99	17	26	8
	10^5	99.96	81.00	99.92	22	25	9
20	10^3	99.99	99.94	100.00	22	24	12
	10^4	100.00	99.79	100.00	32	33	13
	10^5	100.00	99.39	100.00	42	45	14
30	10^3	100.00	100.00	100.00	30	31	16
	10^4	100.00	100.00	100.00	45	46	18
	10^5	100.00	100.00	100.00	64	67	20
40	10^3	100.00	100.00	100.00	39	40	18
	10^4	100.00	100.00	100.00	59	63	21
	10^5	100.00	100.00	100.00	86	92	24
50	10^3	100.00	100.00	100.00	48	49	21
	10^4	100.00	100.00	100.00	74	77	24
	10^5	100.00	100.00	100.00	107	112	28

Table 5. X_i i.i.d. $\mathcal{N}(0, I_d)$: number N_ϵ of iterations performed to reach precision $\epsilon = 10^{-3}$ ($\epsilon = 10^{-6}$ for $\mathrm{A2}_5^*$) — averaged values over 100 repetitions, rounded to the nearest integer.

d	n	A1	$\mathrm{A1}_5$	$\mathrm{A1}_5'$	A2	$\mathrm{A2}_5$	$\mathrm{A2}_5'$	$\mathrm{A2}_5^*$
2	10^3	80	126	95	27	23	22	91
	10^4	84	288	187	38	36	36	291
	10^5	114	132	123	40	35	35	74
3	10^3	84	98	55	36	33	26	86
	10^4	99	83	68	48	40	37	92
	10^5	112	107	100	51	46	45	170
4	10^3	107	87	59	46	40	34	385
	10^4	110	86	66	47	40	37	142
	10^5	127	90	82	74	63	61	288
5	10^3	104	82	54	50	46	37	191
	10^4	122	93	74	64	52	47	199
	10^5	136	100	91	88	74	71	333
10	10^3	125	93	52	59	49	36	320
	10^4	163	112	91	78	62	56	349
	10^5	175	124	116	86	67	64	465
20	10^3	162	121	57	64	56	35	334
	10^4	194	136	94	87	71	58	602
	10^5	222	156	144	99	82	78	754
30	10^3	169	124	53	72	62	37	465
	10^4	228	153	110	92	81	65	813
	10^5	249	168	155	117	94	89	1200
40	10^3	168	119	61	68	64	35	532
	10^4	229	159	105	93	81	65	1054
	10^5	280	182	170	114	96	93	1472
50	10^3	176	123	63	80	73	40	723
	10^4	234	151	116	106	89	70	1171
	10^5	284	177	169	117	100	95	1856

Table 6. X_i i.i.d. $\mathcal{N}(0, I_d)$: computational time t(QP) (in s) and ratios ρ(M)=t(M)/t(QP) — averaged over 100 repetitions. Italicized figures correspond to the fastest exact method.

