[18] K. E. Tikhomirov. On the distance of polytopes with few vertices to the euclidean ball. *Discrete Comput. Geom.*, to appear.
[19] L. N. Vicente. Worst case complexity of direct search. *Euro J. Comput. Optim.*, 1:143–153, 2013.

M. Dodangeh
Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal (dodangeh@mat.uc.pt).

L. N. Vicente
CMUC, Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal (lnv@mat.uc.pt).

Z. Zhang
CERFACS-IRIT joint lab, 31000 Toulouse, France (zaikun.zhang@irit.fr).

Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 14–38

ON THE OPTIMAL ORDER OF WORST CASE COMPLEXITY OF DIRECT SEARCH

M. DODANGEH, L. N. VICENTE AND Z. ZHANG

Abstract: It was recently analyzed the worst case complexity of direct-search methods based on positive spanning sets and sufficient decrease for acceptance of new iterates. Assuming that the objective function is smooth, it is now known that such methods require at most $\mathcal{O}(n^2\epsilon^{-2})$ function evaluations to compute a gradient of norm below $\epsilon \in (0,1)$, where n is the dimension of the problem. Such a maximal effort is reduced to $\mathcal{O}(n^2\epsilon^{-1})$ if the function is convex. The factor n^2 has been derived using the positive spanning set formed by the coordinate vectors and their negatives at all iterations.

In this paper, we prove that such a factor of n^2 is optimal in these worst case complexity bounds, in the sense that no other positive spanning set will yield a better order of n. The proof is based on an observation that reveals the connection between cosine measure in positive spanning and sphere covering.

Keywords: Direct search, worst case complexity, optimal order, sphere covering, positive spanning set, cosine measure.
AMS Subject Classification (2010): 90C56, 90C25, 90C30.

1. Introduction

In this paper, we focus on direct-search methods of directional type applied to an unconstrained minimization problem

$$\min_{x\in\mathbb{R}^n} f(x).$$

At each iteration of such a method, one typically evaluates the objective function using a finite number of directions (the polling directions) multiplied by a certain step size. No derivatives or approximations thereof are required. If a point of lower objective function value is found, it becomes the new iterate (the iteration is declared successful). Otherwise, the method stays at the current point and the step size is decreased.

Date: November 14, 2014.
Support for this research was provided by FCT under grants PTDC/MAT/116736/2010 and PEst-C/MAT/UI0324/2011. Support for the first author was provided by FCT under the scholarship SFRH/BD/51168/2010. The third author works now within the framework of the project FILAOS funded by RTRA STAE.

There are essentially two ways of ensuring convergence to stationary points for direct-search methods of directional type: (i) by exploring integer lattices, insisting on generating points in grids or meshes (which refine only with the decrease of the step size), or (ii) by imposing a sufficient decrease condition (specified by a forcing function of the step size) to the acceptance of a new iterate. To our knowledge, only the latter approach has led to analyses of worst case complexity (WCC).

In fact, Vicente [19] analyzed, for smooth functions, the WCC of direct-search methods for which the polling directions are formed by positive spanning sets and new iterates are accepted based on a sufficient decrease condition. A positive spanning set (PSS) is a set of non-zero directions that spans $\mathbb{R}^n$ with non-negative coefficients. The WCC bounds given there depend on the choice of the positive spanning set used at each iteration and on the forcing function specifying the sufficient decrease. When the forcing function is the square of the step size, it was proved in [19] that direct search drives the norm of the gradient of f below $\epsilon \in (0,1)$ within at most $\mathcal{O}(n^2\epsilon^{-2})$ function evaluations. In the convex case, Dodangeh and Vicente [9] improved the bound in [19] to $\mathcal{O}(n^2\epsilon^{-1})$. Konečný and Richtárik [16] rederived these bounds, in a simpler way, but for a restricted version of direct search where step size increases are not allowed.

A closer look at these WCC bounds reveals that they are of the form $\mathcal{O}(m\kappa^{-2}\epsilon^{-a})$, with $a = 1, 2$, where m is an upper bound on the cardinal of the PSSs and κ is a lower bound for their cosine measure. The factor n^2 results from using the PSS $D_\oplus = [I \ -I]$ formed by the coordinate vectors and their negatives at all iterations. In fact, $D_\oplus$ has $2n$ directions and a cosine measure of $1/\sqrt{n}$, see [15].

