# A HYBRID DIRECT SEARCH AND PROJECTED SIMPLEX GRADIENT METHOD FOR CONVEX CONSTRAINED MINIMIZATION

A. L. CUSTÓDIO*, E. H. M. KRULIKOVSKI†, AND M. RAYDAN‡

*Dedicated to the memory of our friend Oleg Burdakov*

**Abstract.** We propose a new Derivative-free Optimization (DFO) approach for solving convex constrained minimization problems. The feasible set is assumed to be the nonempty intersection of a finite collection of closed convex sets, such that the projection onto each of these individual convex sets is simple and inexpensive to compute. Our iterative approach alternates between steps that use Directional Direct Search (DDS), considering adequate poll directions, and a Spectral Projected Gradient (SPG) method, replacing the real gradient by a simplex gradient, under a DFO approach. In the SPG steps, if the convex feasible set is simple, then a direct projection is computed. If the feasible set is the intersection of finitely many convex simple sets, then Dykstra's alternating projection method is applied. Convergence properties are established under standard assumptions usually associated to DFO methods. Some preliminary numerical experiments are reported to illustrate the performance of the proposed algorithm, in particular by comparing it with a classical DDS method. Our results indicate that the hybrid algorithm is a robust and effective approach for derivative-free convex constrained optimization.

**Key words.** Derivative-free Optimization, Convex Constrained Optimization, Directional Direct Search, Spectral Projected Gradient, Simplex Gradient, Dykstra's Algorithm

## 1. Introduction.

In this work, we consider constrained optimization problems of the form

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a continuously differentiable function on an open set that contains $\Omega$, and $\Omega \subset \mathbb{R}^n$ is a nonempty, closed, convex set. Throughout this work we assume that $\Omega$ is either a closed convex simple set (i.e., it is easy to project onto it, such is the case of boxes, spheres, or half-spaces) or it can be obtained as the nonempty intersection of a finite collection of closed and convex simple sets. Although the objective function is smooth, we assume that the derivatives of $f$ are not available and cannot be numerically approximated, due to the expensive nature of function evaluation (see [11, 25]).

Derivative-free Optimization (DFO) problems appear in many applications, related to different scientific areas. In [5], the authors addressed the optimization of molecular geometries. Thermal insulation systems were the focus of the work reported in [1]. Many other examples could be provided in such different areas as Aerospace Engineering [40], Nanotechnology [48], or Medicine [13] (see also the recent survey [4]). Such practical applications have contributed to the development of research on DFO methods. For a first introduction to the subject see [11, 25] and the references therein.

In general nonlinear optimization problems, penalty methods or augmented Lagrange multipliers techniques are often considered to address constraints [30]. In DFO, these techniques were first introduced by Lewis and Torczon [39], in the context of Directional Direct Search (DDS). Unfortunately, the adequate choice of penalty parameters (or the associated multipliers) adds serious difficulties in a derivative-free setting. However, if the constraints define a convex feasible set in which it is easy to project, then projection schemes could be combined with DFO methods to address constraints in an effective and inexpensive way.

Optimization problems defined on easy-to-project convex sets appear frequently in different domains (see [18] for a list of different real applications). There are several derivative-based methods for solving these convex constrained optimization problems that only depend on the convexity of the feasible region and on the capacity to easily project onto it [14, 15, 18]. Some are Newton-type methods and others are gradient-type algorithms. The common denominator is that all generate sequences of feasible iterates by searching along descent directions. In this sense, they can be regarded as constrained versions of their unconstrained derivative-based counterparts.

*Center for Mathematics and Applications (NOVA Math) and Department of Mathematics, NOVA FCT, 2829-516 Caparica, Portugal (alcustodio@fct.unl.pt).

†Center for Mathematics and Applications (NOVA Math), NOVA FCT, 2829-516 Caparica, Portugal (e.krulikovski@fct.unl.pt).

‡Center for Mathematics and Applications (NOVA Math), NOVA FCT, 2829-516 Caparica, Portugal (m.raydan@fct.unl.pt).

In this work, the main goal is to explore the combination of DFO methods, specifically the DDS class, with standard inexpensive projection schemes, such as the Spectral Projected Gradient (SPG) method [15, 16], to produce effective DFO algorithms for solving convex constrained optimization problems. DFO approaches for convex constrained problems have already been proposed in the literature [22], but not for DDS methods, neither by directly considering projection techniques. In DDS, previous works have addressed bounds [37] and linear constraints [3, 38] by adapting the set of poll directions to the geometry of the nearby feasible region, but again not considering projection approaches on general convex sets.

We will analyze the properties of simplex gradients [19, 33], computed by reusing previous evaluations of the objective function, combined with a SPG scheme, to improve the performance of DDS when solving convex constrained optimization problems. Whenever the SPG method is used, a projection on $\Omega$ is required. If $\Omega$ is the nonempty intersection of a finite collection of convex simple sets, then Dykstra's alternating projection method [20] will be used to obtain the required projection.

The paper is organized as follows. In Section 2 we briefly introduce DDS (see Subsection 2.1) and the SPG method (see Subsection 2.2). Simplex gradients and their main properties are revised in Subsection 2.3. The proposed algorithmic structure is detailed in Section 3 and its convergence is analyzed in Section 4. Numerical experiments are reported in Section 5, for a set of smooth problems, indicating that the use of simplex gradients in a hybrid SPG and DDS scheme leads to significant reductions in the overall number of function evaluations required to solve the problems. Concluding remarks are presented in Section 6.

**Notation.** Throughout this paper, we denote by $\|\cdot\|$ the Euclidean norm and by $I$ the identity matrix. We also denote by $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ the sets of natural, integer, and rational numbers, respectively, and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For a function $f : \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x)$ and $\nabla_S f(x)$ represent the gradient and the simplex gradient of function $f$ computed at point $x$, respectively.

## 2. Preliminaries.

In this section, we revise the basic structure of a DDS algorithm [11, 25], as well as the main lines of the SPG method [15, 16, 17], which are the two main blocks of our hybrid approach. Simplex gradients [19, 33] and their quality as approximations to the true gradient of a smooth function will also be introduced, since they will be used as replacement of the true gradient in the SPG approach.

### 2.1. Directional Direct Search methods.

Directional Direct Search (DDS) was a term coined in [25] to represent any DFO algorithm that proceeds only by sampling the objective function at points corresponding to sets of directions. The first convergence analysis for an algorithm belonging to this class was provided in [45]. After, the algorithmic structure was generalized, being each iteration organized in a search step and a poll step [8].

The search step is optional and not required for ensuring convergence. It can encompass any finite strategy, with some additional requirements depending on the globalization strategy considered (generating points in an implicit mesh or requiring sufficient decrease to accept new points [35]).

If the search step fails in improving the current iterate, the poll step is obligatory performed. At this step, a set of directions with an adequate geometry, directly related to the geometry of the nearby feasible region and the level of smoothness of the objective function, needs to be selected. Typically, positive spanning sets [28] are used. The type of directions considered will give rise to different instances of DDS methods.

Coordinate search characterizes by using as poll directions the columns of the matrix $[I; -I]$, being particularly suited for bound constrained optimization. General linear constraints require specific techniques that adapt the set of directions to the geometry of the nearby constraints [3, 38]. In any case, a finite number of different sets of poll directions will be considered during the optimization process.

For general feasible regions, different strategies have been proposed, some using augmented Lagrangian approaches [39], others using extreme barrier approaches [8, 9], only evaluating feasible points, or progressive barrier approaches, based on filter methods [10]. In the last two, the union of all the sets of normalized poll directions considered should be asymptotically dense in the unit sphere [2, 9].

At the poll step, points associated to poll directions, scaled by a step size parameter, are evaluated, again in an attempt of improving the function value at the current iterate. Let $x_k$ be the current iterate and $\delta_k \in \mathbb{R}_+$ the current value of the step size parameter. The poll step evaluates the function $f$ at the points in the set

$$P_k = \{x_k + \delta_k d : d \in D_k\},$$

where $D_k$ denotes the set of poll directions considered at iteration $k$. This evaluation procedure can be performed under opportunistic or complete strategies. In the former, the polling procedure is stopped, once that a poll point that improves the function value at the current iterate is found. In the latter, all poll points are evaluated.

If an iteration is successful, meaning that a point was found, either at the search or the poll steps, with a better function value than the one of the current iterate, the new point is accepted as the current iterate and the step size is kept constant or can be increased. Unsuccessful iterations, obligatory reduce the step size parameter. For $\mathcal{C}^1$ functions, unless the current iterate is a stationary point, since any positive spanning set is guaranteed of having a descent direction [25], an improvement in the objective function value would be obtained for a sufficiently small step size parameter.

**2.2. Spectral Projected Gradient method.** The Spectral Projected Gradient (SPG) method was proposed in [15], for solving problem (1.1), when derivatives are available. For completeness, we present a brief review of it; see [15, 16, 17, 18] for additional details.

At each iteration of the SPG method, a step is attempted along the feasible projected gradient direction $d_k = \widehat{P}_\Omega(x_k - \lambda_k \nabla f(x_k)) - x_k$, where $\lambda_k > 0$ is the well-known non-monotone Barzilai-Borwein (also known as the spectral) choice of step length [12, 43], and $\widehat{P}_\Omega(z)$ represents the Euclidean projection of $z$ onto $\Omega$.

A key feature of the SPG method is to accept the initial spectral step length $\lambda_k$ as frequently as possible. For that, rather than imposing a function value decrease in $f$, the new iterate $x_{k+1} = x_k + \alpha d_k$ is accepted according to the non-monotone sufficient decrease condition

$$f(x_{k+1}) \leq f_{\max} + \gamma\alpha\nabla f(x_k)^\top d_k, \tag{2.1}$$

where, $0 < \gamma \ll 1$, $0 < \alpha \leq 1$, and $f_{\max} = \max_{1\leq j\leq \min\{k+1,M\}} f(x_{k-j+1})$ represents the maximum function value obtained in the last $M$ iterations. In case of rejection of the first trial point, $x_k + d_k$, a backtracking procedure is initialized, by testing $x_k + \alpha d_k$, with $0 < \alpha < 1$. As a consequence, only one projection onto $\Omega$ is required per iteration.

In many applications, either $\Omega$ is a simple set (i.e., easy to project onto it, such as boxes, spheres, or half-spaces, among others) or it can be written as the nonempty intersection of a finite collection of closed and convex simple sets. In this case, similarly to [17], Dykstra's alternating projection algorithm [20] can be used to obtain the required projection.

Consider $\Omega = \cap_{i=1}^p \Omega_i \neq \emptyset$, with $\Omega_i \subset \mathbb{R}^n, i = 1, \ldots, p$ closed and convex sets. Dykstra's algorithm computes the projection of $y_0$ onto $\Omega$ by generating two sequences, $\{y_i^l\}$ and $\{z_i^l\}$, using the recursive formulae:

$$\begin{array}{rcll} y_0^l & = & y_p^{l-1}, & \\ y_i^l & = & \widehat{P}_{\Omega_i}(y_{i-1}^l - z_i^{l-1}), & i = 1, \ldots, p, \\ z_i^l & = & y_i^l - (y_{i-1}^l - z_i^{l-1}), & i = 1, \ldots, p, \end{array} \tag{2.2}$$

for $l \in \mathbb{N}$, with initial values $y_p^0 = y_0$ and $z_i^0 = 0$ for $i = 1, \ldots, p$. The increment vectors $z_i^l$ play a key role in the convergence of the sequence $\{y_i^l\}_{l\in\mathbb{N}}$. In [20] it is proved that the sequence $\{y_i^l\}_{l\in\mathbb{N}}$ generated by (2.2) converges to $\widehat{P}_\Omega(y_0)$, that is, for any $i = 1, \ldots, p$ and any $y_0 \in \mathbb{R}^n$, $\|y_i^l - \widehat{P}_\Omega(y_0)\| \to 0$ as $l \to \infty$. For a full review of Dykstra's algorithm see [29].

**2.3. Simplex Gradients.** Simplex gradients were firstly proposed in [19], when defining the implicit filtering method [34] to optimize functions subject to numerical noise. Since then, they have been used to define new classes of simplex-based direct search methods [46], to develop convergent variants of the simplex of Nelder-Mead [32], or as descent indicators for ordering poll directions [27], when exploring opportunistic variants of DDS.

Simplex gradients are computed using a simplex, that is a set of $n + 1$ affinely independent points such that the corresponding convex hull has a nonempty interior. Let $\{y_0, y_1, \ldots, y_n\}$ be the simplex and assume that the function values $\{f(y_0),\dots,f(y_n)\}$ are known. Define

$$S = [y_1 - y_0 \cdots y_n - y_0] \text{ and } \delta(f;S) = [f(y_1) - f(y_0) \dots f(y_n) - f(y_0)]^\top.$$

The simplex gradient, $\nabla_S f(y_0)$, is the solution of the linear system

$$S^\top \nabla_S f(y_0) = \delta(f;S).$$

There could be situations in which we would have more or less than $n+1$ points available for the simplex gradient computation, corresponding to overdetermined linear systems, that could be solved using least squares approaches when $S$ is full rank, or underdetermined linear systems, for which a common approach is to compute a minimum Frobenious norm solution.

In general, considering a poised set $\{y_0, y_1, \dots, y_k\}$, with $k+1$ points, for a simplex gradient computation, meaning $rank(S) = \min\{n,k\}$, the simplex gradient can be expressed as

$$\nabla_S f(y_0) = V\Sigma^{-1}U^\top \delta(f;S)/\Delta \text{ where } \Delta = \max_{1\le i\le k} \|y_i - y_0\| \tag{2.3}$$

and $U\Sigma V^\top$ is the reduced singular value decomposition (SVD) of $S^\top/\Delta$.

Bounds for the error between simplex gradients and the true function gradient were established in [33], for determined simplex gradients, and extended in [24] to the nondetermined cases. The nonsmooth case was addressed in [26] and calculus rules were provided in [44]. In Theorem 2.1, we reproduce the bounds for simplex gradients of smooth functions, omitting the case of underdetermined simplex gradients ($k < n$), since it will be irrelevant in what follows.

THEOREM 2.1. *Let $\{y_0, y_1, \dots, y_k\}$ be a poised sample set for computing a simplex gradient in $\mathbb{R}^n$, with $k \ge n$. Consider the smallest (closed) ball, $B(y_0;\Delta)$, with $\Delta = \max_{1\le i\le k}\|y_i - y_0\|$, that contains the sample set. Let $S = [y_1 - y_0 \cdots y_k - y_0]$ and $U\Sigma V^\top$ be the reduced SVD of $S^\top/\Delta$. Suppose that $f$ has a Lipschitz continuous gradient in an open domain containing $B(y_0;\Delta)$, with Lipschitz constant $L > 0$. The simplex gradient error of $f$ at $y_0$ satisfies*

$$\|\nabla f(y_0) - \nabla_S f(y_0)\| \le \left(\sqrt{k}\frac{L}{2}\|\Sigma^{-1}\|\right)\Delta. \tag{2.4}$$

In particular, the error bound depends on the constant $\|\Sigma^{-1}\|$, which measures the quality of the geometry of the set of points used for computation. To control this quality, the notion of $\Lambda$-poisedness was introduced in [23], meaning that there is a constant $\Lambda > 0$ such that $\|\Sigma^{-1}\| \le \Lambda$. Once that the quality of the geometry of the sampling set is controlled, good quality simplex gradients can be computed by reducing the size of $\Delta$, corresponding to the radius of the sampling set.

## 3. The hybrid approach.

In DDS methods, a search step is often used to implement some heuristic procedures, in an attempt of improving the algorithmic efficiency [7, 41, 47]. However, the convergence analysis requires the evaluation of a finite number of points at this step, which should be projected on the implicit mesh, when considering a globalization strategy based on integer lattices [25].

In this work, we analyze the combination of the SPG scheme with DDS, which should not be regarded as a simple search step based on the SPG. In fact, simplex gradients, as accurate approximations of the true gradient for sufficiently smooth functions, allow to establish stronger convergence properties for the algorithm, than the ones derived for classical DDS. If after a given iteration all the new points are generated from a successful SPG step, then all limit points of the sequence of iterates generated by the algorithm will be stationary (see Section 4.2). Another advantage of this combination is that the projection on the implicit mesh is not required, when the new iterate is generated at the SPG step. Additionally, since the SPG step is systematically projected on the feasible set $\Omega$, the need for a feasible initialization (required for the convergence of DDS) is overcome. We emphasize that a preliminary search step could also be considered, but we omitted it for simplicity.

In the SPG approach, the gradient will be replaced by a simplex gradient, computed by reusing points evaluated at the poll step, including infeasible ones, avoiding function evaluations with the solely purpose of computing approximations to derivatives. After each unsuccessful poll step, the poll points with the corresponding function values, allow the computation of an overdetermined simplex gradient, as described in Section 2.3, which can be used to improve the efficiency of the algorithm by incorporating it in a SPG step.

There may be some iterations, particularly at the beginning of the optimization process, where the step size is large and consequently the poll points are considerably distant from the current iterate, conducting to a poor quality simplex gradient. Thus, if the SPG step fails in improving the function value at the current iterate, the reduction of the step size parameter, occurred after the previous unsuccessful poll step, will contribute to increase the quality of the next computed simplex gradient. Algorithm 3 formalizes the described procedure.

---

A hybrid DDS and SPG algorithm for convex minimization.