d	n	t(QP)	QP$_0$	A1	A1$_5$	A1$'_5$	A2	A2$_5$	A2$'_5$	A1$'_5$-QP	A2$'_5$-QP	A2$^*_5$
2	10^3	0.005	*0.52*	4.01	5.98	4.66	1.43	1.24	1.16	4.87	1.39	4.15
	10^4	0.030	*0.18*	1.61	1.60	1.15	0.71	0.27	0.27	1.18	0.32	1.71
	10^5	0.27	*0.09*	1.79	0.16	0.15	0.58	0.08	0.08	0.16	*0.09*	0.10
3	10^3	0.004	*0.73*	4.16	4.61	2.72	1.87	1.65	1.37	2.95	1.66	3.69
	10^4	0.029	*0.28*	2.03	0.61	0.53	0.92	0.31	0.30	0.56	0.34	0.59
	10^5	0.29	0.16	2.18	0.17	0.17	0.85	0.10	0.10	0.17	*0.11*	0.17
4	10^3	0.005	*0.83*	4.89	3.74	2.69	2.14	1.80	1.60	2.95	1.88	14.28
	10^4	0.030	0.43	2.18	0.65	0.54	0.88	0.32	0.31	0.58	*0.36*	0.84
	10^5	0.30	0.31	2.54	0.20	0.20	1.33	0.13	0.13	0.20	*0.14*	0.24
5	10^3	0.004	*0.95*	4.83	3.70	2.62	2.37	2.04	1.74	2.93	2.08	7.14
	10^4	0.032	0.56	2.43	0.70	0.61	1.20	0.40	0.37	0.66	*0.42*	1.06
	10^5	0.32	0.42	3.06	0.25	0.25	1.79	0.16	0.16	0.26	*0.16*	0.28
10	10^3	*0.006*	1.10	4.58	3.11	1.99	2.16	1.66	1.40	2.43	1.84	8.18
	10^4	0.039	0.96	3.41	0.85	0.76	1.57	0.45	0.42	0.83	*0.49*	1.55
	10^5	0.40	0.92	6.19	0.63	0.63	2.77	0.30	0.30	0.64	*0.31*	0.46
20	10^3	*0.010*	1.08	4.23	2.51	1.44	1.70	1.23	0.88	2.00	1.43	5.35
	10^4	0.061	1.12	6.87	1.12	1.01	3.01	0.58	0.54	1.12	*0.65*	1.89
	10^5	0.58	1.16	10.94	1.21	1.21	4.45	0.51	0.52	1.22	*0.53*	0.69
30	10^3	*0.019*	1.06	2.78	1.51	0.83	1.17	0.82	0.56	1.40	1.11	4.06
	10^4	0.081	1.12	9.63	1.36	1.27	3.83	0.75	0.72	1.42	*0.87*	2.15
	10^5	0.77	1.19	16.42	1.75	1.75	7.18	0.80	0.79	1.77	*0.81*	1.02
40	10^3	*0.028*	1.04	2.09	1.09	0.69	0.83	0.61	0.42	1.37	1.09	3.24
	10^4	*0.10*	1.14	11.30	1.63	1.54	4.33	0.90	0.88	1.77	1.09	2.38
	10^5	*0.97*	1.20	20.39	2.15	2.15	7.83	1.05	1.04	2.18	1.06	1.27
50	10^3	*0.038*	1.04	1.79	0.89	0.60	0.77	0.55	0.38	1.34	1.10	3.21
	10^4	*0.13*	1.16	12.71	1.84	1.78	5.54	1.11	1.10	2.05	1.36	2.46
	10^5	*1.20*	1.22	22.43	2.48	2.49	8.74	1.33	1.33	2.52	1.36	1.56

$\pi_{AY}(\xi_2)$ is significatively larger than $\pi(\xi_2)$ for $d \lesssim 5$ only. As in Table 4, κ(A1$_5$)$<$ κ(A2$_5$), but the figures are now much larger, indicating that the algorithms have difficulties with providing small 0-core sets. As a consequence, here A1$'_5$ (respectively, A2$'_5$) does not stop earlier than A1$_5$ (respectively, A2$_5$), and the results for A1$'_5$ and A2$'_5$ are omitted in Tables 8 and 9 since they are identical to those for A1$_5$ and A2$_5$. The number of iterations for given d and n in Table 8 is significantly larger than in Tables 2 and 5, with N_ϵ(A1$_5$)$>N_\epsilon$(A2$_5$) for $d \leq 10$ and N_ϵ(A2$_5$) slightly larger than N_ϵ(A1$_5$) for $d \geq 30$. The number of iterations of A2$^*_5$ is now very large. Table 9 shows that QP$_0$ is generally the fastest among exact methods for $d \leq 5$ and is only slightly slower than QP for larger d. On the other hand, A2$'_5$-QP is much slower than QP for $d \geq 10$ and A2$^*_5$ is by far the slowest exact method is all cases considered.