In this paper, we will show that there exists a universal constant $c > 0$ such that $m\kappa^{-2} \geq c\, n$, and consequently that the optimal order for these two WCC bounds in terms of n is indeed n^2. Our analysis is based on the following observation connecting cosine measure in positive spanning and the topic of sphere covering in Discrete Geometry (see Section 4 for more details): Given a PSS with normalized vectors, the unit sphere is covered by a union of spherical caps centered at such vectors and of angle equal to the one defined by the cosine measure of the PSS.

The structure of the paper is as follows. In Section 2, we introduce the algorithmic details of direct search. Section 3 summarizes the WCC bounds established in [19, 9]. Our main result is given in Section 4, where we prove

Instead of PSSs, Gratton *et al.* [14] propose to use random polling directions in Algorithm 2.1. When minimizing a smooth (possibly non-convex) objective function, the resulting algorithm enjoys an $\mathcal{O}(mn\epsilon^{-2})$ WCC bound for the number of function evaluations (with overwhelmingly high probability), where m is the number of random directions used in each poll step. Such a bound, when m is much smaller than n, is better than the optimal situation that we already proved for the bound (4). The advantage was also observed in the numerical results of [14]. This comparison, more rigorously made given the contribution of our paper, suggests that randomization has the potential to improve the efficiency of some classical algorithms.

References

[1] I. Bárány and Z. Füredi. Approximation of the sphere by polytopes having few vertices. *Proc. Amer. Math. Soc.*, 102:651–659, 1988.

[2] K. Böröczky, Jr. *Finite Packing and Covering*. Cambridge University Press, New York, 2004.

[3] K. Böröczky, Jr. private communication, April 25, 2014.

[4] K. Böröczky, Jr. and G. Wintsche. Covering the sphere by equal spherical balls. In B. Aronov, S. Basu, J. Pach, and M. Sharir, editors, *Discrete Comput. Geom.*, volume 25 of *Algorithms and Combinatorics*, pages 235–251. Springer Berlin, 2003.

[5] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162:73–141, 1989.

[6] B. Carl and A. Pajor. Gelfand numbers of operators with values in a hilbert space. *Invent. Math.*, 94:479–504, 1988.

[7] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[8] L. Dalla, D. G. Larman, P. Mani-Levitska, and C. Zong. The blocking numbers of convex bodies. *Discrete Comput. Geom.*, 24:267–278, 2000.

[9] M. Dodangeh and L. N. Vicente. Worst case complexity of direct search under convexity. Technical Report 13-10, Dept. Mathematics, Univ. Coimbra, 2013.

[10] E. D. Dolan, R. M. Lewis, and V. Torczon. On the local convergence of pattern search. *SIAM J. Optim.*, 14:567–583, 2003.

[11] L. Fejes Tóth. *Regular Figures*. Pergman Press, London, 1964.

[12] O. Fercoq and P. Richtárik. Accelerated, parallel and proximal coordinate descent. Technical Report arxiv:1312.5799v2, School of Mathematics, University of Edinburgh, December, 2013.

[13] E. D. Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces. *Mathematics of the USSR-Sbornik*, 64:85–96, 1989.

[14] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Direct search based on probabilistic descent. Technical Report 14-11, Dept. Mathematics, Univ. Coimbra, 2014.

[15] T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45:385–482, 2003.

[16] J. Konečný and P. Richtárik. Simple complexity analysis of simplified direct search. Technical Report arXiv:1410.0390v2, School of Mathematics, University of Edinburgh, November, 2014.

[17] Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM J. Optim.*, 22:341–362, 2012.

With the help of Lemmas 4.1 and 4.2, we obtain the desired upper bound for the cosine measure.

Lemma 4.3. *Any PSS D in $\mathbb{R}^n$ satisfies*

$$\mathrm{cm}(D) \;\leq\; \zeta \sqrt{n^{-1}\log(n^{-1}|D|)} \tag{10}$$

for the same constant ζ as in Lemma 4.2.