**Initialization**

Choose a set (possibly infinite) of positive spanning sets, $\mathcal{D}$, and the initial step size $\delta_0 > 0$. Define $0 < \epsilon_1 < 1$, the coefficient for step size contraction. Set the value of the constants related to the SPG step: $n_{SPG} = 0, x_{\lambda_1} = 0, x_{\lambda_2} = 0, g_{\lambda_1} = 0, g_{\lambda_2} = 0, M \in \mathbb{N}, 0 < \gamma < 1, \lambda_{\max} > \lambda_{min} > 0, 0 < \tilde{\epsilon} \ll 1$, $0 < \sigma_1 < \sigma_2 < 1$ and the sequence $\{\tilde{\eta}_k\}_{k \in \mathbb{N}_0}$. Consider $x_0 = \widehat{P}_\Omega(x_0)$ as initialization.

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

1. **Poll step:** Choose a positive spanning set $D_k$ from $\mathcal{D}$. Order the poll set $P_k = \{x_k + \delta_k d : d \in D_k\}$. Evaluate the objective function at the poll set $P_k$. If a poll point is found such that $f(x_k + \delta_k d) < f(x_k)$, then set $x_{k+1} = x_k + \delta_k d$, $n_{SPG} = 0$, and declare the poll step as successful. Otherwise, the poll step is declared as unsuccessful.

2. **Step size parameter update:** If the poll step was successful, then skip the SPG step. Otherwise, reduce the step size parameter, $\delta_{k+1} = \epsilon_1 \delta_k$.

3. **SPG step:** Compute a simplex gradient $\nabla_S f(x_k)$ using the evaluated poll set, $P_k$.
   Let $f_{\max} = \max_{1 \le j \le \min\{k+1, M\}} f(x_{k-j+1})$. Compute a trial point using the SPG strategy:

   $$(x_{trial}, f_{trial}) = SPG(x_k, \nabla_S f(x_k), \delta_k, x_{\lambda_1}, x_{\lambda_2}, g_{\lambda_1}, g_{\lambda_2}, f_{max}, n_{SPG}, \tilde{\eta}_k, \tilde{\epsilon}, \gamma, \lambda_{min}, \lambda_{max}, \sigma_1, \sigma_2).$$

   If $f(x_{trial}) < f(x_k)$, then set $x_{k+1} = x_{trial}$, $n_{SPG} = n_{SPG} + 1$, $x_{\lambda_1} = x_{\lambda_2}$, $x_{\lambda_2} = x_k$, $g_{\lambda_1} = g_{\lambda_2}$, $g_{\lambda_2} = \nabla_S f(x_k)$, and declare the SPG step as successful. Otherwise, set $x_{k+1} = x_k$.

**EndFor**

---

Some comments concerning Algorithm 3 are in order. The poll directions are computed as in any DDS algorithm. If the feasible region is a box, then coordinate directions corresponding to the columns of $[I; -I]$ are considered. If the optimization domain is defined by linear constraints, then the procedure described in [3] can be adopted to generate a set of directions that conform to the geometry of the nearby constraints. For general convex feasible regions, the set of poll directions will be asymptotically dense in the unit sphere [2]. To guarantee the quality of the simplex gradients (see Section 4.2), directions will always be normalized.

In the SPG step, the algorithm starts by computing the spectral step length and, using it and the current simplex gradient, defines a new point, that is projected on the feasible region. An important remark is the fact that the SPG scheme is a fast low-cost method, extremely simple to code, only requiring a specialized procedure to compute the projection onto $\Omega$. This projection is easy to compute for simple sets (e.g. boxes, spheres, or half-spaces) and Dykstra's alternating projection algorithm [20] will be used for sets that are the nonempty intersection of a finite collection of closed and convex simple sets (see Section 2.2).

At Step 3, a line search is performed, in an attempt of computing an acceptable point, that satisfies condition (2.1), with $\nabla f(x_k)$ replaced by $\nabla_S f(x_k)$ and adding the term $\tilde{\eta}_k \ge 0$ on the right hand side of the inequality. The addition of the term $\tilde{\eta}_k$ is inspired by the line search developed in [36] to deal with ascent directions. The sequence $\{\tilde{\eta}_k\}$ is chosen such that $\tilde{\eta}_k > 0$ for all $k$ and $\sum_{k=0}^{\infty} \tilde{\eta}_k < \infty$. For theoretical purposes, detailed in Section 4, in our case $\tilde{\eta}_k$ should be equal to zero for $k$ sufficiently large. In practice, large values of $\tilde{\eta}_k$, for small values of $k$, favor the early acceptance of new points in the line search, significantly reducing the number of unnecessary function evaluations, and guaranteeing the finite termination of the backtracking process in case of a poor quality simplex gradient. Algorithm 3 details the procedure.

---

$(x_{trial}, f_{trial}) = SPG(x_k, \nabla_S f(x_k), \delta_k, x_{\lambda_1}, x_{\lambda_2}, g_{\lambda_1}, g_{\lambda_2}, f_{max}, n_{SPG}, \tilde{\eta}_k, \tilde{\epsilon}, \gamma, \lambda_{min}, \lambda_{max}, \sigma_1, \sigma_2)$

Compute the spectral parameter by calling:

$$\tilde{\lambda}_k = lambda\_spect(x_k, \nabla_S f(x_k), \delta_k, x_{\lambda_1}, x_{\lambda_2}, g_{\lambda_1}, g_{\lambda_2}, n_{SPG}, \lambda_{min}, \lambda_{max}).$$

Compute the search direction $d_{trial} = \widehat{P}_\Omega(x_k - \tilde{\lambda}_k \nabla_S f(x_k)) - x_k$. If $\tilde{\eta}_k \leq \tilde{\epsilon}$, then set $\tilde{\eta}_k = 0$. Set $\alpha = 1$ and evaluate the trial point $f_{trial} = f(x_k + d_{trial})$.

**While** $f_{trial} > f_{max} + \gamma\alpha \nabla_S f(x_k)^\top d_{trial} + \tilde{\eta}_k$

Choose $\alpha \in [\sigma_1\alpha, \sigma_2\alpha]$ and evaluate $f_{trial} = f(x_k + \alpha d_{trial})$.

**EndWhile**

$x_{trial} = x_k + \alpha d_{trial}$

---

We note that, since the line search is performed along the feasible direction $d_{trial}$, no additional projections onto $\Omega$ will be required during the backtracking process. The computation of the spectral step length is detailed in Algorithm 3.

---

$\tilde{\lambda}_k = lambda\_spect(x_k, \nabla_S f(x_k), \delta_k, x_{\lambda_1}, x_{\lambda_2}, g_{\lambda_1}, g_{\lambda_2}, n_{SPG}, \lambda_{min}, \lambda_{max})$

**If** $n_{SPG} \geq 2$

Compute $s = x_{\lambda_1} - x_{\lambda_2}$ and $y = g_{\lambda_1} - g_{\lambda_2}$.

**If** $s^\top y \leq 0$ then $\tilde{\lambda}_k = \delta_k + \lambda_{max}$

**Else** $\tilde{\lambda}_k = \min\left(\delta_k + \lambda_{max}, \max\left(\lambda_{min}, \frac{s^\top s}{s^\top y}\right)\right)$

**EndIf**

**Else** $\tilde{\lambda}_k = \min\left(\delta_k + \lambda_{max}, \max\left(\lambda_{min}, \frac{1}{\|\widehat{P}_\Omega(x_k - \nabla_S f(x_k)) - x_k\|_\infty}\right)\right)$

**EndIf**

---

In resume, after each unsuccessful poll step, the algorithm takes advantage of the evaluated poll points to compute a simplex gradient, used to generate a feasible direction in the SPG step (after a single projection on the feasible region). This direction is explored in a non-monotone linesearch, based on the spectral step, in a attempt of finding a better point than the current iterate. Independently of succeeding or not on this task, the algorithm returns to the poll step. If the SPG step was successful, in case of failure of the poll step, new poll points will be available that would allow to compute a simplex gradient at the new point. If the SPG step was unsuccessful and the poll step continues to be unsuccessful, as mentioned before, the reduction of the step size would confer more quality to the new simplex gradient that would be computed at the same point, but using the new poll points.

**4. Convergence analysis.** In this section, we analyze the convergence properties of Algorithm 3, taking into account the results obtained in [8, 9, 15, 17]. At this point, it is worth recalling that $x_* \in \Omega$ is stationary for problem (1.1), if $\nabla f(x_*)^\top d \geq 0$ for all $d \in \mathbb{R}^n$ such that $x_* + d \in \Omega$. Equivalently, a point $x_* \in \Omega$ is stationary for problem (1.1) if $\|\widehat{P}_\Omega(x_* - \nabla f(x_*)) - x_*\| = 0$ (see, e.g., [14]).

There are only two possible cases to be considered in the convergence analysis of Algorithm 3: either the number of iterates $x_k$ obtained from a successful poll step (Step 1) is infinite or finite. In any case, we will assume that $f(x_0)$ is finite and also that the following hypothesis holds:

*Hypothesis* 4.1. The level set $L(x_0) = \{x \in \Omega : f(x) \leq f(x_0)\}$ is compact.

We note that, since the sequence of iterates $\{x_k\}_{k\in\mathbb{N}}$ is such that $\{f(x_k)\}_{k\in\mathbb{N}}$ is monotonically decreasing, all iterates generated by the algorithm (either obtained from Step 1 or Step 3) belong to the compact set $L(x_0)$. We are interested in studying the convergence of $\{x_k\}_{k\in\mathbb{N}}$ to a stationary point $x_*$, regardless of the starting point.

First, we will establish that the number of times that the SPG step (Step 3) is performed in Algorithm 3 cannot be finite.

LEMMA 4.2. *Under the Hypothesis 4.1, the number of times that the SPG step (Step 3) is performed in Algorithm 3 is infinite.*

*Proof.* Let us suppose, by way of contradiction, that the number of times that the SPG step (Step 3) is performed in Algorithm 3 is finite. This means that after $\bar{k} \in \mathbb{N}_0$ all iterates are generated at the poll step. Since a SPG step is always performed after an unsuccessful poll step, this means that after $\bar{k}$ only successful poll steps occur. Poll points are only generated in the implicit mesh of size greater or equal than $\delta_{\bar{k}} > 0$ (see Subsection 4.1). Since the intersection of the mesh with a compact set is finite, there is only a finite number of distinct options for new iterates. Thus, at some $k > \bar{k}$ the iteration needs to be unsuccessful, which leads to a contradiction. ∎

We note that performing Step 3 an infinite number of times does not necessarily mean that there will be an infinite number of successful iterates $x_k$ obtained from Step 3. This issue will be clarified in the next two subsections.

**4.1. The first case: Convergence from the poll step.** Let us consider the following hypothesis:

*Hypothesis* 4.3. The number of successful poll steps (Step 1) in Algorithm 3 is infinite.

Note that Hypothesis 4.3 implies that for any $\bar{k} \geq 0$ there exists some $k > \bar{k}$ such that the iterate $x_k$ is obtained from a successful poll step. A second possibility would be that there exists some $\bar{k} \geq 0$ such that the iterate $x_k$ is never updated for $k \geq \bar{k}$, i.e., $x_k = x_{\bar{k}}, \forall k \geq \bar{k}$.

*Hypothesis* 4.4. There exists some $\bar{k} \geq 0$ such that for all $k \geq \bar{k}$ the iterate $x_k$ is never updated.

In both cases, based on the standard convergence analysis of DDS methods, we will show that there exists a limit point of the sequence of iterates $\{x_k\}_{k\in\mathbb{N}}$ which is a stationary point. The last situation to be analyzed, where there exists some $\bar{k} \geq 1$ such that for all $k \geq \bar{k}$ the iterate $x_k$ is obtained from a successful SPG step will be the subject of Section 4.2.

Let us start by recalling that there are two types of strategies to globalize DDS methods [35]: to require a *sufficient decrease* in the objective function value for successful iterations, or to generate iterates in *integer lattices*, requiring only simple decrease of the function value to accept new points. We will adopt this last globalization strategy.

The convergence analysis is essentially divided in two parts. First, we need to show that the step size becomes infinitely fine, ensuring that there is at least one subsequence of step size parameters $\{\delta_k\}_{k\in K}$ satisfying $\lim_{k\in K} \delta_k = 0$. In our case, since the increase of the step size parameter will not be allowed at successful poll steps, we will be able to establish that the entire sequence $\{\delta_k\}_{k\in\mathbb{N}}$ converges to zero. Second, we need to analyze the behavior of the algorithm at limit points of the sequence of iterates associated with feasible unsuccessful poll steps.

According with the updating strategy of the step size parameter, described in Algorithm 3, we consider the following Hypothesis 4.5:

*Hypothesis* 4.5. Let $\tau \in \mathbb{Q}$, with $\tau > 1$, and $t_{min} \in \mathbb{Z}$, with $t_{min} \leq -1$. At unsuccessful poll steps, the step size parameter is updated as $\delta_{k+1} = \tau^{t_k}\delta_k$ with $t_k \in \{t_{min}, \ldots, -1\}$.

For establishing that $\{\delta_k\}_{k\in\mathbb{N}}$ converges to zero, additional requirements exist on the poll directions [8, 9].

*Hypothesis* 4.6. The set $D$ of positive spanning sets is finite and each $d \in D$ is obtained as the product $Gz_j$, $j = 1, \ldots, |D|$, of some fixed non-singular matrix $G \in \mathbb{R}^{n\times n}$ times a vector $z_j \in \mathbb{Z}^n$.

To address general nonlinear constraints, the following assumption will be considered.

*Hypothesis* 4.7. Let $D$ represent a finite set of positive spanning sets satisfying Hypothesis 4.6. The set $\mathcal{D}$ is such that:

1. $d_k$ is a nonnegative integer combination of the columns of $D$.
2. The distance between $x_k$ and the point $x_k + \delta_k d_k$ tends to zero if and only if $\delta_k$ does:
$$\lim_{k\in K} \delta_k \|d_k\| = 0 \;\Leftrightarrow\; \lim_{k\in K} \delta_k = 0,$$
for any infinite subsequence $K$.
3. The limits of all convergent subsequences of $\bar{D}_k = \{d_k/\|d_k\| \mid d_k \in D_k\}$ are positive spanning sets for $\mathbb{R}^n$.

Now, we recall the definition of current mesh, according to [9]. Specifically, at iteration $k$, the current mesh is given by $M_k = \cup_{x\in S_k}\{x + \delta_k Dz \mid z \in \mathbb{N}^{|D|}\}$, where $S_k$ represents the set of all the points evaluated previously to iteration $k$. The fact that the poll step only generates iterates in the mesh and the way we update the step size parameter are central for establishing that the sequence of step size parameters converges to zero.

Theorem 4.8. *Under the Hypotheses 4.1, one of 4.3 or 4.4, 4.5, and one of 4.6 or 4.7, Algorithm 3 generates a sequence of iterates satisfying*
$$\lim_{k\to+\infty} \delta_k = 0.$$

*Proof.* Let us suppose, by way of contradiction, that there exists $\tilde{\delta}$ such that $\delta_k > \tilde{\delta} > 0$ for all $k \geq \bar{k}$. First, let us assume that for any $k \geq \bar{k}$ the iterate $x_{k+1}$ is obtained from a successful poll step. Classical arguments [8, 45] allow us to conclude that all the iterates are generated in the mesh $M_{\bar{k}}$. Since the intersection of $M_{\bar{k}}$ with the compact set $L(x_0)$ is finite, there is only a finite number of distinct points for new iterates. Thus, at some $k > \bar{k}$ the iteration needs to be unsuccessful, and hence $\delta_{k+1} < \delta_k$. Repeating the argument recursively, this reduction of the step size will occur at a subsequence $\{k_i\}_{i\in\mathbb{N}}$ of unsuccessful poll step iterations for which $\lim_{i\to+\infty} \delta_{k_i} = 0$, and this leads to a contradiction.

Let us now consider the case in which there exists a finite number of iterates obtained from the SPG step between each two iterates obtained from the poll step (this finite number could be zero). In this case, although the iterates obtained from the SPG step are not generally in $M_{\bar{k}}$, since a SPG step is always performed after decreasing the step size $\delta_k$ and the step size is kept constant for successful poll steps, we can guarantee that the next iterate, obtained from a poll step, belongs to a more refined mesh. Consequently, there will be a subsequence $\{k_i\}_{i\in\mathbb{N}}$ of poll step iterations for which $\lim_{i\to+\infty} \delta_{k_i} = 0$, and once again this leads to a contradiction.

Therefore, we obtain that $\liminf_{k\to+\infty} \delta_k = 0$. Considering that $\{\delta_k\}_{k\in\mathbb{N}}$ is decreasing and nonnegative, thus convergent, it follows that $\lim_{k\to+\infty} \delta_k = 0$.

The case corresponding to Hypothesis 4.4 respects to consecutive unsuccessful poll steps. Thus, it is trivial that $\lim_{k\to+\infty} \delta_k = 0$. ∎

The following definition characterizes the subsequences of iterates that will be the subject of our analysis (see [8, 9]).

Definition 4.9. *A subsequence $\{x_k\}_{k\in K}$ of iterates corresponding to unsuccessful poll steps is said to be a refining subsequence if $\{\delta_k\}_{k\in K}$ converges to zero.*

As discussed in [8], there exists at least one convergent refining subsequence.

THEOREM 4.10. *Consider the sequence $\{x_k\}_{k\in\mathbb{N}}$ produced by Algorithm 3. Under the Hypotheses 4.1, one of 4.3 or 4.4, 4.5, and one of 4.6 or 4.7, Algorithm 3 generates at least one convergent refining subsequence $\{x_k\}_{k\in K}$.*

We will focus on limit points of convergent refining subsequences, analyzing the behavior of the algorithm along refining directions [9].