Finally, one may notice that, for given d and n, the computational times for QP (and thus of QP$_0$) are quasi identical in Tables 3 and 6 and are only increased by a small factor in Table 9, enhancing the interest of using QP with elimination of inessential points to solve smallest enclosing ball problems with moderate d.

6. Conclusions

An inequality has been derived that permits to remove inessential (interior) points during the computation of the smallest enclosing ball of a set of points. The inequality is, in some sense, the best possible, and is given by a simple expression depending on the mean and the (trace of the) variance matrix of a probability measure placed on the set of points. Any probability measure gives such an an inequality. Algorithms for the

Table 7. X_i uniform in $\mathscr{B}_d(0,1)$: proportion π (in %) of points not eliminated and number $\kappa(\mathrm{M})$ of remaining points after applying method M — averaged values over 100 repetitions, rounded to the nearest integer.

d	n	π^*	$\pi(\xi_u)$	$\pi(\xi_2)$	$\pi_{AY}(\xi_2)$	$\kappa(\mathrm{A1}_5)$	$\kappa(\mathrm{A2}_5)$	$\kappa(\mathrm{A2}_5^*)$
2	10^3	86.60	87.04	36.00	52.84	67	87	9
	10^4	86.60	86.78	24.82	36.35	644	978	35
	10^5	86.60	86.66	17.45	25.34	6384	16627	235
3	10^3	91.06	91.61	63.29	83.69	100	112	9
	10^4	91.06	91.26	49.12	67.89	948	1061	44
	10^5	91.06	91.13	40.78	57.17	9399	17555	339
4	10^3	93.52	94.21	80.24	95.49	127	140	10
	10^4	93.52	93.74	70.89	89.16	1229	1375	54
	10^5	93.52	93.60	62.08	81.02	12245	15449	445
5	10^3	95.06	95.64	89.42	98.78	157	164	11
	10^4	95.06	95.29	85.33	97.54	1522	1667	64
	10^5	95.06	95.15	77.66	93.34	15173	17407	553
10	10^3	98.21	98.70	99.79	100.00	287	297	19
	10^4	98.21	98.39	99.61	100.00	2795	2920	113
	10^5	98.21	98.28	99.19	100.00	27925	29054	1080
20	10^3	99.54	99.74	100.00	100.00	488	499	39
	10^4	99.54	99.64	100.00	100.00	4799	4859	221
	10^5	99.54	99.58	100.00	100.00	47892	48562	2084
30	10^3	99.84	99.93	100.00	100.00	632	638	56
	10^4	99.84	99.89	100.00	100.00	6237	6276	329
	10^5	99.84	99.86	100.00	100.00	62624	62577	3053
40	10^3	99.93	99.98	100.00	100.00	738	740	72
	10^4	99.93	99.96	100.00	100.00	7286	7284	433

solution of the dual problem construct sequences of probability measures (defined by the Lagrange coefficients), which can thus straightforwardly be used to progressively eliminate inessential points. A two-point measure ξ_2, already proposed in the literature to efficiently initialise such dual algorithms [28], has been shown to efficiently directly remove a significant proportion of points in various situations with reasonably small dimension d. Several numerical experiments have indicated that this simple pre-filtering of the input set is clearly beneficial to a QP solver when enough inessential points are removed (d is small enough) and that the extra cost (slow-down factor) due to pre-filtering is marginal otherwise (for large d). Other methods, like those in [9, 26][4] might also benefit from the input-size reduction offered by this pre-filtering. Notice, finally, that these methods rely on the computation of a sequence of smallest enclosing balls for sets of $d+1$ points, from which a sequence of probability measures, and thus of eliminating inequalities, could easily be deduced; see [9, Sect. 3].

Acknowledgments

The author thanks the two referees for their comments that helped to improve the presentation of the paper. He also thanks the referee who pointed out the existence of reference [2].