Proof: Without loss of generality, we assume that all the directions in D are normalized. Then inequality (10) follows immediately from Lemmas 4.1 and 4.2. ∎

Theorem 4.1 is a straightforward consequence of Lemma 4.3, since

$$\sqrt{n^{-1}\log(n^{-1}|D|)} \;\leq\; \sqrt{n^{-1}(n^{-1}|D|-1)} \;\leq\; n^{-1}|D|^{\frac{1}{2}}.$$

Summarizing, in terms of the orders of n and ϵ, the right-hand sides of (4) and (5) are optimal when setting $p = 2$ (a fact already known) and when $D_k = D_\oplus$ for all $k \geq 0$ (a fact directly resulting from Theorem 4.1).

Theorem 4.2. *With the same constant $c > 0$ as in Theorem 4.1, we have*

$$m\kappa^{-\hat{p}}\epsilon^{-\hat{p}} \;\geq\; c\, n^{-2}\epsilon^{-2},$$

where m and κ are defined in Assumption 3.3, and $\hat{p} = p/\min\{1, p-1\}$ $(p > 1)$.

5. Final remarks

In Section 4, we established a lower bound for the optimal value of problem (6), but this problem itself is still open. A closely related problem is

$$\max_{D\in\mathcal{D}(m)} \mathrm{cm}(D), \tag{11}$$

where $\mathcal{D}(m)$ is the set of all the PSSs consisting of m directions in $\mathbb{R}^n$ $(m \geq n+1)$. This problem is also widely open. In the language of sphere covering, problem (11) is to find the most "economical" covering of the sphere by m equal spherical caps. In the special case of $m = 2n$, it is intuitive to conjecture that $D_\oplus$ is the solution to problem (11). This is clear when $n = 2$, but it becomes non-trivial when $n \geq 3$. The optimality of $D_\oplus$ for the case $m = 2n$ is already proved when $n = 3$ (see [11, Theorem 5.4.1]) and $n = 4$ (see [8, Theorem 6.7.1]), but it is open when $n \geq 5$ according to [3] (see also [2, Page 194] and [4, Conjecture 1.3]).

the optimality of the factor involving n in the bounds $\mathcal{O}(n^2\epsilon^{-a})$, with $a = 1, 2$. The paper is concluded with some remarks in Section 5.

2. A class of direct-search methods

First we describe the direct-search method under analysis. Algorithm 2.1 below follows the lines of the presentation in [7, Chapter 7].

Algorithm 2.1 (Directional direct-search method).

Initialization

Choose x_0 with $f(x_0) < +\infty$, $\alpha_0 > 0$, $0 < \beta_1 \leq \beta_2 < 1$, and $\gamma \geq 1$.

For $k = 0, 1, 2, \ldots$ **do**

(1) **Search step:** Try to compute a point with $f(x) < f(x_k) - \rho(\alpha_k)$ by evaluating the function f at a finite number of points. If such a point is found, then set $x_{k+1} = x$, declare the iteration and the search step successful, and skip the poll step.

(2) **Poll step:** Choose a positive spanning set D_k. Order the set of poll points $P_k = \{x_k + \alpha_k d : d \in D_k\}$. Evaluate f at the poll points following the chosen order. If a poll point $x_k + \alpha_k d_k$ is found such that $f(x_k + \alpha_k d_k) < f(x_k) - \rho(\alpha_k)$, then stop polling, set $x_{k+1} = x_k + \alpha_k d_k$, and declare the iteration and the poll step successful. Otherwise, declare the iteration (and the poll step) unsuccessful and set $x_{k+1} = x_k$.

(3) **Mesh parameter update:** If the iteration was successful, then maintain or increase the step size parameter: $\alpha_{k+1} \in [\alpha_k, \gamma\alpha_k]$. Otherwise, decrease the step size parameter: $\alpha_{k+1} \in [\beta_1\alpha_k, \beta_2\alpha_k]$.