DEFINITION 4.11. *Given the limit point $x_*$ of a convergent refining subsequence $\{x_k\}_{k\in K}$, a direction $d$ is said to be a refining direction if there exists an infinite subset $L \subseteq K$ with poll directions $d_k \in D_k$ such that $x_k + \delta_k d_k \in \Omega$ and $\lim_{k\in L} d_k/\|d_k\| = d/\|d\|$.*

Now, we will recall the definitions of Clarke tangent and hypertangent cones, which are used to establish the stationarity results.

DEFINITION 4.12. *[9, Definition 3.5] A vector $d \in \mathbb{R}^n$ is said to be a Clarke tangent vector to the set $\Omega \subset \mathbb{R}^n$ at the point $x$ in the closure of $\Omega$ if for every sequence $\{y_k\}$ of elements of $\Omega$ that converges to $x$, and for every sequence of positive real numbers $\{t_k\}$ converging to zero, there exists a sequence of vectors $\{w_k\}$ converging to $d$ such that $y_k + t_k w_k \in \Omega$. The set $T^{Cl}_\Omega(x)$ of all tangent vectors to $\Omega$ at $x$ is called the Clarke tangent cone to $\Omega$ at $x$.*

DEFINITION 4.13. *[9, Definition 3.3] A vector $d \in \mathbb{R}^n$ is said to be a hypertangent vector to the set $\Omega \subset \mathbb{R}^n$ at the point $x \in \Omega$ if and only if there exists a scalar $\epsilon > 0$ such that*

$$y + tw \in \Omega \text{ for all } y \in \Omega \cap B(x;\epsilon), w \in B(d;\epsilon) \text{ and } 0 < t < \epsilon.$$

*The set of all hypertangent vectors to $\Omega$ at $x$ is called the hypertangent cone to $\Omega$ at $x$, and is denoted by $T^H_\Omega(x)$.*

The interior of a Clarke tangent cone defines the hypertangent cone. Reciprocally, the Clarke tangent cone is the closure of the corresponding hypertangent cone.

Since Algorithm 3 is applied to smooth objective functions, we will establish the non-negativity of the directional derivatives, computed at the limit point of a convergent refining subsequence generated by the algorithm, for the whole set of directions belonging to the tangent cone to $\Omega$. At this point, we need to recall that $f$ is strictly differentiable at $x_*$ if and only if it is Lipschitz continuous near $x_*$ and

$$f'(x_*; v) = \lim_{x\to x_*, t\downarrow 0} \frac{f(x+tv) - f(x)}{t} = \nabla f(x_*)^\top v \text{ for all } v \in \mathbb{R}^n.$$

Using the previous definitions, we aim at establishing the stationarity result formalized in the following definition.

DEFINITION 4.14. *[21] Let $f$ be strictly differentiable at a point $x_* \in \Omega$. We say that $x_*$ is a Clarke-KKT critical point of $f$ in $\Omega$ if, for all directions $d \in T^{CL}_\Omega(x_*)$, $\nabla f(x_*)^\top d \geq 0$.*

A similar result is first established for refining directions associated to the convergent refining subsequence.

THEOREM 4.15. *[9, Theorem 3.12] Consider a refining subsequence $\{x_k\}_{k\in K}$ converging to $x_* \in \Omega$ and let $d \in T^H_\Omega(x_*)$ be a refining direction for $x_*$. Assume that $f$ is strictly differentiable at $x_*$. Then $f'(x_*; d) \geq 0$.*

If we assume the density of the refining directions in $T^H_\Omega(x_*)$, we can extend Theorem 4.15 to the whole set of directions belonging to the Clarke tangent cone.

THEOREM 4.16. *[9, Corollary 3.14] Consider a refining subsequence $\{x_k\}_{k\in K}$ converging to $x_* \in \Omega$ and such that $T^H_\Omega(x_*) \neq \emptyset$. If $f$ is strictly differentiable at $x_*$ and the set of refining directions for $x_*$ is dense in $T^H_\Omega(x_*)$ then $x_*$ is a Clarke-KKT critical point.*

Summing up, under Hypotheses 4.1 and one of 4.3 or 4.4, we stated that there is a limit point $x_*$ of the sequence of iterates generated by Algorithm 3, such that $\nabla f(x_*)^\top d \geq 0$ for every direction $d \in T^{Cl}_\Omega(x_*)$. Therefore, $\nabla f(x_*)^\top d \geq 0$ for all $d$ such that $x + d \in \Omega$. Hence, $x_*$ is a stationary point for problem (1.1).

**4.2. The second case: Convergence from the SPG step.** For this last case, we consider the following hypothesis.

*Hypothesis* 4.17. There exists some $\bar{k} \geq 1$ such that for all $k \geq \bar{k}$ the iterate $x_k$ is obtained from a successful SPG step.

Thus, under the Hypothesis 4.17, for all $k \in J = \{k \in \mathbb{N} \mid k \geq \bar{k}\}$ our iterates have the form

$$x_{k+1} = x_k + \alpha_k \tilde{d}_k,$$

where the search direction $\tilde{d}_k$ is given by

$$\tilde{d}_k = \widehat{P}_\Omega(x_k - \tilde{\lambda}_k \nabla_S f(x_k)) - x_k,$$

$\tilde{\lambda}_k$ is described in Algorithm 3, and $0 < \alpha_k \leq 1$.

Let us start by establishing some auxiliary results that only depend on the convexity of $\Omega$.

LEMMA 4.18. *For all $x \in \Omega$ and $\tilde{d} = \widehat{P}_\Omega(x - \tilde{\lambda}\nabla_S f(x)) - x$,*

1. $\nabla_S f(x)^\top \tilde{d} \leq -\dfrac{1}{\tilde{\lambda}}\|\tilde{d}\|^2$.
2. *The vector $\tilde{d}$ vanishes at $x_*$ if and only if $\nabla_S f(x_*)^\top (y - x_*) \geq 0$ for all $y \in \Omega$.*

*Proof.* For the first statement, we have that $\Omega$ is a nonempty closed convex set, and so

$$(x - \tilde{\lambda}\nabla_S f(x) - \widehat{P}_\Omega(x - \tilde{\lambda}\nabla_S f(x)))^\top (y - \widehat{P}_\Omega(x - \tilde{\lambda}\nabla_S f(x))) \leq 0$$

for all $y \in \Omega$ (see, e.g., [29, Theorem 2.8]). Thus, choosing $y = x$ we obtain

$$\tilde{\lambda}(\nabla_S f(x))^\top (\widehat{P}_\Omega(x - \tilde{\lambda}\nabla_S f(x)) - x) \leq -\|\tilde{d}\|^2,$$

and the result follows.

For the last item, if $\tilde{d}$ vanishes at $x_*$, that is, $\widehat{P}_\Omega(x_* - \tilde{\lambda}\nabla_S f(x_*)) = x_*$, we obtain that

$$(x_* - \tilde{\lambda}\nabla_S f(x_*) - x_*)^\top (y - x_*) \leq 0,$$

for all $y \in \Omega$. This implies that

$$-\tilde{\lambda}\nabla_S f(x_*)^\top (y - x_*) \leq 0 \text{ then } \nabla_S f(x_*)^\top (y - x_*) \geq 0 \text{ for all } y \in \Omega.$$

Reciprocally, if $\nabla_S f(x_*)^\top (y - x_*) \geq 0$ for all $y \in \Omega$, then multiplying by $-\tilde{\lambda}$ and adding and subtracting $x_*$ we obtain

$$(x_* - \tilde{\lambda}\nabla_S f(x_*) - x_*)^\top (y - x_*) \leq 0 \text{ for all } y \in \Omega.$$

Therefore, since $\Omega$ is nonempty closed convex set, $x_* = \widehat{P}_\Omega(x_* - \tilde{\lambda}\nabla_S f(x_*))$, and so $\tilde{d}$ vanishes at $x_*$. ∎

At each poll step, Algorithm 3 stores the poll points and the corresponding objective function values. If the poll step is unsuccessful, we use these points, including infeasible ones, in the construction of a simplex gradient. The negative simplex gradient vector is then used in the SPG step. Thus, at iteration $k$, with $k \in J$, once the poll step fails, $|D_k|$ points were stored (the poll points $x_k + \delta_k d$ for all $d \in D_k$).

Now, we need to establish some key results, that will guarantee the quality of the simplex gradients, computed using these sets of points, as approximations to the true gradients. The analysis will be based on [2, 26, 35, 45].

We start by recalling that the cosine measure of a positive spanning set, $cm(D_k)$, which was introduced in [35], is given by

$$cm(D_k) = \min_{0 \neq v \in \mathbb{R}^n} \max_{d \in D_k} \frac{v^\top d}{\|v\|\|d\|}.$$

The following result states that if the cosine measure of a positive spanning set is greater than a strictly positive constant then the corresponding poll set is $\Lambda$-poised.

LEMMA 4.19. *[26, Proposition 1] Let $D_k$ be a positive spanning set for $\mathbb{R}^n$. Let $\|d\| \geq d_{min} > 0$ for all $d \in D_k$. Then $D_k$ is full rank and*

$$\|\Sigma^{-1}\| \leq \frac{1}{d_{\min} cm(D_k)} \quad \textit{where} \quad D_k^\top = U\Sigma V^\top.$$

As discussed in Section 3, for general nonlinear constraints, poll directions are computed according to the procedure described in [2], meaning that poll directions have the form $D_k = H_k[I; -I]$, with $H_k$ an integer orthogonal basis for $\mathbb{R}^n$ (see [2, Proposition 3.5]).

LEMMA 4.20. *Let $H_k$ be a orthogonal matrix and $D_k = H_k[I; -I]$. Then,*

$$cm(D_k) \geq \frac{1}{\sqrt{n}}.$$

*Proof.* From [45] we obtain that

$$cm(D_k) \geq \frac{1}{\kappa(BM)\sqrt{n}},$$

for $D_k = B[M; -M]$ with $B \in \mathbb{R}^{n\times n}$ and $M \in \mathbb{Z}^{n\times n}$, both nonsingular, and $\kappa(BM)$ the condition number of the matrix $BM$.

Since $H_k$ is orthogonal, $\kappa(H_k) = 1$. Hence

$$cm(D_k) \geq \frac{1}{\kappa(H_k)\sqrt{n}} = \frac{1}{\sqrt{n}}.$$

The matrix $S_k$, used for the simplex gradient computation is $S_k = \delta_k D_k$. Thus, according to Lemma 4.19, $\|\Sigma^{-1}\| \leq \frac{\sqrt{n}}{d_{min}}$, with $\frac{S_k^\top}{\Delta_k} = U\Sigma V^\top$ a reduced singular value decomposition, $\Delta_k = \max_{d\in D_k} \delta_k \|d\|$, and $d_{min} = \min_{d\in D_k} \frac{\delta_k}{\Delta_k}\|d\|$.

Therefore, by (2.4), we obtain

$$\|\nabla f(x_k) - \nabla_S f(x_k)\| \leq c\Delta_k \quad \text{with} \quad c = \left(\frac{\sqrt{2}Ln}{2d_{\min}}\right),$$

establishing the quality of the simplex gradients used as search directions in the SPG step. As previously discussed, $\lim_{k\in J} \delta_k = 0$. In addition, normalized poll directions have been considered.

Under the Hypothesis 4.1, the sequence $\{x_k\}_{k\in J}$ is bounded. Therefore, there exists a subsequence $\{x_k\}_{k\in K'\subseteq J}$, converging to a limit point, say $x_*$. Theorem 3.5 in [26] allows us to establish the existence of a subsequence of overdetermined simplex gradients, converging to the true gradient computed at this limit point.

THEOREM 4.21. *Consider the refining subsequence $\{x_k\}_{k\in K'\subseteq J}$ converging to $x_*$. Let $\nabla_S f(x_k)$ denote an overdetermined simplex gradient computed using $x_k$ and the poll points. Assume that this set is $\Lambda$-poised and let $f$ be strictly differentiable at $x_*$. Then, there exists a subsequence of indices $\tilde{K} \subseteq K' \subseteq J$ such that*

$$\lim_{k\to+\infty, k\in\tilde{K}} \nabla_S f(x_k) = \nabla f(x_*).$$

Notice that, assuming that $f$ is continuously differentiable, as a by-product, we obtain the safeguard that $\lim_{k\to+\infty, k\in\tilde{K}} \nabla_S f(x_k) = 0$ only if $\lim_{k\to+\infty, k\in\tilde{K}} \nabla f(x_k) = 0$.

We also note that under Hypothesis 4.17, we can obtain a stronger result.

COROLLARY 4.22. *Under the Hypothesis 4.17, consider $x_*$ a limit point of $\{x_k\}_{k\in J}$. Let $\nabla_S f(x_k)$ denote an overdetermined simplex gradient computed using $x_k$ and the poll points. Assume that this set is $\Lambda$-poised and let $f$ be strictly differentiable at $x_*$. Then, there exists a subsequence of indices $\tilde{K} \subseteq J$ such that $\lim_{k\to+\infty, k\in\tilde{K}} \nabla_S f(x_k) = \nabla f(x_*)$.*

*Proof.* For all $k \geq \bar{k}$, the poll step in Algorithm 3 always fails and so $\delta_{k+1} < \delta_k$. Thus, in our case, the entire sequence $\{x_k\}_{k \in J}$ is a refining subsequence, and hence from Theorem 4.21 the result follows. ∎

Now, we are able to study the convergence properties of our algorithm using simplex gradients. Our analysis follows the arguments developed in [17]. Given $x_k \in \Omega$, we consider the two subproblems (4.1) and (4.2). The first is defined as:

$$\begin{array}{rl} \text{minimize} & Q_k(d) \\ \text{subject to} & x_k + d \in \Omega, \end{array} \tag{4.1}$$

where, in the SPG algorithmic setting, $Q_k(d) = \frac{1}{2\lambda_k^{spg}}\|d\|^2 + \nabla f_k^\top d$, $\nabla f_k = \nabla f(x_k)$, and $\lambda_k^{spg}$ is the spectral step length [17].

By the strict convexity of problem (4.1) there exists a unique global minimizer $\bar{d}_k$. In addition, the optimal direction $\bar{d}_k$ is obtained by projecting $x_k - \lambda_k^{spg}\nabla f_k$ onto $\Omega$, with respect to the Euclidean norm. We have that the sequence $\{\bar{d}_k\}_{k\in\mathbb{N}}$ converges to $d_*^{spg} = \widehat{P}_\Omega(x_* - \lambda_*^{spg}\nabla f_*) - x_*$, and so $Q_k(\bar{d}_k)$ converges to $Q(d_*^{spg}) = \frac{1}{2\lambda_*^{spg}}\|d_*^{spg}\|^2 + \nabla f_*^\top d_*^{spg}$, with $\nabla f_* = \nabla f(x_*)$.

Consider now the following subproblem in which simplex gradients are used

$$\begin{array}{rl} \text{minimize} & \tilde{Q}_k(d) \\ \text{subject to} & x_k + d \in \Omega, \end{array} \tag{4.2}$$

with $\tilde{Q}_k(d) = \frac{1}{2\tilde{\lambda}_k}\|d\|^2 + \nabla_S f_k^\top d$, $\nabla_S f_k = \nabla_S f(x_k)$, and $\tilde{\lambda}_k$ given by Algorithm 3.

Lemma 4.23. *The sequence $\{\tilde{\lambda}_k\}_{k\in\mathbb{N}}$ is uniformly bounded.*

*Proof.* Since $\delta_k \leq \delta_0$ for all $k \in \mathbb{N}$, then in Algorithm 3 we obtain that $\tilde{\lambda}_k$ belongs to $[\lambda_{\min}, \delta_0 + \lambda_{\max}]$ for all $k \in \mathbb{N}$. ∎

Since $\tilde{Q}_k$ is strictly convex and the feasible set of (4.2) is convex, then problem (4.2) also has a unique global minimizer, $\tilde{d}_k$. Moreover, by Lemmas 4.18 and 4.23, we obtain

$$\tilde{Q}_k(\tilde{d}_k) = \frac{1}{2\tilde{\lambda}_k}\|\tilde{d}_k\|^2 + \nabla_S f_k^\top \tilde{d}_k \leq \left(\frac{1}{2\tilde{\lambda}_k} - \frac{1}{\tilde{\lambda}_k}\right)\|\tilde{d}_k\|^2 \leq 0. \tag{4.3}$$

For proving convergence of Algorithm 3 under the Hypothesis 4.17, we need some similar results to the ones established in [17]. To be precise, we need to prove convergence of an inexact SPG approach, that is, to prove convergence when we consider the SPG method using the simplex gradient instead of the exact real gradient, and using inexact projections (using Dykstra's algorithm) instead of the exact projection onto $\Omega$. Notice that by Theorem 4.21 we have that $\lim_{k\to+\infty, k\in\tilde{K}} \nabla_S f_k = \nabla f_*$. Since the projection on a convex set is a non-expansive operator, hence it is continuous, we obtain

$$\lim_{k\to+\infty, k\in\tilde{K}} \tilde{d}_k = \lim_{k\to+\infty, k\in\tilde{K}} \widehat{P}_\Omega(x_k - \tilde{\lambda}_k\nabla_S f_k) - x_k = \tilde{d}_* = \widehat{P}_\Omega(x_* - \tilde{\lambda}_*\nabla f_*) - x_* \tag{4.4}$$

and

$$\lim_{k\to+\infty, k\in\tilde{K}} \tilde{Q}_k(\tilde{d}_k) = \tilde{Q}(\tilde{d}_*) = \frac{1}{2\tilde{\lambda}_*}\|\tilde{d}_*\|^2 + \nabla f_*^\top \tilde{d}_*, \tag{4.5}$$

where $\tilde{\lambda}_*$ is the limit point of the subsequence of $\tilde{\lambda}_k$, for $k \in \tilde{K}$, that lies in the compact interval $[\lambda_{\min}, \delta_0 + \lambda_{\max}]$.