[4]See also the implementation in `http://doc.cgal.org/latest/Bounding_volumes/classCGAL_1_1Min_sphere__d.html`

Table 8. X_i uniform in $\mathscr{B}_d(0,1)$: number N_ϵ of iterations performed to reach precision $\epsilon = 10^{-3}$ ($\epsilon = 10^{-6}$ for $A2_5^*$) — averaged values over 100 repetitions, rounded to the nearest integer.

d	n	A1	$A1_5$	A2	$A2_5$	$A2_5^*$
2	10^3	336	172	216	146	1764
	10^4	505	220	223	87	14658
	10^5	522	238	17	4	41055
3	10^3	363	182	189	144	5624
	10^4	530	219	270	167	18211
	10^5	522	216	231	69	50772
4	10^3	403	198	220	172	5222
	10^4	527	227	232	163	27298
	10^5	528	219	285	130	58736
5	10^3	405	207	240	191	5823
	10^4	546	240	228	178	23909
	10^5	528	219	257	144	60436
10	10^3	454	242	297	228	9757
	10^4	570	274	278	216	37559
	10^5	535	234	255	193	78289
20	10^3	474	262	364	273	26375
	10^4	589	292	363	276	80345
	10^5	539	237	310	235	133569
30	10^3	496	275	414	303	48385
	10^4	591	290	412	301	131301
	10^5	539	217	364	267	199827
40	10^3	508	284	437	316	64644
	10^4	595	278	426	315	174531

References

[1] S.D. Ahipaşaoğlu, P. Sun, and M.J. Todd. Linear convergence of a modified Frank-Wolfe algorithm for computing minimum-volume enclosing ellipsoids. *Optimization Mehods and Software*, 23:5–19, 2008.

[2] S.D. Ahipaşaoğlu and E.A. Yildirim. Identification and elimination of interior points for the minimum enclosing ball problem. *SIAM Journal on Optimization*, 19(3):1392–1396, 2008.

[3] C.L. Atwood. Sequences converging to D-optimal designs of experiments. *Annals of Statistics*, 1(2):342–352, 1973.

[4] M. Bădoiu and K.L. Clarkson. Optimal core-sets for balls. *Computational Geometry*, 40(1):14–22, 2008.

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

[6] K.M. Clarkson. Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. *ACM Transactions on Algorithms (TALG)*, 6(4):63, 2010.

[7] D.J. Elzinga and D.W. Hearn. The minimum covering sphere problem. *Manag. Sci.*, 19:96–104, 1972.

[8] V.V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.

[9] B. Gärtner. Fast and robust smallest enclosing balls. In J. Nesetril, editor, *Proc. 7th Annual European Symposium on Algorithms (ESA)*, pages 325–338. Springer, LNCS 1643, Berlin, 1999.

[10] D. Goldfarb and A. Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, 27(1):1–33, 1983.

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

[12] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

[13] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.

[14] A. Pázman. *Foundations of Optimum Experimental Design*. Reidel (Kluwer group), Dordrecht (co-pub. VEDA, Bratislava), 1986.

Table 9. X_i uniform in $\mathscr{B}_d(0,1)$: computational time t(QP) (in s) and ratios ρ(M)=t(M)/t(QP) — averaged over 100 repetitions. Italicized figures correspond to the fastest exact method.