End

In the poll step, the objective function is evaluated at points of the form $x_k + \alpha_k d$ for directions d belonging to a PSS D_k. The poll step is successful if the value of the objective function can be sufficiently decreased relatively to the step size α_k, in the sense that there exists a $d_k \in D_k$ such that $f(x_k+\alpha_k d_k) < f(x_k)-\rho(\alpha_k)$, where $\rho(\cdot)$ is a forcing function [15], meaning a non-decreasing (typically continuous) function $\rho : (0,\infty) \to (0,\infty)$ satisfying $\lim_{\alpha\downarrow 0} \rho(\alpha)/\alpha = 0$. Typical examples of forcing functions are $\rho(\alpha) = \mathcal{C}\alpha^p$ ($\mathcal{C} > 0$ and $p > 1$). Polling can be complete (where the point with the lowest function value is then chosen) or opportunistic (moving to the first point satisfying the sufficient decrease condition).

The purpose of the search step is solely to improve the practical performance of the overall algorithm. It is left unspecified since it does not interfere in the analyses of convergence or complexity of the algorithm. It is the poll step that determines these properties and gives a directional character to the algorithm. The step size is possibly increased in successful iterations and decreased at unsuccessful iterations by a factor uniformly bounded away from 0 and 1.

3. Worst case complexity of direct search

To analyze the WCC of direct search, one needs to make some assumptions on the objective function f, the forcing function ρ, and the positive spanning sets $\{D_k\}$. For convenience, we use $L(x_0)$ to denote the level set $\{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$, X_*^f to denote the set of all the minimizers of f (possibly empty), and $|\cdot|$ to denote the cardinal of a set. The assumptions on f and ρ are as follows.

Assumption 3.1. *f is bounded from below by $f_{\inf} > -\infty$ and continuously differentiable in $\mathbb{R}^n$, and ∇f is Lipschitz continuous in $\mathbb{R}^n$ with a Lipschitz constant ν.*

Assumption 3.2. *$\rho(\alpha) = \mathcal{C}\alpha^p$ for some constants $\mathcal{C} > 0$ and $p > 1$.*

To state the assumptions on $\{D_k\}$, one must look at a key feature of a PSS, its cosine measure [15]. Given a PSS D, its cosine measure is defined by

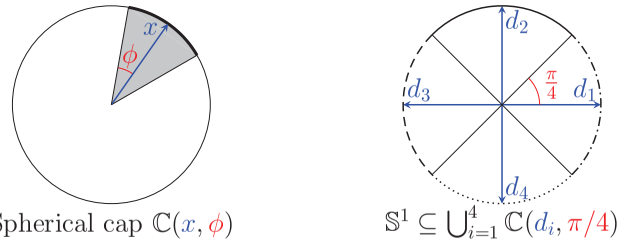
$$\mathrm{cm}(D) \;=\; \min_{0 \neq v \in \mathbb{R}^n} \max_{d \in D} \frac{v^\top d}{\|v\|\|d\|}. \tag{1}$$

Since any PSS has a positive cosine measure, there is at least one direction in D making an acute angle with any considered non-zero vector. Such a property is at the heart of the fact that the norm of the gradient is of the order of the step size when an unsuccessful iteration occurs [10, 15] (see also [7, Theorem 2.4 and inequality (7.14)]), which is critical for analyzing global convergence and WCC of direct search. The assumption on $\{D_k\}$ can be presented as follows.

Assumption 3.3. *There exist positive constants κ, m, μ_1, and μ_2 such that*

$$\mathrm{cm}(D_k) \geq \kappa, \quad |D_k| \leq m, \quad \text{and} \quad \mu_1 \leq \|d\| \leq \mu_2$$

for each $k \geq 0$ and each $d \in D_k$.



Spherical cap $\mathbb{C}(x, \phi)$ $\quad$ $\mathbb{S}^1 \subseteq \bigcup_{i=1}^4 \mathbb{C}(d_i, \pi/4)$

FIGURE 2. The thick arc of the left picture represents a spherical cap in $\mathbb{R}^2$. The right picture depicts the covering of the unit sphere given in Lemma 4.1 for $D_\oplus$ when $n = 2$.