Moreover, as $\lim_{k\in\mathbb{N}} \delta_k = 0$ and the spectral step length is given by

$$\lambda_k^{spg} = \begin{cases} \min(\lambda_{\max}, \max(\lambda_{\min}, s_k^\top s_k / s_k^\top w_k)), & \text{if } s_k^\top w_k > 0, \\ \lambda_{\max}, & \text{otherwise} \end{cases}$$

with $s_k = x_k - x_{k-1}$ and $w_k = \nabla f_k - \nabla f_{k-1}$, then by the definition of $\tilde{\lambda}_k$ in Algorithm 3, it follows that

$$\lim_{k\to+\infty, k\in\tilde{K}} \tilde{\lambda}_k = \tilde{\lambda}_* = \lambda_*^{spg} = \lim_{k\to+\infty, k\in\tilde{K}} \lambda_k^{spg}$$

and by (4.4) and (4.5) we obtain that

$$\lim_{k\to+\infty, k\in\tilde{K}} \tilde{d}_k = \tilde{d}_* = d_*^{spg} = \widehat{P}_\Omega(x_* - \lambda_*^{spg}\nabla f_*) - x_*$$

and

$$\lim_{k\to+\infty, k\in\tilde{K}} \tilde{Q}_k(\tilde{d}_k) = \tilde{Q}(\tilde{d}_*) = Q(d_*^{spg}) = \lim_{k\to+\infty, k\in\tilde{K}} Q_k(\bar{d}_k). \tag{4.6}$$

In [17] it was established that Dykstra's algorithm can be used to obtain a direction $d_k^{spg}$ such that $x_k + d_k^{spg} \in \Omega$ and

$$Q_k(d_k^{spg}) \leq \eta_1 Q_k(\bar{d}_k), \tag{4.7}$$

where $\eta_1 \in (0, 1]$ (note that $\eta_1 = 1$ corresponds to the case where (4.1) is solved exactly). Similarly, by the convergence properties of Dykstra's algorithm, we can also obtain a direction $d_k$ such that $x_k + d_k \in \Omega$ and

$$\tilde{Q}_k(d_k) \leq \eta_2 \tilde{Q}_k(\tilde{d}_k), \tag{4.8}$$

where $\eta_2 \in (0, 1]$ (note that once again $\eta_2 = 1$ corresponds to the case where (4.2) is solved exactly).

The following preliminary result establishes that our Algorithm 3 is well defined, which means that in Algorithm 3 we stop the backtracking process after a finite number of trials.

LEMMA 4.24. *Under the Hypothesis 4.17, Algorithm 3 is well defined.*

*Proof.* If $\tilde{d}_k = 0$ then $\tilde{Q}_k(\tilde{d}_k) = 0$. Since $d_k$ is feasible we obtain $\tilde{Q}_k(d_k) \leq 0$. On the other hand, $\tilde{d}_k$ is the unique minimizer of (4.2) so $d_k = \tilde{d}_k = 0$. Hence, clearly

$$f_k \leq f_{\max} + \gamma\alpha\nabla_S f_k^\top d_k + \tilde{\eta}_k. \tag{4.9}$$

Now, if $\tilde{d}_k \neq 0$ then since $\tilde{Q}_k(0) = 0$ and the solution of (4.2) is unique, it follows that $\tilde{Q}_k(\tilde{d}_k) < 0$. By (4.8) we obtain $\tilde{Q}_k(d_k) < 0$. Therefore, by convexity of $\tilde{Q}_k$ and $\tilde{Q}_k(0) = 0$, it follows that $\tilde{Q}_k(d_k) - \tilde{Q}_k(0) \geq \nabla\tilde{Q}_k(0)^\top(d_k - 0)$ and consequently $0 > \tilde{Q}_k(d_k) \geq \nabla_S f_k^\top d_k$, that is, $\nabla_S f_k^\top d_k < 0$. Since $x_{k+1} = x_k + \alpha d_k$ then for $\alpha$ sufficiently small if $d_k$ is a descent direction (i.e., there exist $\zeta$ such that for all $\alpha \in (0, \zeta)$ we have $f(x_k + \alpha d_k) < f(x_k)$) we obtain that (4.9) holds after a finite number of trials. In addition, if $d_k$ is not a descent direction (e.g., when $k$ is not large enough and $\nabla_S f_k$ is still not a sufficiently good approximation of $\nabla f_k$) then the parameter $\tilde{\eta}_k > 0$ guarantees that (4.9) holds after a finite number of trials. ∎

The following lemmas play a key role in our convergence analysis. In all of them we use the fact that the sequence of iterates $\{x_k\}$ generated by Algorithm 3 belongs to the compact set $L(x_0)$.

LEMMA 4.25. *Under the Hypothesis 4.1 and 4.17, let $f$ be a continuously differentiable function in $L(x_0)$ and consider the refining subsequence $\{x_k\}_{k\in K'\subseteq J}$ converging to $x_*$. Then $\{d_k\}_{k\in K'\subseteq J}$ is bounded.*

*Proof.* For all $k \in J$, $\tilde{Q}_k(d_k) = \dfrac{1}{2\tilde{\lambda}_k}\|d_k\|^2 + \nabla_S f_k^\top d_k \leq 0$. So, by the Cauchy-Schwarz inequality

$$\|d_k\|^2 \leq -2\tilde{\lambda}_k\nabla_S f_k^\top d_k \leq 2\tilde{\lambda}_k\|\nabla_S f_k\|\|d_k\| \text{ for all } k \in K' \subseteq J.$$

Therefore, if $\|d_k\| \neq 0$,

$$\|d_k\| \leq 2\tilde{\lambda}_k\|\nabla_S f_k\| \text{ for all } k \in K' \subseteq J.$$

Now, since $f$ is Lipschitz continuous near $x_*$, $\{\nabla_S f_k\}_{k\in K'\subseteq J}$ is bounded [26, Lemma 3.1].

Moreover, from Lemma 4.23 the sequence $\{\tilde{\lambda}_k\}_{k\in K'\subseteq J}$ is also bounded, and the result follows. ∎

Our next result plays a key role in the convergence analysis of this subsection.

LEMMA 4.26. *Under the Hypotheses 4.1 and 4.17,* $\lim_{k\in J} \alpha_k \tilde{Q}_k(\tilde{d}_k) = 0$.

*Proof.* For any $k \in J$, the iterate $x_{k+1}$ obtained from a successful SPG step satisfies $f(x_{k+1}) < f(x_k)$. Consequently, since $\sum_{k=0}^{\infty} \tilde{\eta}_k < \infty$ and $f$ is continuous and bounded below, from the line search condition (4.9), imposed in Algorithm 3, we obtain that

$$\lim_{k\in J} \alpha_k \nabla_S f_k^\top d_k = 0. \tag{4.10}$$

Now, by (4.3), we have that,

$$0 \geq \tilde{Q}_k(d_k) = \frac{1}{2\tilde{\lambda}_k} d_k^\top d_k + \nabla_S f_k^\top d_k \geq \nabla_S f_k^\top d_k \text{ for all } k \in J.$$

Using (4.8), $0 \geq \eta_2 \tilde{Q}_k(\tilde{d}_k) \geq \tilde{Q}_k(d_k) \geq \nabla_S f_k^\top d_k$ for all $k \in J$. Therefore,

$$0 \geq \alpha_k \eta_2 \tilde{Q}_k(\tilde{d}_k) \geq \alpha_k \tilde{Q}_k(d_k) \geq \alpha_k \nabla_S f_k^\top d_k \text{ for all } k \in J.$$

Hence, by (4.10), it follows that $\lim_{k\in J} \alpha_k \tilde{Q}_k(\tilde{d}_k) = 0$. ∎

From Lemma 4.26 we note that if for a subsequence $\widehat{K} \subset J$ converging to a limit point, one of the two subsequences $\{\alpha_k\}_{k\in\widehat{K}}$ or $\{\tilde{Q}_k(\tilde{d}_k)\}_{k\in\widehat{K}}$ is bounded away from zero, then the other one converges to zero. This fact justifies the next two lemmas.

LEMMA 4.27. *Under the Hypotheses 4.1 and 4.17, assume that $K_1 \subset J$ is a sequence of indices such that*

$$\lim_{k\in K_1} x_k = x_* \in \Omega, \ \textit{and} \ \lim_{k\in K_1} \tilde{Q}_k(\tilde{d}_k) = 0.$$

*Then, $x_*$ satisfies $\nabla_S f_*^\top d \geq 0$ for all $d \in \mathbb{R}^n$ such that $x_* + d \in \Omega$.*

*Proof.* Since $\tilde{\lambda}_k \in [\lambda_{\min}, \delta_0 + \lambda_{\max}]$ for all $k \geq 0$, we can obtain a subsequence of indices $K_2 \subset K_1$ such that $\lim_{k\in K_2} \tilde{\lambda}_k = \tilde{\lambda}$ and so $\lim_{k\in K_2} \frac{1}{\tilde{\lambda}_k} = \frac{1}{\tilde{\lambda}}$.

We define $\tilde{Q}(d) = \frac{1}{2\tilde{\lambda}}\|d\|^2 + \nabla_S f_*^\top d$ for all $d \in \mathbb{R}^n$. Suppose that there exists $\hat{d} \in \mathbb{R}^n$ such that $x_* + \hat{d} \in \Omega$ and $\tilde{Q}(\hat{d}) < 0$. Define $\hat{d}_k = x_* + \hat{d} - x_k$ for all $k \in K_2$. Clearly, $x_k + \hat{d}_k \in \Omega$ for all $k \in K_2$. By continuity, since $\lim_{k\in K_2} x_k = x_*$, we have that

$$\lim_{k\in K_2} \tilde{Q}_k(\hat{d}_k) = \tilde{Q}(\hat{d}) < 0. \tag{4.11}$$

But, by the definition of $\tilde{d}_k$, we have that $\tilde{Q}_k(\tilde{d}_k) \leq \tilde{Q}_k(\hat{d}_k)$. Therefore, by (4.11),

$$\tilde{Q}_k(\tilde{d}_k) < 0,$$

for $k \in K_2$ large enough. This contradicts the fact that $\lim_{k\in K_2} \tilde{Q}_k(\tilde{d}_k) = 0$. The contradiction comes from the assumption that there exists $\hat{d}$ such that $\tilde{Q}(\hat{d}) < 0$. Then, $\tilde{Q}(d) \geq 0$ for all $d \in \mathbb{R}^n$ such that $x_* + d \in \Omega$. Thus, $\nabla_S f_*^\top d \geq 0$ for all $d \in \mathbb{R}^n$ such that $x_* + d \in \Omega$. ∎

LEMMA 4.28. *Under the Hypotheses 4.1 and 4.17, assume that $K_3 \subset J$ is a sequence of indices such that*

$$\lim_{k\in K_3} x_k = x_* \in \Omega \ \textit{and} \ \lim_{k\in K_3} \alpha_k = 0.$$

*Then,*

$$\lim_{k\in K_3} \tilde{Q}_k(\tilde{d}_k) = 0. \tag{4.12}$$

*Proof.* Let us suppose that (4.12) is not true. Then for some infinite set of indices $K_4 \subset K_3$, $\tilde{Q}_k(\tilde{d}_k)$ is bounded away from zero. According to the definition of $\tilde{\eta}_k$ in Algorithm 3 we have that there exists $\hat{k}$ such that $\tilde{\eta}_k = 0$ for all $k \geq \hat{k}$. Since $\alpha_k \to 0$, we have that for such $k \in K_4$ sufficiently large, there exists $\alpha'_k$ such that $\lim_{k\in K_4} \alpha'_k = 0$, thus we get

$$f(x_k + \alpha'_k d_k) > \max_{1\leq j\leq \min\{k+1,M\}} \{f(x_{k-j+1})\} + \gamma\alpha'_k \nabla_S f_k^\top d_k.$$

Hence,

$$f(x_k + \alpha'_k d_k) > f(x_k) + \gamma\alpha'_k \nabla_S f_k^\top d_k \text{ for all } k \in K_4.$$

Therefore, for all $k \in K_4$

$$\frac{f(x_k + \alpha'_k d_k) - f(x_k)}{\alpha'_k} > \gamma \nabla_S f_k^\top d_k.$$

By the mean value theorem, there exists $\zeta_k \in (0, 1)$ such that for $k \in K_4$

$$\nabla f(x_k + \zeta_k \alpha'_k d_k)^\top d_k > \gamma \nabla_S f_k^\top d_k. \tag{4.13}$$

From Lemma 4.25, the sequence $\{d_k\}_{k\in K_4}$ is bounded, and so there exists a sequence of indices $K_5 \subset K_4$ such that $\lim_{k\in K_5} d_k = d$. We also have that $\lim_{k\in K_5} \frac{1}{\tilde{\lambda}_k} = \frac{1}{\tilde{\lambda}}$, for some $\tilde{\lambda} > 0$. Taking limits for $k \in K_5$ in both sides of (4.13), we obtain $\nabla f_*^\top d \geq \gamma \nabla_S f_*^\top d$. Now, by Corollary 4.22, there exists a subsequence $\{x_k\}_{k\in \tilde{K}\subset J}$ such that $\lim_{k\in \tilde{K}} \nabla_S f_k = \nabla f_*$, and so $\nabla f_*^\top d \geq \gamma \nabla f_*^\top d$. This implies that $\nabla f_*^\top d \geq 0$. Hence,

$$\frac{1}{2\lambda^{spg}} \|d\|^2 + \nabla f_*^\top d \geq 0.$$

On the other hand, by [15, Lemma 2.1] we obtain

$$\nabla f_*^\top d \leq -\frac{1}{\lambda^{spg}} \|d\|^2, \text{ and hence } \frac{1}{2\lambda^{spg}} \|d\|^2 + \nabla f_*^\top d \leq \left( \frac{1}{2\lambda^{spg}} - \frac{1}{\lambda^{spg}} \right) \|d\|^2 \leq 0.$$

As a consequence,

$$\lim_{k\in K_5} \frac{1}{2\lambda_k^{spg}} \|d_k\|^2 + \nabla f_k^\top d_k = 0.$$

By (4.7) this implies that $\lim_{k\in K_5} Q_k(\bar{d}_k) = 0$. Since $\bar{d}_k$ is feasible for problem (4.2), we have that

$$\|\tilde{Q}_k(\bar{d}_k) - Q_k(\bar{d}_k)\| = \left\| \frac{1}{2\tilde{\lambda}_k} \|\bar{d}_k\|^2 + \nabla_S f_k^\top \bar{d}_k - \frac{1}{2\lambda_k^{spg}} \|\bar{d}_k\|^2 - \nabla f_k^\top \bar{d}_k \right\|.$$

Using the fact that $\lambda_k^{spg}$ and $\tilde{\lambda}_k$ are positive we obtain

$$\|\tilde{Q}_k(\bar{d}_k) - Q_k(\bar{d}_k)\| \leq \frac{1}{2} \frac{|\lambda_k^{spg} - \tilde{\lambda}_k|}{\tilde{\lambda}_k \lambda_k^{spg}} \|\bar{d}_k\|^2 + \left|(\nabla_S f_k - \nabla f_k)^\top \bar{d}_k\right|,$$

and by Cauchy-Schwarz inequality we have that

$$\|\tilde{Q}_k(\bar{d}_k) - Q_k(\bar{d}_k)\| \leq \left[ \frac{1}{2} \frac{|\lambda_k^{spg} - \tilde{\lambda}_k|}{\tilde{\lambda}_k \lambda_k^{spg}} \|\bar{d}_k\| + \|(\nabla_S f_k - \nabla f_k)\| \right] \|\bar{d}_k\|.$$

Taking limits for $k \in K_5$, as $\lim_{k\in K_3} \nabla_S f_k = \nabla f_*$ and by (4.6) it follows that

$$0 = \lim_{k\in K_5} Q_k(\bar{d}_k) = \lim_{k\in K_5} \tilde{Q}_k(\bar{d}_k) = \lim_{k\in K_5} \tilde{Q}_k(\tilde{d}_k).$$

This contradicts the assumption that $\tilde{Q}_k(\tilde{d}_k)$ is bounded away from zero for $k \in K_4$. Therefore, (4.12) holds. Moreover, we note that the hypotheses of Lemma 4.27 hold, with $K_3$ replacing $K_1$. So, by Lemma 4.27, $x_*$ satisfies $\nabla_S f_*^\top d \geq 0$ for all $d \in \mathbb{R}^n$ such that $x_* + d \in \Omega$. ∎

THEOREM 4.29. *Assume that Hypotheses 4.1 and 4.17 hold. Then every limit point of the sequence* $\{x_k\}_{k\in J}$ *is a stationary point for problem (1.1).*

*Proof.* Let us assume that $x_*$ is a limit point of $\{x_k\}_{k\in J}$ and, using Corollary 4.22, let $\{x_k\}_{k\in\tilde{K}}$ be the convergent subsequence of $\{x_k\}_{k\in J}$ whose limit point is $x_*$. Hence, $\lim_{k\to+\infty,k\in\tilde{K}} \nabla_S f_k = \nabla f_*$. By Lemma 4.26, the thesis follows applying Lemmas 4.27 and 4.28, and we conclude that $x_*$ is a stationary point for problem (1.1). ∎

**5. Numerical experiments.** To give further insight into the proposed hybrid scheme, and to illustrate its performance when solving problem (1.1), we present some numerical comparisons between Algorithm 3 (DDS-SPG) and DDS variants. All runs were carried out in Matlab R2022a, on an Intel® Quad-Core i7-1165G7 at 4.70 GHz with 16 GB of RAM memory, using Windows 10 Pro with 64 Bits.