d	n	t(QP)	QP$_0$	A1	A1$_5$	A2	A2$_5$	A1$'_5$-QP	A2$'_5$-QP	A2$^*_5$
2	10^3	0.004	*0.74*	16.20	9.06	10.15	6.83	9.64	7.14	73.0
	10^4	0.028	*0.34*	8.69	1.96	3.75	0.72	2.10	0.88	81.2
	10^5	0.28	*0.24*	9.37	0.61	0.25	0.07	0.68	*0.24*	24.1
3	10^3	0.006	*0.88*	12.40	6.55	6.05	4.47	6.63	4.78	146.2
	10^4	0.031	*0.59*	9.09	2.18	4.44	1.34	2.38	1.51	92.2
	10^5	0.31	0.51	9.77	0.93	3.43	0.25	1.04	*0.44*	29.1
4	10^3	0.007	*0.99*	12.75	6.73	6.53	4.89	6.89	5.28	144.9
	10^4	0.035	*0.85*	8.03	2.27	3.43	1.37	2.48	1.58	120.2
	10^5	0.34	0.77	10.47	1.35	4.57	0.47	1.50	*0.65*	33.4
5	10^3	*0.008*	1.02	11.68	6.37	6.63	4.93	6.66	5.53	126.2
	10^4	0.039	*0.99*	8.36	2.44	3.27	1.55	2.67	1.80	98.8
	10^5	0.37	*0.91*	11.54	1.82	4.48	0.74	1.99	0.93	34.5
10	10^3	*0.012*	1.05	9.10	5.04	6.07	4.31	6.04	5.17	134.5
	10^4	*0.059*	1.06	7.62	3.02	3.54	2.08	3.46	2.58	110.7
	10^5	*0.54*	1.10	15.55	4.04	6.09	2.34	4.36	2.66	49.9
20	10^3	*0.025*	1.08	5.29	3.15	4.02	2.88	4.06	3.91	173.4
	10^4	*0.11*	1.05	11.52	4.22	6.85	3.45	4.90	4.16	148.5
	10^5	*0.96*	1.08	16.96	6.74	8.44	5.10	7.24	5.59	112.9
30	10^3	*0.049*	1.03	3.17	1.94	2.59	1.97	3.06	3.00	165.9
	10^4	*0.17*	1.06	11.72	5.62	7.78	5.14	6.45	5.97	191.2
	10^5	*1.65*	1.09	17.83	7.39	10.77	7.01	7.97	7.61	232.3
40	10^3	*0.087*	1.02	2.03	1.29	1.68	1.27	2.34	2.33	132.2
	10^4	*0.27*	1.05	11.50	5.53	7.58	5.32	6.42	6.19	204.7

[15] L. Pronzato. A delimitation of the support of optimal designs for Kiefer's ϕ_p-class of criteria. *Statistics & Probability Letters*, 83:2721–2728, 2013.

[16] L. Pronzato. Minimax and maximin space-filling designs: some properties and methods for construction. *Journal de la Société Française de Statistique*, 158(1):7–36, 2017.

[17] L. Pronzato and A. Pázman. *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties.* Springer, LNS 212, New York, 2013.

[18] F. Pukelsheim. *Optimal Experimental Design.* Wiley, New York, 1993.

[19] R. Sibson. Discussion on a paper by H.P. Wynn. *Journal of Royal Statistical Society*, B34:181–183, 1972.

[20] S.D. Silvey and D.M. Titterington. A geometric approach to optimal design theory. *Biometrika*, 60(1):21–32, 1973.

[21] D.M. Titterington. Optimal design: some geometrical espects of D-optimality. *Biometrika*, 62(2):313–320, 1975.

[22] D.M. Titterington. Algorithms for computing D-optimal designs on a finite design space. In *Proc. of the 1976 Conference on Information Science and Systems*, pages 213–216, Baltimore, 1976. Dept. of Electronic Engineering, John Hopkins University.

[23] D.M. Titterington. Estimation of correlation coefficients by ellipsoidal trimming. *Journal of Royal Statistical Society*, C27(3):227–234, 1978.

[24] M.J. Todd. *Minimum-Volume Ellipsoids. Theory and Algorithms.* SIAM, Philadelphia, 2016.

[25] M.J. Todd and E.A. Yildirim. On Khachiyan's algorithm for the computation of minimum volume enclosing ellipsoids. *Discrete Applied Math.*, 155:1731–1744, 2007.

[26] E. Welzl. Smallest enclosing disks (balls and ellipsoids). In *Lecture Notes in Computer Science, vol. 555*, pages 359–370. Springer, Heidelberg, 1991.

[27] C.F.J. Wu. Some algorithmic aspects of the theory of optimal designs. *Annals of Statistics*, 6(6):1286–1301, 1978.