We remark that Lemma 4.1 holds even when D is not a PSS. In such a case, $\mathrm{cm}(D)$ is not positive and thus $\arccos(\mathrm{cm}(D))$ not acute, but the covering is still a valid one. We also notice that $\mathrm{cm}(D)$ is actually the largest κ that satisfies (13). In other words,

$$\mathrm{cm}(D) \;=\; \max\left\{\kappa : \mathbb{S}^{n-1} \;\subseteq\; \bigcup_{i=1}^{m} \mathbb{C}(d_i, \arccos(\kappa))\right\},$$

which can be seen as an equivalent definition of cosine measure using the language of sphere covering.

To prove Theorem 4.1 we need to establish an appropriate upper bound for the cosine measure $\mathrm{cm}(D)$ in terms of the dimension n and the number of directions in D. In light of Lemma 4.1, that is equivalent to find a lower bound for ϕ (or an upper bound for $\cos(\phi)$) in terms of n and m when $\mathbb{S}^{n-1}$ is covered by m equal spherical caps with geodesic radius ϕ. Such a bound is fortunately already established in the research community of *Discrete Geometry*. The conclusion of the following lemma is proved by Tikhomirov [18] for $n + 1 \leq m \leq 2n$. The case $m \geq 2n$ was established much earlier (for more details see [1], [2, Chapter 6], [4], [5, Corollary 9.5], [6], [13] and [18])[1].

Lemma 4.2 ([18])**.** *Any covering of $\mathbb{S}^{n-1}$ by $m \geq n + 1$ spherical caps of geodesic radius ϕ satisfies*

$$\cos(\phi) \;\leq\; \zeta\sqrt{n^{-1}\log(n^{-1}m)}$$

for some universal constant $\zeta > 0$.

[1]We are grateful to Professor Károly Böröczky, Jr. for drawing our attention to these references.

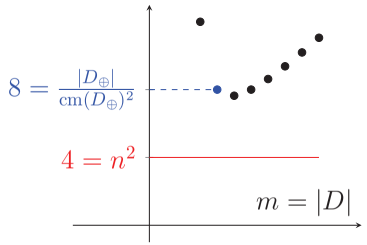


FIGURE 1. Illustration of inequality (8) in $\mathbb{R}^2$ for PSSs making uniform adjacent angles. The plot depicts the values of $|D|/\,\mathrm{cm}(D)^2$ for $|D| = 3, 4, \ldots, 10$.

where $\mathbb{S}^{n-1}$ is the unit sphere in $\mathbb{R}^n$, x is a fixed point on $\mathbb{S}^{n-1}$, ϕ is a constant in $[0, \pi]$, and $\mathrm{d}(\cdot, \cdot)$ is the geodesic distance on $\mathbb{S}^{n-1}$. We will call $\mathbb{C}(x, \phi)$ a spherical cap centered at x with geodesic radius ϕ (see Figure 2). The result says essentially that for any PSS (with normalized vectors) one can cover the unit sphere by the union of the spherical caps centered at its vectors and of geodesic radius equal its cosine measure (see also Figure 2).

Lemma 4.1. *Suppose that $D = [d_1 \cdots d_m]$ is a PSS in $\mathbb{R}^n$ consisting of unit vectors. If $\mathrm{cm}(D) = \kappa$, then*

$$\mathbb{S}^{n-1} \subseteq \bigcup_{i=1}^{m} \mathbb{C}(d_i, \arccos(\kappa)).$$

In other words, $\mathbb{S}^{n-1}$ is covered by the spherical caps centered at d_i ($i = 1, 2, \ldots, m$) with geodesic radius $\arccos(\kappa)$.

Proof: According to definition (1), for any $v \in \mathbb{S}^{n-1}$, there exists an $i \in \{1, 2, \ldots, m\}$ such that

$$v^\top d_i \geq \mathrm{cm}(D) = \kappa. \tag{9}$$

Since the radius of the sphere is 1, the geodesic distance between v and d_i is equal to the angle between them. Hence inequality (9) implies that

$$\mathrm{d}(v, d_i) \leq \arccos(\kappa),$$

which is equivalent to

$$v \in \mathbb{C}(d_i, \arccos(\kappa)).$$

This is sufficient to conclude the proof as v is arbitrary. ∎

Notice that a poll step takes at most m function evaluations. We suppose that the number of function evaluations per search step is at most a multiple of m, which is not restrictive because this step is optional.