The test functions have been previously used in numerical experiments in the literature with unconstrained DDS methods. In our case, there was the need of adding a constraint set to the problems, that could be written as the intersection of a finite number of convex sets on which it is trivial to project. To build the feasible set $\Omega$, we combined boxes, spheres, half-spaces, and ellipses. We notice that an ellipse is given by $\{x \in \mathbb{R}^n \mid x^\top Ax \le r\}$, where $A$ is symmetric positive definite. So, it can be reduced to a ball ($A = I$) by an adequate change of variables. Hence, for the small values of $n$ usually considered in DFO, projections onto ellipses can also be trivially computed.

The selected test set includes only smooth objective functions: 22 from CUTEr [31] and 20 from [27]. Additionally, we tested a simple quadratic function named as *Quadratic*, described in Experiment 5.1, and a strictly convex function denoted by *Strictly Convex 2 (SC2)*, which is described in Experiment 5.5. These two functions were used in a numerical study related to problem dimension. The 44 functions are listed in the first column of Tables 5–10.

The variants initially considered for the methods are enumerated in Table 1. For each one of them, we describe the type of positive basis selected and the polling strategy adopted (opportunistic or complete).

TABLE 1
*Variants of DDS and DDS-SPG considered in the numerical experiments.*

| Method | | Positive Basis | Strategy |
|---|---|---|---|
| 1 | DDS | Maximal | Opportunistic |
| 2 | | | Complete |
| 3 | DDS-SPG | Maximal | Opportunistic |
| 4 | | | Complete |
| 5 | DDS | ABDP/Maximal | Opportunistic |
| 6 | | | Complete |
| 7 | | ABDP/Minimal | Opportunistic |
| 8 | | | Complete |
| 9 | DDS-SPG | ABDP/Maximal | Opportunistic |
| 10 | | | Complete |
| 11 | | ABDP/Minimal | Opportunistic |
| 12 | | | Complete |
| 13 | DDS | Dense | Opportunistic |
| 14 | | | Complete |
| 15 | DDS-SPG | Dense | Opportunistic |
| 16 | | | Complete |

The choice of positive basis follows what has been described in Section 2.1: for bound constrained problems we considered coordinate search $[I; -I]$, the technique proposed by [3] (here denoted by ABDP) was used for general linear constraints, and asymptotically dense generation in the unit sphere, following the ORTHOMADS approach of [2], was considered for general nonlinear constrained problems. In this first set of experiments, to address constraints, both in DDS and in the poll step of DDS-SPG, we used the extreme barrier approach, that is, if $x \notin \Omega$, then $f(x) = +\infty$.

Concerning the selection of parameters for the algorithms, we used the typical defaults of DDS, $\delta_0 = 1$ and $\epsilon_1 = 1/2$. For the DDS-SPG method, we set $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, $\gamma = 10^{-4}$, $\lambda_{\min} = 10^{-3}$, $\lambda_{\max} = \delta_0$,

$M = 10$, and $\tilde{\eta}_k = |f(x_0)|/(k^{1.1})$ for all $k \in \mathbb{N}$ such that $\tilde{\eta}_k > 10^{-6}$, whereas for $\tilde{\eta}_k \leq 10^{-6}$ we set $\tilde{\eta}_k = 0$. In all experiments, the algorithms stopped if the number of iterations or the number of function evaluations reached $1000 \times n$ or if $\delta_k$ was less than $10^{-5}$. Considering the theoretical results derived in Section 4.2, DDS-SPG also stops if $\|\widehat{P}_\Omega(x_k - \tilde{\lambda}_k \nabla_S f(x_k)) - x_k\| < 10^{-7}$.

In what follows, we describe a variety of experiments combining different objective functions and different feasible sets $\Omega$. For each experiment, we also provide an initialization. It is worth mentioning that DDS methods require a feasible initial guess, which is not necessarily easy to obtain when the feasible region is the intersection of several different sets. To guarantee a fair comparison between the different methods, the initialization provided should be the same. In the numerical results reported, we favor the DDS schemes, by using the initial point easily obtained by the DDS-SPG methods, via Dykstra's algorithm.

*Experiment* 5.1. For our first experiment, we considered $f(x) = \sum_{i=1}^n x_i^2$ and $\Omega = \{x \in \mathbb{R}^n \mid -1 \leq x_i \leq 4\}$. The initial point was $x_0 = \left[\frac{3}{2}, \ldots, \frac{3}{2}\right]^\top$. Clearly, the unique global minimizer is given by $x_* = [0, \ldots, 0]^\top$.

*Experiment* 5.2. For this experiment, we used the objective function described in Experiment 5.1, and the feasible set is the intersection of a box and a half-space: $\{x \in \mathbb{R}^2 \mid -1 \leq x_i \leq 4 \text{ and } x_1 + x_2 \leq 5\}$. We started from the feasible point $x_0 = [2.63, 2.37]^\top$.

*Experiment* 5.3. We considered the objective function described in Experiment 5.1, but now $\Omega$ is the intersection of the box $-1 \leq x_i \leq 4$, the sphere $(x_1 - 4)^2 + (x_2 - 4)^2 \leq 4^2$, and the half-space $x_1 + x_2 \leq 5$. We started from the feasible point $x_0 = [2, 2]^\top$. In this case, the optimal function value is 2.7452.

*Experiment* 5.4. Again, we considered the objective function described in Experiment 5.1, but here the feasible set was the ellipse $10x_1^2 + x_2^2 \leq 1$. We initialized from the feasible point $x_0 = [0.17, 0.78]^\top$.

*Experiment* 5.5. In this case, we considered the problem of minimizing the strictly convex non-quadratic function $f(x) = \sum_{i=1}^n (i/10)(e^{x_i} - x_i)$, subject to $1 \leq x \leq 3$, and with initialization $x_0 = [2, \ldots, 2]^\top$. The unique global solution is given at $x^* = [1, \ldots, 1]^\top$.

*Experiment* 5.6. This experiment respects to $f(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1)\cos(4\pi x_2) + 0.3$, known as the Bohachevsky function [6]. The feasible set is the box $-50 \leq x_i \leq 50$, for $i = 1, 2$, the minimum global value is 0, and the global minimizer is $[0, 0]^\top$. We start from the feasible point $x_0 = [5, 5]^\top$.

In Table 2, we report the performance of several DDS-SPG variants, when applied to the problem described in Experiment 5.1, for a variety of feasible sets, stating the robustness of the approach in converging to the problem solution.

TABLE 2

*Performance of Algorithm 3 (DDS-SPG), considering different feasible sets, for the two-dimensional problem described in Experiment 5.1.*

| Feasible sets and initial point | Method | $f_{evals}$ | $f_{best}$ | $f_*$ | $\|\widehat{P}_\Omega(x_* - \nabla_S f_*) - x_*\|_\infty$ |
|---|---|---|---|---|---|
| $\Omega$ and $x_0$ as in Expt. (5.1) | 3 | 28 | 0.00 | 0.00 | 2.5E-16 |
| Only the sphere of Expt. (5.3); $x_0$ as in Expt. (5.3) | 15 | 19 | 2.75 | 2.75 | 8.9E-16 |
| $\Omega$ and $x_0$ as in Expt. (5.4) | 15 | 12 | 0.00 | 0.00 | 2.5E-16 |
| Only the half-space of Expt. (5.2); $x_0$ as in Expt. (5.2) | 9 | 35 | 0.00 | 0.00 | 8.3E-17 |
| $\Omega$ and $x_0$ as in Expt. (5.2) | 9 | 24 | 0.00 | 0.00 | 1.1E-16 |
| $\Omega$ and $x_0$ as in Expt. (5.3) | 15 | 14 | 2.75 | 2.75 | 8.9E-16 |

Table 3 reports the performance of several DDS-SPG and DDS variants on the problems corresponding to Experiments 5.1 and 5.5, considering different dimensions.

All variants converged to the unique global minimizer, in case of Experiment 5.5, attained on the boundary of the feasible set. DDS-SPG variants required lower number of function evaluations ($f_{evals}$) than the DDS approaches. It is also clear the dramatic increase in the difference of the number of function evaluations required with the increase of the problem dimension. In fact, it is worth noticing the controlled increase in the number of function evaluations required by the DDS-SPG variants.

The results for Experiments 5.6, 5.2, 5.3, and 5.4 are summarized in Table 4. In all cases, we clearly observe that the DDS-SPG variants require less function evaluations than the DDS appraches. The last column of the table reports the number of simplex gradients ($\nabla_S f_{evals}$) that were computed during the

TABLE 3

*Performance of several variants of DDS-SPG and DDS methods for solving the problems reported in Experiments 5.1 and 5.5, for different dimensions. In boldface, we highlight lower numbers of function evaluations.*

| Experiment | Method | $n=2$ $f_{evals}$ | $n=2$ $f_{best}$ | $n=3$ $f_{evals}$ | $n=3$ $f_{best}$ | $n=4$ $f_{evals}$ | $n=4$ $f_{best}$ | $n=5$ $f_{evals}$ | $n=5$ $f_{best}$ |
|---|---|---|---|---|---|---|---|---|---|
| 5.1 | 1 | 93 | 0.00 | 142 | 0.00 | 198 | 0.00 | 261 | 0.00 |
| | 2 | 82 | 0.00 | 139 | 0.00 | 207 | 0.00 | 286 | 0.00 |
| | 3 | **28** | 0.00 | **40** | 0.00 | **50** | 0.00 | **60** | 0.00 |
| | 4 | **27** | 0.00 | **43** | 0.00 | **56** | 0.00 | **69** | 0.00 |
| | | $n=10$ | | $n=20$ | | $n=30$ | | $n=40$ | |
| | 1 | 681 | 0.00 | 2046 | 0.00 | 4111 | 0.00 | 6876 | 0.00 |
| | 2 | 846 | 0.00 | 2791 | 0.00 | 5836 | 0.00 | 9981 | 0.00 |
| | 3 | **110** | 0.00 | **210** | 0.00 | **310** | 0.00 | **410** | 0.00 |
| | 4 | **134** | 0.00 | **264** | 0.00 | **394** | 0.00 | **524** | 0.00 |
| | | $n=2$ | | $n=3$ | | $n=4$ | | $n=5$ | |
| 5.5 | 1 | 38 | 0.52 | 61 | 1.03 | 87 | 1.72 | 116 | 2.58 |
| | 2 | 39 | 0.52 | 64 | 1.03 | 93 | 1.72 | 126 | 2.58 |
| | 3 | **13** | 0.52 | **18** | 1.03 | **23** | 1.72 | **28** | 2.58 |
| | 4 | **14** | 0.52 | **20** | 1.03 | **26** | 1.72 | **32** | 2.58 |
| | | $n=10$ | | $n=20$ | | $n=30$ | | $n=40$ | |
| | 1 | 306 | 9.45 | 911 | 36.08 | 1816 | 79.90 | 3021 | 140.9 |
| | 2 | 351 | 9.45 | 1101 | 36.08 | 2251 | 79.90 | 3801 | 140.9 |
| | 3 | **53** | 9.45 | **103** | 36.08 | **153** | 79.90 | **203** | 140.9 |
| | 4 | **62** | 9.45 | **122** | 36.08 | **182** | 79.90 | **242** | 140.9 |

iterative process (the symbol '–' represents cases where the SPG step was never performed in Algorithm 3).

TABLE 4

*Performance of several variants of DDS and DDS-SPG methods for the Experiments 5.6, 5.2, 5.3, and 5.4. In boldface, we highlight lower numbers of function evaluations.*

| Experiment | Method | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
|---|---|---|---|---|
| 5.6 | 1 | 113 | 0.00 | – |
| | 2 | 97 | 0.00 | – |
| | 3 | **43** | 0.00 | 1 |
| | 4 | **45** | 0.00 | 1 |
| 5.2 | 5 | 181 | 0.00 | – |
| | 6 | 178 | 0.00 | – |
| | 7 | 148 | 0.00 | – |
| | 8 | 139 | 0.00 | – |
| | 9 | **24** | 0.00 | 4 |
| | 10 | **28** | 0.00 | 4 |
| | 11 | 101 | 0.00 | 23 |
| | 12 | 111 | 0.00 | 23 |
| 5.3 | 13 | 115 | 2.7657 | – |
| | 14 | 120 | 2.7628 | – |
| | 15 | **14** | 2.7452 | 2 |
| | 16 | **14** | 2.7452 | 2 |
| 5.4 | 13 | 102 | 0.00 | – |
| | 14 | 104 | 0.00 | – |
| | 15 | **12** | 0.00 | 2 |
| | 16 | **11** | 0.00 | 2 |

Several figures illustrate the practical behavior of the DDS-SPG method, for problems with two variables. We present plots of the progression of the iterates in Figure 1 for Experiment 5.2 and in Figure 2 for

Experiment 5.3.

FIGURE 1. *Level sets of f for Experiment 5.2, and plot of the produced iterates, using DDS variant 5 (left) and DDS-SPG variant 9 (right).*

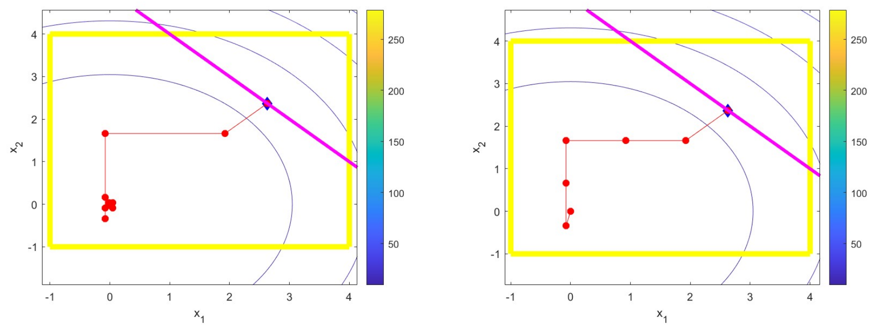
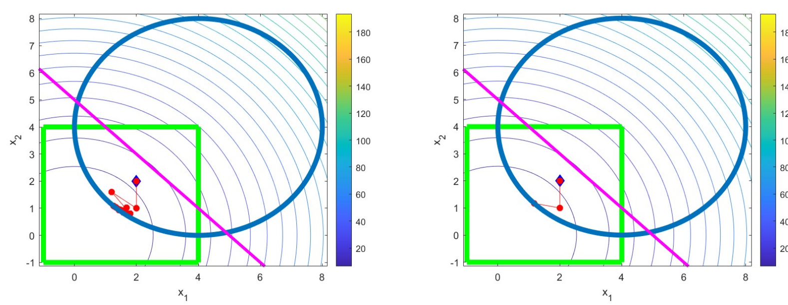


FIGURE 2. *Level sets of f for Experiment 5.3, and plot of the produced iterates, using DDS variant 14 (left) and DDS-SPG variant 16 (right).*



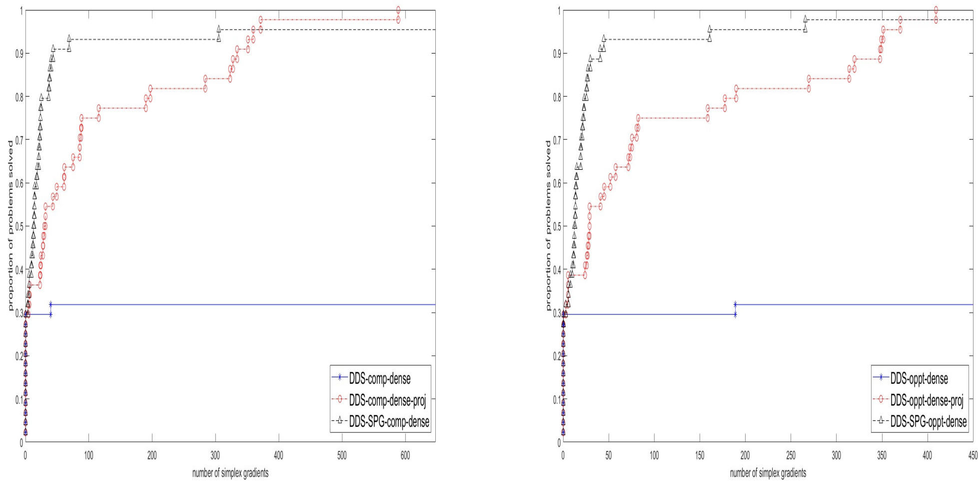
We note that Figures 1 and 2, as well as Tables 3 and 4, seem to indicate that DDS-SPG presents the best performance, by comparison with DDS. It is worthwhile to investigate if this is the result of the SPG step, or simply results from the projection on the feasible region. For that, we considered the set of 44 original problems (not including the Bohachevsky function) and a feasible region corresponding to a sphere. An additional variant to DDS and DDS-SPG was considered, namely DDS-Proj, that projects the infeasible points generated in $\Omega$, before function evaluation. The same approach was adopted at the poll step of DDS-SPG, whereas DDS continues to use the extreme barrier approach.

Therefore, considering sets of directions asymptotically dense in the unit sphere, Figure 3 reports the data profiles [42], corresponding to complete and opportunistic approaches. The corresponding numerical results can be found in Tables 5 and 6, respectively. In the tables we report the problem dimension $n$, which ranges from 2 to 20; the number of function evaluations $f_{evals}$ required; the number of simplex gradient evaluations $\nabla_S f_{evals}$ used; the best known unconstrained minimum $f(x_*)$; and $f_{best}$, which is the best function value obtained by each method (defined as 0 if its absolute value is below $10^{-6}$). All results were obtained starting from $x_0 = \hat{P}_\Omega([2,\ldots,2]^\top)$. For the computation of the data profiles, we considered the convergence test $f(x_k) \leq prob_{\min}(p) + gate(f(x_0) - prob_{\min}(p))$ with tolerance parameter $gate = 10^{-5}$ and $prob_{\min}(p)$ the smallest objective function value computed by any method on problem $p$.