[28] E.A. Yildirim. Two algorithms for the minimum enclosing ball problem. *SIAM Journal on Optimization*, 19(3):1368–1391, 2008.

[29] Y. Yu. Monotonic convergence of a general algorithm for computing optimal designs. *Annals of Statistics*, 38(3):1593–1606, 2010.

Appendix A.

LEMMA A.1 *For any set $\mathscr{X}_n \subset \mathbb{R}^d$ and any $X_1 \in \mathscr{X}_n$, the measure ξ_2 defined by (26) is such that*

$$\max_{i=1,\ldots,n} \|X_i - c(\xi_2)\| \leq \frac{\sqrt{7}}{2} \|X_{i_1} - X_{i_2}\| , \tag{A1}$$

and the bound is tight.

Proof. Let X_{i^*} denote the furthest point in $\mathscr{X}_n$ from $c(\xi_2)$. Without any loss of generality we can take X_{i_1} as the origin, $X_{i_2} - X_{i_1}$ proportional to the first basis vector e_1, X_1 in the plane formed by e_1 and e_2, and X_{i^*} in the three dimensional space generated by e_1, e_2 and e_3. With a suitable rescaling, we can also assume that $\|X_{i_1} - X_{i_2}\| = 1$. Therefore, we can write, in polar coordinates,

$$X_{i_1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; X_{i_2} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; X_1 = r \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } X_{i^*} = \rho \begin{pmatrix} \cos(\varphi)\cos(\omega) \\ \cos(\varphi)\sin(\omega) \\ \sin(\varphi) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{A2}$$

for some r and ρ in $\mathbb{R}^+$, $\theta \in [0, \pi]$, $\omega \in [0, 2\pi]$ and $\varphi \in [-\pi/2, \pi/2]$. Then, since X_{i_2} is the furthest point in $\mathscr{X}_n$ from X_{i_1}, $\mathscr{X}_n \subset \mathscr{B}_d(0, 1)$ and $r, \rho < 1$. Also, $\mathscr{X}_n \subset \mathscr{B}_d(X_1, r)$ since X_{i_1} is the furthest point in $\mathscr{X}_n$ from X_1, and $\|X_1 - X_{i_2}\| \leq r$ gives $2r\cos(\theta) \geq 1$, which implies $\theta \leq \pi/3$. Direct calculation shows that $\|X_1 - X_{i^*}\| \leq r$ yields the constraint

$$\rho \leq 2r\cos(\varphi)\cos(\omega - \theta) . \tag{A3}$$

We also obtain $\|X_{i^*} - c(\xi_2)\|^2 = \rho^2 - \rho\cos(\varphi)\cos(\omega) + 1/4$, which we want to maximise with respect to r, ρ, θ, ω and φ satisfying all the constraints above. The value of $\|X_{i^*} - c(\xi_2)\|^2$ will be maximum when $r = 1$ in (A3). Simple calculation shows that one cannot consider the two constraints (i) $\rho \leq 1$ and (ii) $\rho \leq 2\cos(\varphi)\cos(\omega - \theta)$ separately: when using (i) only, $\|X_{i^*} - c(\xi_2)\|^2$ is maximum at $\rho = 1$, $\varphi = 0$ and $\omega = \pi$, which violates (ii) since $\cos(\omega - \theta) = -\cos(\theta) < 0$; when using (ii) only, the maximum is when the constraint is saturated, with $\varphi = 0$, $\theta = \pi/3$ and $\omega = 5\pi/12$, but then $\rho = 2\cos(\pi/12) > 1$. Enforcing (i) and (ii) simultaneously, we obtain the solution $\varphi = 0$, $\theta = \pi/3$ and $\omega = 2\pi/3$, showing that $\|X_{i^*} - c(\xi_2)\|^2 \leq 7/4$. The bound (A1) is tight since equality is obtained when setting $r = \rho = 1$, $\varphi = 0$, $\theta = \pi/3$ and $\omega = 2\pi/3$ in (A2). ■