Suppose that $\epsilon \in (0, 1)$. Let

$$k_\epsilon = \min\{k \in \mathbb{N} : \|\nabla f(x_k)\| \leq \epsilon\}.$$

Then the worst case complexity bounds obtained by Vicente [19] and Dodangeh and Vicente [9] can be stated as follows.

Theorem 3.1 ([19, Theorems 2 and 3], [9, Theorem 4.2])**.** *Under Assumptions 3.1, 3.3, and 3.2, there exists a constant c_1 such that*

$$k_\epsilon \leq c_1 \kappa^{-\hat{p}} \epsilon^{-\hat{p}}, \tag{2}$$

where $\hat{p} = p/\min\{1, p-1\}$.

If we assume additionally that f is convex, X_*^f is nonempty, and $\sup_{y \in L(x_0)} \mathrm{dist}(y, X_*^f) < \infty$, then there exists a constant c_2 such that

$$k_\epsilon \leq c_2 \kappa^{-\hat{p}} \epsilon^{1-\hat{p}}. \tag{3}$$

The constants c_1 and c_2 are fully determined by $f(x_0)$, $f_{\inf}$, ν, α_0, β_1, β_2, γ, $\mathcal{C}$, p, μ_1, μ_2, and, in the case of c_2, by $\sup_{y \in L(x_0)} \mathrm{dist}(y, X_*^f)$. For the definitions of c_1 and c_2, we refer to [19, 9].

It is noticeable that κ appears in the bounds (2) and (3). The reason is that, in the worst case, the smallest angle between $-\nabla f(x_k)$ and the directions in D_k is $\arccos(\kappa)$. In other words, κ can be interpreted as the price to pay for the absence of gradient information. When $p = 2$, the factor in κ is of the form κ^{-2} which is then equal to n if we use the PSS $D_\oplus$.

A similar phenomenon happens in the WCC of stochastic coordinate descent methods (see, for instance, [12, 17]). In Nesterov [17], the randomized coordinate descent method (RCDM) has an $\mathcal{O}(n_b k^{-1})$ global decaying rate for the expectation of $f(x_k) - f_*$ (see (2.14) of [17], where n_b is the number of blocks in [17]). This is because, on average, the angle between $-\nabla f(x_k)$ and $\mathcal{S}_k$ is $\arccos(1/n_b)$, where $\mathcal{S}_k$ denotes the subspace corresponding to the coordinate block chosen at the k-th iteration. Thus, n_b can be seen here as the price to pay for the absence of the full gradient. Although the methods are different, the factor $\kappa^{-\hat{p}}$ in (2) and (3) (which is equal to n when we use $p = 2$ and $D_\oplus$ at each iteration) and the factor n_b in the global rate of RCDM appear for similar reasons. It is possible to reduce such a price

(by putting more directions into D_k for direct search; by using less blocks in RCDM), which will potentially lead to lower complexity of iterations, but meanwhile increase the expense of each iteration. An important question is how to maintain such a balance in order to have the algorithms exhibiting optimal overall measures of performance? We will answer this question for direct search in Section 4. To do this, we consider the WCC in terms of the number of function evaluations, which is a reasonable indicator for the overall performance of direct search.

Let k_ϵ^f be the number of function evaluations within k_ϵ iterations. Then it is easy to derive bounds for k_ϵ^f from Theorem 3.1.

Corollary 3.1. *Under Assumptions 3.1, 3.3, and 3.2, there exists a constant c_3 such that*

$$k_\epsilon^f \ \leq \ c_3 m \kappa^{-\hat{p}} \epsilon^{-\hat{p}}. \tag{4}$$

If we assume additionally that f is convex, X_*^f is nonempty, and $\sup_{y\in L(x_0)} \operatorname{dist}(y, X_*^f) < \infty$, then there exists a constant c_4 such that

$$k_\epsilon^f \ \leq \ c_4 m \kappa^{-\hat{p}} \epsilon^{1-\hat{p}}. \tag{5}$$

As pointed out in [19] it is obvious that the optimal choice for p in (4) is 2, which leads to $\hat{p} = 2$ (and the same happens in (5)). Thus let us fix $p = 2$ henceforth.