The differences between DDS-SPG and DDS-Proj are clear, indicating that the good numerical performance observed is not the single result of the projection, but also benefits from the spectral step based on the simplex gradients.

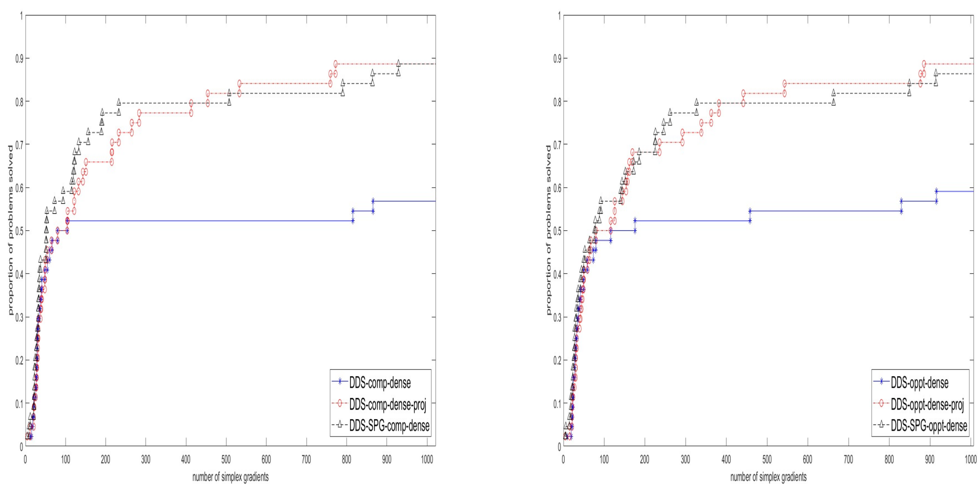
However, let us now consider a more complicated feasible region, corresponding to the convex set resulting from the intersection of the box $\{x \in \mathbb{R}^n \mid -1 \leq x \leq 4\}$, the half-space $\{x \in \mathbb{R}^n \mid w^\top x \leq 5\}$, where $w = [1/n,\ldots,1/n]^\top$, and the sphere $\{x \in \mathbb{R}^n \mid \|x - c\|^2 \leq 81\}$, where $c = [4,\ldots,4]^\top$. Again, all results

FIGURE 3. *Data profiles for DDS, DDS-proj, and DDS-SPG methods, using complete (on the left) and opportunistic (on the right) polling on asymptotically dense poll sets. The feasible region is the sphere* $\{x \in \mathbb{R}^n \mid \|x - c\|^2 \leq 4\}$*, where* $c = [4, \ldots, 4]^\top$.




were obtained with the initialization $x_0 = \widehat{P}_\Omega([2, \ldots, 2]^\top)$. Figure 4 displays the data profiles obtained, corresponding to the results reported in Tables 7 and 8, for complete and opportunistic versions, applied to the set of 44 smooth functions.

FIGURE 4. *Data profiles for DDS, DDS-proj, and DDS-SPG methods, using complete (on the left) and opportunistic (on the right) polling on asymptotically dense poll sets. The feasible region is the intersection of a box, a half-space, and a sphere.*




In this case, the advantages of DDS-SPG are clear by comparison with DDS, but not so evident when comparing with DDS-Proj. The more complicated feasible region forces the frequent use of the projection, causing more similarities between these two approaches.

Finally, at the beginning of Section 3 it was mentioned that DDS-SPG should not be regarded as a simple search step implemented in DDS, using the SPG approach. Thus, our last numerical experiments report the comparison between DDS-SPG and a variant of DDS, with a search step based on SPG, named as DDS-Search. In this case, the stopping criteria disregarded the maximum number of iterations or function

TABLE 5

*Results for the 44 smooth functions, obtained by DDS, DDS-proj, and DDS-SPG methods, using complete polling on asymptotically dense poll sets. The feasible region is the sphere* $\{x \in \mathbb{R}^n \mid \|x-c\|^2 \leq 4\}$*, where* $c = [4, \ldots, 4]^\top$.

| Problem | $n$ | $f(x_*)$ unconstrained | DDS | | DDS-proj | | DDS-SPG | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
| ARWHEAD | 2 | 0.00 | 29 | 171.48 | 153 | 171.46 | 73 | 171.46 | 13 |
| ARWHEAD | 10 | 0.00 | 1343 | 3008.46 | 2681 | 3002.91 | 734 | 3002.91 | 14 |
| ARWHEAD | 20 | 0.00 | 3091 | 6821.93 | 7281 | 6812.77 | 3334 | 6812.77 | 14 |
| BDQRTIC | 5 | 0.00 | 605 | 1.89E+04 | 801 | 1.89E+04 | 254 | 1.89E+04 | 14 |
| BDQRTIC | 10 | 1187 | 2258 | 1.44E+05 | 2481 | 1.43E+05 | 454 | 1.43E+05 | 14 |
| BDQRTIC | 20 | 35.41 | 3965 | 4.44E+05 | 10001 | 4.43E+05 | 974 | 4.43E+05 | 14 |
| BDVALUE | 2 | 0.00 | 29 | 82.20 | 121 | 81.62 | 98 | 81.62 | 14 |
| BROWNAL | 2 | 0.00 | 29 | 54.97 | 353 | 53.86 | 948 | 53.86 | 12 |
| BROWNAL | 10 | 0.00 | 404 | 3.56E+10 | 2441 | 3.52E+10 | 294 | 3.52E+10 | 14 |
| BROWNAL | 20 | 0.00 | 254 | 1.05E+22 | 561 | 1.05E+22 | 574 | 1.05E+22 | 14 |
| BROYDN3D | 2 | 0.00 | 29 | 147.64 | 113 | 147.46 | 90 | 147.46 | 14 |
| CHROSEN | 2 | 0.00 | 177 | 1 | 129 | 1 | 239 | 1 | 11 |
| FLETCHER | 2 | 1.82E-3 | 29 | 2.94E+04 | 153 | 4.43E+03 | 106 | 2.92E+04 | 14 |
| GENPOWELL3 | 2 | 0.00 | 119 | 0.68 | 73 | 0.68 | 67 | 0.68 | 11 |
| GENPOWELL3 | 4 | 0.00 | 352 | 1.38 | 3217 | 1.38 | 524 | 1.38 | 12 |
| INTEGREQ | 2 | 0.00 | 29 | 81.81 | 105 | 81.61 | 81 | 81.61 | 13 |
| PENALTY1 | 3 | 1.52E-5 | 43 | 577.79 | 85 | 577.79 | 7 | 577.79 | 1 |
| PENALTY1 | 4 | 2.25E-5 | 57 | 1278.06 | 113 | 1278.06 | 9 | 1278.06 | 1 |
| PENALTY1 | 10 | 7.09E-5 | 819 | 1.28E+04 | 2121 | 1.28E+04 | 21 | 1.28E+04 | 1 |
| PENALTY1 | 20 | 1.58E-4 | 254 | 6.36E+04 | 561 | 6.36E+04 | 41 | 6.36E+04 | 1 |
| PENALTY2 | 4 | 1.52E-5 | 233 | 7142.98 | 465 | 7133.50 | 222 | 7133.50 | 14 |
| PENALTY2 | 5 | 3.27E-5 | 359 | 1.86E+04 | 801 | 1.86E+04 | 194 | 1.86E+04 | 14 |
| PENALTY2 | 10 | 4.00E-4 | 1419 | 3.56E+05 | 2561 | 3.55E+05 | 354 | 3.55E+05 | 14 |
| PENALTY2 | 20 | 8.31E-3 | 3977 | 6.58E+06 | 9921 | 6.58E+06 | 574 | 6.58E+06 | 14 |
| PENALTY3 | 2 | 0.00 | 29 | 90.36 | 105 | 90.34 | 85 | 90.34 | 13 |
| POWELLSG | 4 | 0.00 | 406 | 889.03 | 577 | 887.92 | 278 | 887.92 | 14 |
| POWELLSG | 12 | 0.00 | 2730 | 4090.02 | 3745 | 4088.03 | 1862 | 4088.03 | 14 |
| POWELLSG | 20 | 0.00 | 7184 | 7658.95 | 10561 | 7655.81 | 2134 | 7655.81 | 14 |
| SROSENBR | 2 | 0.00 | 120 | 1 | 177 | 1 | 234 | 1 | 14 |
| SROSENBR | 4 | 0.00 | 287 | 1322.66 | 625 | 1307.19 | 342 | 1307.19 | 14 |
| SROSENBR | 10 | 0.00 | 1057 | 1.56E+04 | 2921 | 1.56E+04 | 1014 | 1.56E+04 | 14 |
| SROSENBR | 20 | 0.00 | 2727 | 5.37E+04 | 11841 | 5.33E+04 | 1854 | 5.33E+04 | 14 |
| TRIDIA | 2 | 0.00 | 82 | 9.03 | 129 | 9 | 230 | 9 | 10 |
| TRIDIA | 10 | 0.00 | 1237 | 569.17 | 3081 | 569.01 | 1254 | 569.01 | 14 |
| TRIDIA | 20 | 0.00 | 5039 | 2516.63 | 10481 | 2515.03 | 3214 | 2515.03 | 14 |
| TRIGO | 2 | 0.00 | 87 | 1.36 | 169 | 1.36 | 85 | 1.36 | 13 |
| VARDIM | 2 | 0.00 | 72 | 453.53 | 153 | 446.71 | 90 | 446.71 | 14 |
| VARDIM | 10 | 0.00 | 710 | 2.53E+08 | 2801 | 2.50E+08 | 354 | 2.50E+08 | 14 |
| VARDIM | 20 | 0.00 | 9446 | 7.49E+10 | 10961 | 7.47E+10 | 574 | 7.47E+10 | 14 |
| WOODS | 4 | 0.00 | 483 | 1666.26 | 609 | 1657.62 | 262 | 1657.62 | 14 |
| WOODS | 12 | 0.00 | 2560 | 2.22E+04 | 3889 | 2.22E+04 | 1478 | 2.22E+04 | 14 |
| WOODS | 20 | 0.00 | 6854 | 5.25E+04 | 11761 | 5.25E+04 | 2694 | 5.25E+04 | 14 |
| QUADRATIC | 2 | 0.00 | 29 | 13.37 | 57 | 13.37 | 5 | 13.37 | 1 |
| SC2 | 2 | 0.30 | 134 | 3.09 | 225 | 3.09 | 53 | 3.09 | 9 |

evaluations, since we would like to evaluate the robustness of the two algorithms. Figure 5 reports the two data profiles corresponding to complete and opportunistic approaches. The corresponding numerical results are in Tables 9 and 10. Again, the better performance of DDS-SPG is clear.

**6. Concluding remarks.** We proposed and analyzed a novel hybrid DFO method for the minimization of smooth functions over a feasible set that can be written as the intersection of a finite collection of convex and closed sets, such that it is easy and inexpensive to project onto each one of them. The hybrid scheme takes advantage of DDS methods and, when it fails in generating a new point, explores a derivative-free version of the SPG method, which is an inexpensive gradient-type projection scheme known for its fast linear convergence and low computational cost when real gradients are used. In our DFO proposal, instead of real gradients, we use simplex gradients, conveniently obtained by reusing previous evaluations of the objective function. Moreover, to avoid unnecessary function evaluations when the computed simplex gradient is not sufficiently close to the real gradient, we include some convenient features into the algorithmic framework of the SPG scheme. Furthermore, at those iterations where the derivative-free SPG scheme is used, the required projections onto the feasible set are approximated at a low cost, using Dykstra's alternating projection method.

From a theoretical point of view, we established that if there exists an infinite number of successful poll step iterations, then there exists at least a subsequence of the generated sequence of iterates converging to a stationary point. On the other hand, if after some iteration all the forthcoming iterates are obtained from a successful SPG step, then every limit point of the generated sequence of iterates is a stationary point.

Table 6

*Results for the 44 smooth functions, obtained by DDS, DDS-proj and DDS-SPG methods, using opportunistic polling on asymptotically dense poll sets. The feasible region is the sphere* $\{x \in \mathbb{R}^n \mid \|x - c\|^2 \leq 4\}$*, where* $c = [4, \ldots, 4]^\top$.

| Problem | $n$ | $f(x_*)$ unconstrained | DDS | | DDS-proj | | DDS-SPG | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
| ARWHEAD | 2 | 0.00 | 38 | 171.48 | 160 | 171.46 | 85 | 171.46 | 13 |
| ARWHEAD | 10 | 0.00 | 976 | 3005.47 | 2638 | 3002.91 | 625 | 3002.91 | 14 |
| ARWHEAD | 20 | 0.00 | 3317 | 6824.69 | 8474 | 6812.77 | 2665 | 6812.77 | 14 |
| BDQRTIC | 5 | 0.00 | 385 | 18897.29 | 765 | 18866.15 | 250 | 18866.15 | 14 |
| BDQRTIC | 10 | 1187 | 1740 | 1.43E+05 | 2601 | 1.43E+05 | 391 | 1.43E+05 | 14 |
| BDQRTIC | 20 | 35.41 | 3164 | 4.44E+05 | 12136 | 4.43E+05 | 894 | 4.43E+05 | 14 |
| BDVALUE | 2 | 0.00 | 38 | 82.20 | 129 | 81.62 | 103 | 81.62 | 14 |
| BROWNAL | 2 | 0.00 | 38 | 54.97 | 337 | 53.86 | 520 | 53.86 | 12 |
| BROWNAL | 10 | 0.00 | 724 | 3.57E+10 | 3046 | 3.52E+10 | 308 | 3.52E+10 | 14 |
| BROWNAL | 20 | 0.00 | 261 | 1.05E+22 | 575 | 1.05E+22 | 588 | 1.05E+22 | 14 |
| BROYDN3D | 2 | 0.00 | 38 | 147.64 | 125 | 147.46 | 99 | 147.46 | 14 |
| CHROSEN | 2 | 0.00 | 137 | 1.01 | 132 | 1 | 167 | 1 | 11 |
| FLETCHER | 2 | 1.82E-3 | 38 | 2.94E+04 | 161 | 4.43E+03 | 105 | 2.92E+04 | 14 |
| GENPOWELL3 | 2 | 0.00 | 139 | 0.68 | 86 | 0.68 | 73 | 0.68 | 11 |
| GENPOWELL3 | 4 | 0.00 | 325 | 1.38 | 1527 | 1.38 | 1361 | 1.38 | 13 |
| INTEGREQ | 2 | 0.00 | 38 | 81.81 | 117 | 81.61 | 90 | 81.61 | 13 |
| PENALTY1 | 3 | 1.52E-5 | 48 | 577.79 | 99 | 577.79 | 8 | 577.79 | 1 |
| PENALTY1 | 4 | 2.25E-5 | 61 | 1.28E+03 | 127 | 1.28E+03 | 10 | 1.28E+03 | 1 |
| PENALTY1 | 10 | 7.09E-5 | 724 | 1.29E+04 | 2401 | 1.28E+04 | 22 | 1.28E+04 | 1 |
| PENALTY1 | 20 | 1.58E-4 | 261 | 6.36E+04 | 575 | 6.36E+04 | 42 | 6.36E+04 | 1 |
| PENALTY2 | 4 | 1.52E-5 | 218 | 7.14E+03 | 499 | 7.13E+03 | 214 | 7.13E+03 | 14 |
| PENALTY2 | 5 | 3.27E-5 | 409 | 1.87E+04 | 647 | 1.86E+04 | 196 | 1.86E+04 | 14 |
| PENALTY2 | 10 | 4.00E-4 | 1521 | 3.56E+05 | 2828 | 3.55E+05 | 344 | 3.55E+05 | 14 |
| PENALTY2 | 20 | 8.31E-3 | 3925 | 6.57E+06 | 12095 | 6.57E+06 | 588 | 6.57E+06 | 14 |
| PENALTY3 | 2 | 0.00 | 38 | 90.36 | 112 | 90.34 | 89 | 90.34 | 13 |
| POWELLSG | 4 | 0.00 | 366 | 889.25 | 589 | 887.92 | 309 | 887.92 | 14 |
| POWELLSG | 12 | 0.00 | 2222 | 4.09E+03 | 3811 | 4.09E+03 | 1504 | 4.09E+03 | 14 |
| POWELLSG | 20 | 0.00 | 9019 | 7.66E+03 | 11729 | 7.66E+03 | 1658 | 7.66E+03 | 14 |
| SROSENBR | 2 | 0.00 | 623 | 1.00 | 143 | 1 | 237 | 1 | 14 |
| SROSENBR | 4 | 0.00 | 262 | 1.32E+03 | 624 | 1.31E+03 | 254 | 1.31E+03 | 14 |
| SROSENBR | 10 | 0.00 | 1425 | 1.56E+04 | 2649 | 1.56E+04 | 775 | 1.56E+04 | 14 |
| SROSENBR | 20 | 0.00 | 2718 | 5.39E+04 | 10247 | 5.33E+04 | 1217 | 5.33E+04 | 14 |
| TRIDIA | 2 | 0.00 | 101 | 9.01 | 143 | 9 | 195 | 9 | 10 |
| TRIDIA | 10 | 0.00 | 1160 | 569.39 | 2619 | 569.01 | 967 | 569.01 | 14 |
| TRIDIA | 20 | 0.00 | 4013 | 2.52E+03 | 12209 | 2.52E+03 | 2785 | 2.52E+03 | 14 |
| TRIGO | 2 | 0.00 | 138 | 1.37 | 211 | 1.36 | 87 | 1.36 | 11 |
| VARDIM | 2 | 0.00 | 74 | 453.53 | 155 | 446.71 | 95 | 446.71 | 14 |
| VARDIM | 10 | 0.00 | 1485 | 2.51E+08 | 2362 | 2.50E+08 | 336 | 2.50E+08 | 14 |
| VARDIM | 20 | 0.00 | 6213 | 7.48E+10 | 11307 | 7.47E+10 | 588 | 7.47E+10 | 14 |
| WOODS | 4 | 0.00 | 385 | 1.66E+03 | 506 | 1.66E+03 | 233 | 1.66E+03 | 14 |
| WOODS | 12 | 0.00 | 1340 | 2.23E+04 | 3612 | 2.22E+04 | 902 | 2.22E+04 | 14 |
| WOODS | 20 | 0.00 | 4906 | 5.25E+04 | 13091 | 5.25E+04 | 1192 | 5.25E+04 | 14 |
| QUADRATIC | 2 | 0.00 | 38 | 13.37 | 71 | 13.37 | 6 | 13.37 | 1 |
| SC2 | 2 | 0.30 | 118 | 3.09 | 214 | 3.09 | 57 | 3.09 | 9 |

From a practical point of view, the main objective of this suitable combination of derivative-free schemes is to reduce the number of function evaluations that turns out to be the greatest difficulty when dealing with the so-called black-box derivative-free optimization problems. In that sense, in our preliminary numerical experiments, we observed that the expected described advantages take effect, and indeed the number of required function evaluations is clearly reduced when compared to the use of only DDS-type methods.

**Funding.** All authors were financially supported by Fundação para a Ciência e a Tecnologia (FCT) (Portuguese Foundation for Science and Technology) under the scope of the projects UIDB/00297/2020, UIDP/00297/2020 (Centro de Matemática e Aplicações). Additionally, A. L. Custódio was financially supported by project PTDC/MAT-APL/28400/2017 and E. H. M. Krulikovski by project UI/297/2020-5/2021, of the same funding agency.

**Data availability.** The codes and datasets generated during and/or analyzed during the current study are available from the authors on reasonable request.

**Disclosure statement.** No potential conflict of interest was reported by the authors.

REFERENCES

TABLE 7

*Results for the 44 smooth functions, obtained by DDS, DDS-proj, and DDS-SPG methods, using complete polling on asymptotically dense poll sets. The feasible region is the intersection of a box, a half-space, and a sphere.*

| Problem | $n$ | $f(x_*)$ unconstrained | DDS | | DDS-proj | | DDS-SPG | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
| ARWHEAD | 2 | 0.00 | 176 | 0.00 | 193 | 0 | 57 | 0.00 | 9 |
| ARWHEAD | 10 | 0.00 | 1777 | 6.07 | 3081 | 6.05 | 1094 | 6.15 | 14 |
| ARWHEAD | 20 | 0.00 | 1884 | 299.31 | 9841 | 292.02 | 2023 | 296.92 | 14 |
| BDQRTIC | 5 | 0.00 | 910 | 0.00 | 901 | 0 | 189 | 0.00 | 9 |
| BDQRTIC | 10 | 1187 | 1361 | 1.11E+03 | 3441 | 1.09E+03 | 797 | 1.09E+03 | 14 |
| BDQRTIC | 20 | 35.41 | 3993 | 2.95E+04 | 11761 | 2.89E+04 | 1917 | 2.89E+04 | 14 |
| BDVALUE | 2 | 0.00 | 205 | 0.00 | 177 | 0.00 | 120 | 0.00 | 12 |
| BROWNAL | 2 | 0.00 | 223 | 0.00 | 225 | 0.00 | 13 | 0.00 | 1 |
| BROWNAL | 10 | 0.00 | 741 | 37.00 | 2961 | 33.07 | 412 | 35.02 | 14 |
| BROWNAL | 20 | 0.00 | 556 | 8.53E+11 | 1921 | 1.08E+04 | 574 | 8.55E+11 | 14 |
| BROYDN3D | 2 | 0.00 | 196 | 0.00 | 201 | 0.00 | 117 | 0.00 | 13 |
| CHROSEN | 2 | 0.00 | 175 | 0.00 | 177 | 0.00 | 53 | 0.00 | 9 |
| FLETCHER | 2 | 1.82E-3 | 175 | 0.00 | 185 | 0.00 | 162 | 5.79E-05 | 14 |
| GENPOWELL3 | 2 | 0.00 | 496 | 0.57 | 537 | 0.57 | 746 | 0.57 | 14 |
| GENPOWELL3 | 4 | 0.00 | 601 | 1.16 | 4001 | 1.16 | 4005 | 1.16 | 8 |
| INTEGREQ | 2 | 0.00 | 177 | 0.00 | 201 | 0.00 | 100 | 0.00 | 12 |
| PENALTY1 | 3 | 1.52E-5 | 2201 | 1.74E-05 | 709 | 2.79E-05 | 123 | 2.19E-05 | 9 |
| PENALTY1 | 4 | 2.25E-5 | 3539 | 2.27E-05 | 4001 | 3.23E-05 | 153 | 3.25E-05 | 9 |
| PENALTY1 | 10 | 7.09E-5 | 1924 | 172.80 | 2961 | 170.72 | 613 | 171.63 | 14 |
| PENALTY1 | 20 | 1.58E-4 | 1924 | 172.80 | 2961 | 170.72 | 613 | 171.63 | 14 |
| PENALTY2 | 4 | 1.52E-5 | 378 | 1.76E-05 | 497 | 1.69E-05 | 208 | 1.89E-05 | 11 |
| PENALTY2 | 5 | 3.27E-5 | 439 | 3.43E-05 | 701 | 3.66E-05 | 432 | 3.42E-05 | 12 |
| PENALTY2 | 10 | 4.00E-4 | 1673 | 3.01E+03 | 3121 | 2.99E+03 | 671 | 2.99E+03 | 14 |
| PENALTY2 | 20 | 8.31E-3 | 5834 | 4.69E+05 | 13841 | 4.68E+05 | 2154 | 4.68E+05 | 14 |
| PENALTY3 | 2 | 0.00 | 225 | 1.00E-03 | 225 | 1.00E-03 | 99 | 1.00E-03 | 11 |
| POWELLSG | 4 | 0.00 | 2507 | 1.82E-05 | 2561 | 1.82E-05 | 99 | 0.00 | 2 |
| POWELLSG | 12 | 0.00 | 3179 | 128.08 | 7345 | 127.76 | 12003 | 139.73 | 12 |
| POWELLSG | 20 | 0.00 | 9686 | 948.24 | 15681 | 947.00 | 20009 | 959.28 | 12 |
| SROSENBR | 2 | 0.00 | 464 | 1.25E-03 | 465 | 1.25E-03 | 63 | 0.00 | 11 |
| SROSENBR | 4 | 0.00 | 4001 | 0.38 | 4001 | 0.37 | 131 | 0.00 | 11 |
| SROSENBR | 10 | 0.00 | 10008 | 2.75 | 10001 | 3.92 | 10002 | 1.25 | 11 |
| SROSENBR | 20 | 0.00 | 20007 | 4.73 | 20001 | 9.84 | 20012 | 6.04 | 10 |
| TRIDIA | 2 | 0.00 | 192 | 0.00 | 193 | 0.00 | 153 | 0.00 | 13 |
| TRIDIA | 10 | 0.00 | 1425 | 22.46 | 4121 | 22.10 | 10003 | 22.85 | 12 |
| TRIDIA | 20 | 0.00 | 2462 | 646.72 | 20001 | 624.20 | 942 | 633.45 | 14 |
| TRIGO | 2 | 0.00 | 168 | 0.00 | 169 | 0.00 | 101 | 0.00 | 13 |
| VARDIM | 2 | 0.00 | 177 | 0.00 | 177 | 0.00 | 13 | 0.00 | 1 |
| VARDIM | 10 | 0.00 | 3112 | 1.42 | 4681 | 1.38 | 6042 | 2.61 | 14 |
| VARDIM | 20 | 0.00 | 11608 | 4.87E+08 | 10801 | 4.78E+08 | 410 | 4.78E+08 | 10 |
| WOODS | 4 | 0.00 | 887 | 3.24E-04 | 4001 | 7.07E-03 | 131 | 0.00 | 11 |
| WOODS | 12 | 0.00 | 8793 | 43.90 | 12001 | 38.11 | 1528 | 49.03 | 14 |
| WOODS | 20 | 0.00 | 14130 | 450.67 | 20001 | 456.17 | 20019 | 469.97 | 12 |
| QUADRATIC | 2 | 0.00 | 173 | 0.00 | 185 | 0.00 | 21 | 0.00 | 1 |
| SC2 | 2 | 0.30 | 189 | 0.30 | 185 | 0.30 | 46 | 0.30 | 6 |


[1] M. A. Abramson, *Mixed variable optimization of a load-bearing thermal insulation system using a filter pattern search algorithm*, Optim. Eng., 5 (2004), pp. 157–177.

[2] M. A. Abramson, C. Audet, J. E. Dennis Jr, and S. Le Digabel, *OrthoMADS: A deterministic MADS instance with orthogonal directions*, SIAM J. Optim., 20 (2009), pp. 948–966.

[3] M. A. Abramson, O. A. Brezhneva, J. E. Dennis Jr, and R. L. Pingel, *Pattern search in the presence of degenerate linear constraints*, Optim. Methods Softw., 23 (2008), pp. 297–319.

[4] S. Alarie, C. Audet, A. E. Gheribi, M. Kokkolaras, and S. Le Digabel, *Two decades of blackbox optimization applications*, EURO J. Comp. Optim., 9 (2021), p. 100011.

[5] P. Alberto, F. Nogueira, H. Rocha, and L. N. Vicente, *Pattern search methods for user-provided points: Application to molecular geometry problems*, SIAM J. Optim., 14 (2004), pp. 1216–1236.

[6] M. M. Ali, C. Khompatraporn, and Z. B. Zabinsky, *A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems*, J. Global Optim., 31 (2005), pp. 635–672.

[7] C. Audet, V. Béchard, and S. Le Digabel, *Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search*, J. Global Optim., 41 (2008), pp. 299–318.

[8] C. Audet and J. E. Dennis Jr, *Analysis of generalized pattern searches*, SIAM J. Optim., 13 (2002), pp. 889–903.

[9] ———, *Mesh adaptive direct search algorithms for constrained optimization*, SIAM J. Optim., 17 (2006), pp. 188–217.

[10] ———, *A progressive barrier for derivative-free nonlinear programming*, SIAM J. Optim., 20 (2009), pp. 445–472.

[11] C. Audet and W. Hare, *Derivative-free and Blackbox Optimization*, Springer, 2017.

[12] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

TABLE 8

*Results for the 44 smooth functions, obtained by DDS, DDS-proj, and DDS-SPG methods, using opportunistic polling on asymptotically dense poll sets. The feasible region is the intersection of a box, a half-space, and a sphere.*

| Problem | $n$ | $f(x_*)$ | DDS | | DDS-proj | | DDS-SPG | | |
|---|---|---|---|---|---|---|---|---|---|
| | | unconstrained | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
| ARWHEAD | 2 | 0.00 | 176 | 0.00 | 195 | 0.00 | 181 | 0.00 | 25 |
| ARWHEAD | 10 | 0.00 | 1744 | 6.08 | 3333 | 6.05 | 1241 | 6.10 | 47 |
| ARWHEAD | 20 | 0.00 | 1675 | 301.28 | 12895 | 292.02 | 1033 | 292.59 | 23 |
| BDQRTIC | 5 | 0.00 | 644 | 0.00 | 823 | 0.00 | 721 | 0.00 | 48 |
| BDQRTIC | 10 | 1187 | 1717 | 1.10E+03 | 3237 | 1.09E+03 | 1104 | 1.09E+03 | 44 |
| BDQRTIC | 20 | 35.41 | 5074 | 2.92E+04 | 14149 | 2.89E+04 | 2110 | 2.89E+04 | 42 |
| BDVALUE | 2 | 0.00 | 207 | 0.00 | 256 | 0.00 | 172 | 0.00 | 25 |
| BROWNAL | 2 | 0.00 | 327 | 0.00 | 331 | 0.00 | 279 | 0.00 | 37 |
| BROWNAL | 10 | 0.00 | 1089 | 37.55 | 3147 | 33.07 | 492 | 35.85 | 21 |
| BROWNAL | 20 | 0.00 | 524 | 8.47E+11 | 1453 | 1.05E+04 | 581 | 8.55E+11 | 14 |
| BROYDN3D | 2 | 0.00 | 190 | 0.00 | 207 | 0.00 | 204 | 0.00 | 28 |
| CHROSEN | 2 | 0.00 | 240 | 0.00 | 388 | 0.00 | 286 | 1.22E-06 | 38 |
| FLETCHER | 2 | 1.82E-3 | 200 | 1.32E-05 | 211 | 1.32e-05 | 222 | 6.06E-06 | 30 |
| GENPOWELL3 | 2 | 0.00 | 503 | 0.57 | 537 | 0.57 | 300 | 0.57 | 41 |
| GENPOWELL3 | 4 | 0.00 | 1006 | 1.16 | 4004 | 1.16 | 1090 | 1.16 | 79 |
| INTEGREQ | 2 | 0.00 | 178 | 0.00 | 255 | 0.00 | 197 | 0.00 | 27 |
| PENALTY1 | 3 | 1.52E-5 | 3006 | 1.88E-05 | 638 | 2.75E-05 | 218 | 2.87E-05 | 23 |
| PENALTY1 | 4 | 2.25E-5 | 4007 | 2.69E-05 | 4003 | 2.46E-05 | 178 | 5.93E-05 | 15 |
| PENALTY1 | 10 | 7.09E-5 | 760 | 176.26 | 2947 | 170.72 | 598 | 171.04 | 25 |
| PENALTY1 | 20 | 1.58E-4 | 760 | 2947 | 170.72 | 176.26 | 598 | 171.04 | 25 |
| PENALTY2 | 4 | 1.52E-5 | 473 | 1.93E-05 | 698 | 1.85E-05 | 438 | 1.91E-05 | 35 |
| PENALTY2 | 5 | 3.27E-5 | 338 | 3.52E-05 | 790 | 3.63E-05 | 395 | 3.53E-05 | 27 |
| PENALTY2 | 10 | 4.00E-4 | 1731 | 3.02E+03 | 3415 | 2.99E+03 | 1085 | 2.99E+03 | 42 |
| PENALTY2 | 20 | 8.31E-3 | 6277 | 4.71E+05 | 14007 | 4.68E+05 | 2125 | 4.67E+05 | 41 |
| PENALTY3 | 2 | 0.00 | 174 | 1.00E-03 | 174 | 1.00E-03 | 161 | 1.00E-03 | 23 |
| POWELLSG | 4 | 0.00 | 982 | 1.34E-05 | 992 | 1.34E-05 | 1037 | 5.49E-06 | 77 |
| POWELLSG | 12 | 0.00 | 2123 | 128.39 | 5337 | 127.76 | 1990 | 129.27 | 64 |
| POWELLSG | 20 | 0.00 | 6071 | 951.26 | 14861 | 947.00 | 3546 | 947.45 | 73 |
| SROSENBR | 2 | 0.00 | 714 | 1.60E-06 | 715 | 1.60E-06 | 858 | 1.32E-06 | 104 |
| SROSENBR | 4 | 0.00 | 4007 | 0.44 | 4007 | 7.51E-02 | 4008 | 0.48 | 264 |
| SROSENBR | 10 | 0.00 | 10013 | 2.28 | 10005 | 2.26 | 10016 | 2.12 | 301 |
| SROSENBR | 20 | 0.00 | 20035 | 5.40 | 20023 | 7.05 | 20026 | 5.37 | 314 |
| TRIDIA | 2 | 0.00 | 206 | 0.00 | 207 | 0.00 | 206 | 0.00 | 29 |
| TRIDIA | 10 | 0.00 | 1144 | 22.41 | 3164 | 22.10 | 998 | 22.34 | 36 |
| TRIDIA | 20 | 0.00 | 2176 | 655.30 | 18343 | 624.20 | 1722 | 628.32 | 35 |
| TRIGO | 2 | 0.00 | 121 | 122 | 0 | 0.00 | 160 | 0.00 | 23 |
| VARDIM | 2 | 0.00 | 210 | 0.00 | 210 | 0.00 | 174 | 0.00 | 24 |
| VARDIM | 10 | 0.00 | 2626 | 1.41 | 3988 | 1.38 | 3390 | 1.61 | 125 |
| VARDIM | 20 | 0.00 | 9052 | 4.86E+08 | 13984 | 4.78E+08 | 414 | 4.78E+08 | 10 |
| WOODS | 4 | 0.00 | 4008 | 1.63E-03 | 4007 | 1.15E-02 | 4002 | 1.52E-03 | 278 |
| WOODS | 12 | 0.00 | 4775 | 49.50 | 12004 | 37.81 | 4945 | 52.31 | 146 |
| WOODS | 20 | 0.00 | 7326 | 462.86 | 20004 | 444.67 | 7302 | 464.49 | 135 |
| QUADRATIC | 2 | 0.00 | 201 | 0.00 | 215 | 0.00 | 43 | 0.00 | 6 |
| SC2 | 2 | 0.30 | 180 | 0.30 | 222 | 0.30 | 141 | 0.30 | 20 |


[13] M. C. BERENBAUM, *Direct search methods in the optimisation of cancer chemotherapy regimens*, British J. Cancer, 61 (1990), pp. 101–109.

[14] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.

[15] E. G. BIRGIN, J. M. MARTÍNEZ, AND M. RAYDAN, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM J. Optim., 10 (2000), pp. 1196–1211.