It was also suggested in [19] to set $D_k = D_\oplus$ for all $k \geq 0$. Since $|D_\oplus| = 2n$ and $\operatorname{cm}(D_\oplus) = 1/\sqrt{n}$, inequality (4) then becomes $k_\epsilon^f = \mathcal{O}(n^2\epsilon^{-2})$. In Section 4 (see Theorem 4.2), we will show that $\mathcal{O}(n^2\epsilon^{-2})$ is the optimal order for the bound (4) and a similar conclusion holds for (5), whose optimal order is $\mathcal{O}(n^2\epsilon^{-1})$.

4. The optimal order of the worst case complexity

In this section, we will discuss how to choose $\{D_k\}$ so that the right-hand sides of (4) and (5) are minimized. As mentioned in Section 3, one can increase κ by using more polling directions (see definition (1)), which is favorable in terms of number of iterations (see Theorem 3.1). However, such a strategy will increase the number of function evaluations at each iteration (in the worst-case scenario). Hence there is a trade-off in the number of directions to use in the PSSs when we consider the bounds for the number of function evaluations.

Let us illustrate this trade-off with a few examples. We know that a minimal positive basis with uniform angles in $\mathbb{R}^n$ has $n+1$ directions and a cosine measure of $1/n$ [7, Corollary 2.6 and Exercise 2.7.7]. If we apply this PSS to Algorithm 2.1 (with $\rho(\alpha) = \mathcal{C}\alpha^2$), then the bound in (4) will become $\mathcal{O}(n^3\epsilon^{-2})$. But if we use $D_\oplus$, the bound is $\mathcal{O}(n^2\epsilon^{-2})$. In this case, increasing the number of directions improves the WCC bound. This is not always true. For example, if we use a PSS with n^4 directions, then the bound will become at least $\mathcal{O}(n^4\epsilon^{-2})$.

According to the definitions of κ and m, in order to minimize the right-hand sides of (4) and (5) (notice that $\hat{p} = 2$), we need to solve

$$\min_{D\in\mathcal{D}} \frac{|D|}{\operatorname{cm}(D)^2}, \tag{6}$$

where $\mathcal{D}$ is the set of all PSSs in $\mathbb{R}^n$. In two dimensions, problem (6) is not difficult, and one can easily prove that the PSSs with five directions and uniform $2\pi/5$ adjacent angles are optimal. When we go to higher dimensions, determining the optimal PSSs for problem (6) is not that easy. It is not clear what are the solutions to this problem when $n \geq 3$. But we are able to show that $D_\oplus$ is "almost optimal" for problem (6) in the sense of that

$$\min_{D\in\mathcal{D}} \frac{|D|}{\operatorname{cm}(D)^2} \ \geq \ c_5 \frac{|D_\oplus|}{\operatorname{cm}(D_\oplus)^2}, \tag{7}$$

for some constant $c_5 > 0$ not depending on n or on any specific PSS. In fact, we will prove the following result.

Theorem 4.1. *There exists a universal constant $c > 0$ such that*

$$\frac{|D|}{\operatorname{cm}(D)^2} \ \geq \ c\, n^2 \tag{8}$$

for each $n \geq 1$ and each PSS D in $\mathbb{R}^n$.

Inequality (7) will then follow directly from (8) with $c_5 = c/2$. We illustrate inequality (8) in Figure 1 for the particular case where $n = 2$ and the vectors of the PSSs make uniform adjacent angles (a case where it is known that one can pick $c = 1$).

To prove Theorem 4.1, we first observe in Lemma 4.1 a connection between cosine measure in positive spanning and sphere covering. For that purpose, let us define

$$\mathbb{C}(x, \phi) = \left\{ y \in \mathbb{S}^{n-1} : \mathrm{d}(y, x) \leq \phi \right\},$$