[16] ———, *Algorithm 813: SPG software for convex-constrained optimization*, ACM Trans. Math. Software, 27 (2001), pp. 340–349.

[17] ———, *Inexact spectral projected gradient methods on convex sets*, IMA J. Numer. Anal., 23 (2003), pp. 539–559.

[18] ———, *Spectral projected gradient methods: Review and perspectives*, J. Stat. Softw., 60 (2014), pp. 1–21.

[19] D. M. BORTZ AND C. T. KELLEY, *The simplex gradient and noisy optimization problems*, in Computational Methods for Optimal Design and Control, vol. 24, Springer, 1998, pp. 77–90.

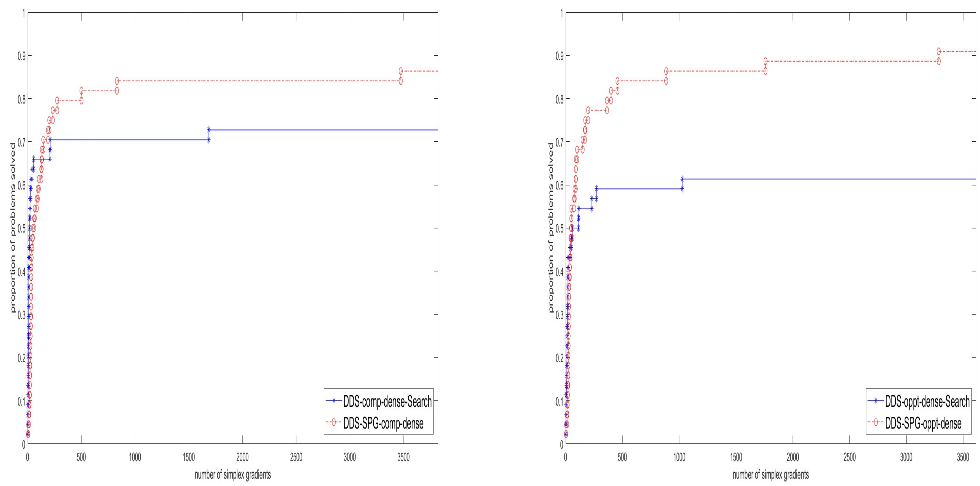
[20] J. P. BOYLE AND R. L. DYKSTRA, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, in Advances in Order Restricted Statistical Inference, Springer, 1986, pp. 28–47.

[21] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983. Reissued by SIAM, Philadelphia, 1990.

[22] P. D. CONEJO, E. W. KARAS, L. G. PEDROSO, A. A. RIBEIRO, AND M. SACHINE, *Global convergence of trust-region algorithms for convex constrained minimization without derivatives*, Appl. Math. Comput., 220 (2013), pp. 324–330.

[23] A. R. CONN, K. SCHEINBERG, AND L. N. VICENTE, *Geometry of interpolation sets in derivative free optimization*, Math. Program., 111 (2008), pp. 141–172.

Figure 5. *Data profiles for DDS-Search and DDS-SPG methods, using complete (on the left) and opportunistic (on the right) polling on asymptotically dense poll sets, respectively. The feasible region is the intersection of a box, a half-space, and a sphere.*





[24] ———, *Geometry of sample sets in derivative free optimization: Polynomial regression and underdetermined interpolation*, IMA J. Numer. Anal., 28 (2008), pp. 721–748.

[25] ———, *Introduction to Derivative-Free Optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.

[26] A. L. Custódio, J. E. Dennis Jr, and L. N. Vicente, *Using simplex gradients of nonsmooth functions in direct search methods*, IMA J. Numer. Anal., 28 (2008), pp. 770–784.

[27] A. L. Custódio and L. N. Vicente, *Using sampling and simplex derivatives in pattern search methods*, SIAM J. Optim., 18 (2007), pp. 537–555.

[28] C. Davis, *Theory of positive linear dependence*, Amer. J. Math., 76 (1954), pp. 733–746.

[29] R. Escalante and M. Raydan, *Alternating Projection Methods*, Fundamentals of Algorithms, SIAM, Philadelphia, 2011.

[30] R. Fletcher, *Pratical Methods of Optimization*, John Wiley, Chichester, 2nd ed., 1987.

[31] N. I. M. Gould, D. Orban, and P. L. Toint, *CUTEr and SifDec: A constrained and unconstrained testing environment, revisited*, ACM Trans. Math. Software, 29 (2003), pp. 373–394.

[32] C. T. Kelley, *Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition*, SIAM J. Optim., 10 (1999), pp. 43–55.

[33] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.

[34] C. T. Kelley, *Implicit Filtering*, Software Environments and Tools, SIAM, Philadelphia, 2011.

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

[36] W. La Cruz, J. M. Martínez, and M. Raydan, *Spectral residual method without gradient information for solving large-scale nonlinear systems of equations*, Math. Comp., 75 (2006), pp. 1429–1448.

[37] R. M. Lewis and V. Torczon, *Pattern search algorithms for bound constrained minimization*, SIAM J. Optim., 9 (1999), pp. 1082–1099.

[38] ———, *Pattern search methods for linearly constrained minimization*, SIAM J. Optim., 10 (2000), pp. 917–941.

[39] ———, *A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, SIAM J. Optim., 12 (2002), pp. 1075–1089.

[40] A. L. Marsden, M. Wang, J. E. Dennis Jr, and P. Moin, *Optimal aeroacoustic shape design using the surrogate management framework*, Optim. Eng., 5 (2004), pp. 235–262.

[41] J. V. Morais, A. L. Custódio, and G. M. Marques, *Calibration of parameters in Dynamic Energy Budget models using direct-search methods*, J. Math. Biology, 78 (2019), pp. 1439–1458.

[42] J. J. Moré and S. M. Wild, *Benchmarking derivative-free optimization algorithms*, SIAM J. Optim., 20 (2009), pp. 172–191.

[43] M. Raydan, *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA J. Numer. Anal., 13 (1993), pp. 321–326.

[44] R. G. Regis, *The calculus of simplex gradients*, Optim. Letters, 9 (2015), pp. 845–865.

[45] V. Torczon, *On the convergence of pattern search algorithms*, SIAM J. Optim., 7 (1997), pp. 1–25.

[46] P. Tseng, *Fortified-descent simplicial search method: A general approach*, SIAM J. Optim., 10 (1999), pp. 269–288.

[47] A. I. F. Vaz and L. N. Vicente, *A particle swarm pattern search method for bound constrained global optimization*, J. Global Optim., 39 (2007), pp. 197–219.

Table 9

*Results for the 44 smooth functions, obtained by DDS-Search and DDS-SPG methods, using complete polling on asymptotically dense poll sets, not considering neither the maximum number of iterations nor or function evaluations as stopping criteria. The feasible region is the intersection of a box, a half-space, and a sphere.*

| Problem | $n$ | $f(x_*)$ unconstrained | DDS-Search | | DDS-SPG | | |
|---|---|---|---|---|---|---|---|
| | | | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
| ARWHEAD | 2 | 0.00 | 82 | 0.00 | 175 | 0.00 | 25 |
| ARWHEAD | 10 | 0.00 | 2664 | 6.20 | 1853 | 6.06 | 62 |
| ARWHEAD | 20 | 0.00 | 878 | 292.10 | 1074 | 293.26 | 22 |
| BDQRTIC | 5 | 0.00 | 244 | 0 | 751 | 0.00 | 42 |
| BDQRTIC | 10 | 1187 | 1742 | 1097.77 | 1316 | 1093.07 | 45 |
| BDQRTIC | 20 | 35.41 | 10226 | 2.89E+04 | 1625 | 2.89E+04 | 31 |
| BDVALUE | 2 | 0.00 | 134 | 0.00 | 129 | 0.00 | 20 |
| BROWNAL | 2 | 0.00 | 78 | 0.00 | 204 | 0.00 | 28 |
| BROWNAL | 10 | 0.00 | 412 | 34.97 | 965 | 33.38 | 35 |
| BROWNAL | 20 | 0.00 | 574 | 8.56E+11 | 574 | 8.55E+11 | 14 |
| BROYDN3D | 2 | 0.00 | 138 | 0.00 | 143 | 0.00 | 21 |
| CHROSEN | 2 | 0.00 | 78 | 0 | 195 | 0.00 | 27 |
| FLETCHER | 2 | 1.82E-3 | 125 | 2.01E-05 | 217 | 6.06E-06 | 30 |
| GENPOWELL3 | 2 | 0.00 | 641 | 0.57 | 319 | 0.57 | 43 |
| GENPOWELL3 | 4 | 0.00 | 1090 | 1.16 | 1025 | 1.16 | 71 |
| INTEGREQ | 2 | 0.00 | 126 | 0.00 | 117 | 0.00 | 18 |
| PENALTY1 | 3 | 1.52E-5 | 3884 | 2.08E-05 | 212 | 3.42E-05 | 21 |
| PENALTY1 | 4 | 2.25E-5 | 31950 | 2.42E-05 | 224 | 4.46E-05 | 18 |
| PENALTY1 | 10 | 7.09E-5 | 481 | 170.75 | 1134 | 171.13 | 40 |
| PENALTY1 | 20 | 1.58E-4 | 481 | 170.75 | 1134 | 171.13 | 40 |
| PENALTY2 | 4 | 1.52E-5 | 246 | 1.89E-05 | 437 | 1.70E-05 | 31 |
| PENALTY2 | 5 | 3.27E-5 | 384 | 3.38E-05 | 575 | 3.45E-05 | 33 |
| PENALTY2 | 10 | 4.00E-4 | 2506 | 3.00E+03 | 1291 | 3.00E+03 | 44 |
| PENALTY2 | 20 | 8.31E-3 | 2734 | 4.68E+05 | 3966 | 4.68E+05 | 69 |
| PENALTY3 | 2 | 0.00 | 194 | 1.00E-03 | 186 | 1.00E-03 | 26 |
| POWELLSG | 4 | 0.00 | 207 | 0.00 | 683 | 5.11E-06 | 46 |
| POWELLSG | 12 | 0.00 | 23250 | 136.25 | 2261 | 128.39 | 64 |
| POWELLSG | 20 | 0.00 | 90854 | 957.58 | 2648 | 949.33 | 49 |
| SROSENBR | 2 | 0.00 | 78 | 0.00 | 868 | 0.00 | 102 |
| SROSENBR | 4 | 0.00 | 158 | 0.00 | 21442 | 1.31E-03 | 1268 |
| SROSENBR | 10 | 0.00 | 375683 | 0.12 | 192318 | 0.05 | 4699 |
| SROSENBR | 20 | 0.00 | 60817 | 5.12 | 89926 | 4.27 | 1284 |
| TRIDIA | 2 | 0.00 | 246 | 0.00 | 169 | 0.00 | 24 |
| TRIDIA | 10 | 0.00 | 53483 | 22.71 | 858 | 22.74 | 30 |
| TRIDIA | 20 | 0.00 | 733 | 629.12 | 1222 | 636.32 | 24 |
| TRIGO | 2 | 0.00 | 114 | 0.00 | 145 | 0.00 | 21 |
| VARDIM | 2 | 0.00 | 78 | 0.00 | 170 | 0.00 | 24 |
| VARDIM | 10 | 0.00 | 6683 | 1.92 | 3187 | 1.58 | 100 |
| VARDIM | 20 | 0.00 | 574 | 4.78E+08 | 410 | 4.78E+08 | 10 |
| WOODS | 4 | 0.00 | 158 | 0.00 | 905 | 1.94E-04 | 60 |
| WOODS | 12 | 0.00 | 3926 | 59.43 | 3106 | 53.00 | 83 |
| WOODS | 20 | 0.00 | 60542 | 481.65 | 8509 | 449.89 | 139 |
| QUADRATIC | 2 | 0.00 | 86 | 0.00 | 44 | 0.00 | 6 |
| SC2 | 2 | 0.30 | 86 | 0.30 | 123 | 0.30 | 18 |

[48] Z. Zhao, J. C. Meza, and M. Van Hove, *Using pattern search methods for surface structure determination of nanomaterials*, J. Phys. Condens. Matter, 18 (2006), pp. 8693–8706.

TABLE 10

*Results for the 44 smooth functions, obtained by DDS-Search and DDS-SPG methods, using opportunistic polling on asymptotically dense poll sets, not considering neither the maximum number of iterations nor or function evaluations as stopping criteria. The feasible region is the intersection of a box, a half-space, and a sphere.*

| Problem | $n$ | $f(x_*)$ | DDS-Search | | DDS-SPG | | |
|---|---|---|---|---|---|---|---|
| | | unconstrained | $f_{evals}$ | $f_{best}$ | $f_{evals}$ | $f_{best}$ | $\nabla_S f_{evals}$ |
| ARWHEAD | 2 | 0.00 | 91 | 0.00 | 181 | 0.00 | 25 |
| ARWHEAD | 10 | 0.00 | 14462 | 6.32 | 1241 | 6.11 | 47 |
| ARWHEAD | 20 | 0.00 | 1139 | 292.16 | 1033 | 292.59 | 23 |
| BDQRTIC | 5 | 0.00 | 225 | 0.00 | 721 | 0.00 | 48 |
| BDQRTIC | 10 | 1187 | 6139 | 1104.35 | 1104 | 1094.55 | 44 |
| BDQRTIC | 20 | 35.41 | 1830 | 2.89E+04 | 2110 | 2.89E+04 | 42 |
| BDVALUE | 2 | 0.00 | 131 | 0.00 | 172 | 0.00 | 25 |
| BROWNAL | 2 | 0.00 | 171 | 0.00 | 279 | 0.00 | 37 |
| BROWNAL | 10 | 0.00 | 6220 | 34.67 | 492 | 35.85 | 21 |
| BROWNAL | 20 | 0.00 | 581 | 8.56E+11 | 581 | 8.55E+11 | 14 |
| BROYDN3D | 2 | 0.00 | 125 | 0.00 | 204 | 0.00 | 28 |
| CHROSEN | 2 | 0.00 | 893 | 0.00 | 286 | 1.22E-06 | 38 |
| FLETCHER | 2 | 1.82E-3 | 123 | 6.47E-06 | 222 | 6.06E-06 | 30 |
| GENPOWELL3 | 2 | 0.00 | 346 | 0.57 | 300 | 0.57 | 41 |
| GENPOWELL3 | 4 | 0.00 | 506 | 1.16 | 1090 | 1.16 | 79 |
| INTEGREQ | 2 | 0.00 | 129 | 0.00 | 197 | 0.00 | 27 |
| PENALTY1 | 3 | 1.52E-5 | 5259 | 1.99E-05 | 218 | 2.87E-05 | 23 |
| PENALTY1 | 4 | 2.25E-5 | 6079 | 2.79E-05 | 178 | 5.93E-05 | 15 |
| PENALTY1 | 10 | 7.09E-5 | 650 | 171.17 | 598 | 171.04 | 25 |
| PENALTY1 | 20 | 1.58E-4 | 650 | 171.17 | 598 | 171.04 | 25 |
| PENALTY2 | 4 | 1.52E-5 | 266 | 1.92E-05 | 438 | 1.91E-05 | 35 |
| PENALTY2 | 5 | 3.27E-5 | 369 | 3.54E-05 | 395 | 3.53E-05 | 27 |
| PENALTY2 | 10 | 4.00E-4 | 2521 | 3.04E+03 | 1085 | 2.99E+03 | 42 |
| PENALTY2 | 20 | 8.31E-3 | 3820 | 4.68E+05 | 2125 | 4.68E+05 | 41 |
| PENALTY3 | 2 | 0.00 | 125 | 0.49 | 161 | 0.00 | 23 |
| POWELLSG | 4 | 0.00 | 1582 | 3.66E-05 | 1037 | 5.49E-06 | 77 |
| POWELLSG | 12 | 0.00 | 27207 | 150.51 | 1990 | 129.27 | 64 |
| POWELLSG | 20 | 0.00 | 84779 | 964.41 | 3546 | 947.43 | 73 |
| SROSENBR | 2 | 0.00 | 89 | 0.00 | 858 | 1.33E-06 | 104 |
| SROSENBR | 4 | 0.00 | 37425 | 8.89E-04 | 22240 | 6.69E-04 | 1475 |
| SROSENBR | 10 | 0.00 | 618265 | 0.15 | 167315 | 0.07 | 4916 |
| SROSENBR | 20 | 0.00 | 333288 | 5.00 | 113456 | 4.03 | 1747 |
| TRIDIA | 2 | 0.00 | 172 | 0.00 | 206 | 0.00 | 29 |
| TRIDIA | 10 | 0.00 | 4561 | 25.17 | 998 | 22.34 | 36 |
| TRIDIA | 20 | 0.00 | 662 | 631.12 | 1722 | 628.32 | 35 |
| TRIGO | 2 | 0.00 | 110 | 0.00 | 160 | 0.00 | 23 |
| VARDIM | 2 | 0.00 | 89 | 0.00 | 174 | 0.00 | 24 |
| VARDIM | 10 | 0.00 | 16244 | 1.79 | 3390 | 1.61 | 125 |
| VARDIM | 20 | 0.00 | 578 | 4.78E+08 | 414 | 4.78E+08 | 10 |
| WOODS | 4 | 0.00 | 11039 | 1.03E-03 | 6097 | 2.79E-04 | 428 |
| WOODS | 12 | 0.00 | 9410 | 56.44 | 4945 | 52.31 | 146 |
| WOODS | 20 | 0.00 | 68174 | 473.87 | 7302 | 464.49 | 135 |
| QUADRATIC | 2 | 0.00 | 94 | 0.00 | 43 | 0.00 | 6 |
| SC2 | 2 | 0.30 | 94 | 0.30 | 141 | 0.30 | 20 |