Worst Case Complexity of Direct Search

L. N. Vicente*

May 13, 2010

Abstract

In this paper we prove that direct search of directional type shares the worst case complexity bound of steepest descent when sufficient decrease is imposed using a quadratic function of the step size parameter. This result is proved under smoothness of the objective function and using a framework of the type of GSS (generating set search). We also discuss the worst case complexity of direct search when only simple decrease is imposed and when the objective function is non-smooth.

Keywords: derivative-free optimization, direct search, worst case complexity, sufficient decrease

1 Introduction

It was shown by Nesterov [11, Page 29] that the steepest descent method, for unconstrained optimization, takes at most $\mathcal{O}(\epsilon^{-2})$ iterations to drive the norm of the gradient of the objective function below ϵ. Similar bounds have been proved by Gratton, Toint, and co-authors [8, 9] for trust-region methods and by Cartis, Gould, and Toint [3] for adaptive cubic overestimation methods, when these algorithms are based on a Cauchy decrease condition. Cartis, Gould, and Toint [4] have shown that such a bound is tight for the steepest descent method by presenting an example where the number of iterations is arbitrarily close to it (these authors have also shown that Newton's method

*CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt). Support for the author was provided by FCT under the grant PTDC/MAT/098214/2008. This work was developed while the author was visiting the Courant Institute of Mathematical Sciences of New York University under a FLAD scholarship.

can also take a number of iterations arbitrarily close to $\mathcal{O}(\epsilon^{-2})$). The above bound reduces to $\mathcal{O}(\epsilon^{-\frac{3}{2}})$ for the cubic regularization of Newton's method (see Nesterov and Polyak [12]) and for the adaptive cubic overestimation method (see Cartis, Gould, and Toint [3]).

Direct search-methods that poll using positive spanning sets are directional methods of descent type, and it is reasonable to expect that they share the same worst complexity bound of steepest descent provided new iterates are only accepted based on a sufficient decrease condition (see Section 2 for a description of these and related methods). In fact, we prove in Section 3 of this paper that such generating set search (GSS) type methods take at most $\mathcal{O}(\epsilon^{-2})$ iterations to drive the norm of the gradient of the objective function below ϵ. Interestingly, this bound is achieved when using a quadratic function of the step size in the sufficient decrease condition, i.e., a function like $\rho(t) = ct^2$, $c > 0$ — which corroborates previous numerical experience [13] where different functions of the form $\rho(t) = ct^p$ (with $2 \neq p > 1$) where tested but leading to a worse performance.

Once we allow new iterates to be accepted based simply on simple decrease (and therefore revert to integer lattices for globalization), such complexity bound seems only provable if the objective function satisfies an appropriate decrease rate (see Section 4). Deviation from smoothness also poses problems to the derivation of an upper complexity bound (see Section 5), as it is unclear how many directions (from a dense set in the unit sphere) must be tried until a descent one is found.

The paper focuses on the unconstrained minimization of an n-dimensional real function. Extension to the constrained case is achievable, in principle, for certain type of constraints like bound and linear constraints.

2 Direct-search algorithmic framework

We will follow the algorithmic description in [6, Chapter 7] for direct-search methods of directional type. Such framework can describe the main features of pattern search, generalized pattern search (GPS), mesh adaptive direct search (MADS), and generating set search (GSS). Following GPS and MADS, each iteration of the algorithm is organized around a search step (optional) and a poll step. Following GSS, we include provision to accept new iterates based on a sufficient decrease condition which uses a forcing function.

Following the terminology in [10], $\rho : (0, +\infty) \to (0, +\infty)$ will represent a forcing function, i.e., a continuous and non decreasing function satisfying $\rho(t)/t \to 0$ when $t \downarrow 0$. Typical examples of forcing functions are $\rho(t) = ct^p$, for $p > 1$. To write the algorithm in general terms we will use $\bar{\rho}(\cdot)$ to either represent a forcing function $\rho(\cdot)$ or the constant, zero function. A relatively minor difference between the presentation below and the one in [6, Chapter 7] (see [13]) is the use of $\bar{\rho}(\alpha_k \|d_k\|)$ instead of $\bar{\rho}(\alpha_k)$.

The evaluation process of the poll step is opportunistic, meaning that one moves to a poll point once simple or sufficient decrease is found, depending on the variant being used.

Algorithm 2.1 (Directional direct-search method)

Initialization

Choose x_0 with $f(x_0) < +\infty$, $\alpha_0 > 0$, $0 < \beta_1 \leq \beta_2 < 1$, and $\gamma \geq 1$. Let $\mathcal{D}$ be a (possibly infinite) set of positive spanning sets.

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

1. **Search step:** Try to compute a point with $f(x) < f(x_k) - \bar{\rho}(\alpha_k)$ by evaluating the function f at a finite number of points (in a mesh if $\bar{\rho}(\cdot) = 0$, see the Appendix). If such a point is found then set $x_{k+1} = x$, declare the iteration and the search step successful, and skip the poll step.
2. **Poll step:** Choose a positive spanning set D_k from the set $\mathcal{D}$. Order the set of poll points $P_k = \{x_k + \alpha_k d \, : \, d \in D_k\}$. Start evaluating f at the poll points following the chosen order. If a poll point $x_k + \alpha_k d_k$ is found such that $f(x_k + \alpha_k d_k) < f(x_k) - \bar{\rho}(\alpha_k \|d_k\|)$ then stop polling, set $x_{k+1} = x_k + \alpha_k d_k$, and declare the iteration and the poll step successful. Otherwise declare the iteration (and the poll step) unsuccessful and set $x_{k+1} = x_k$.
3. **Mesh parameter update:** If the iteration was successful then maintain or increase the step size parameter: $\alpha_{k+1} \in [\alpha_k, \gamma \alpha_k]$. Otherwise decrease the step size parameter: $\alpha_{k+1} \in [\beta_1 \alpha_k, \beta_2 \alpha_k]$.

The global convergence of these methods is heavily based on the analysis of the behavior of the step size parameter α_k which must approach zero as an indication of some form of stationarity. There are essentially two known ways of enforcing the existence of a subsequence of step size parameters converging to zero (a refining subsequence) in direct search of directional type. One way is by ensuring that all new iterates lie on an integer lattice (rigorously speaking only when the step size is bounded away from zero). The other form consists of imposing a sufficient decrease on the acceptance

method	set $\mathcal{D}$ of directions
GPS [1], GSS simple decrease [10]	integer and finite (Ass. A.1)
MADS, non-smooth [2]	integer (Ass. A.4), dense in unit sphere
GSS suff. decrease, smooth [10]	bounded below cosine measure (Ass. 3.1)
suff. decrease, non-smooth [13]	dense in unit sphere

Table 1: Summary of the features of the directions used by the various direct-search methods of directional type.

method	search step	step size update	$\bar{\rho}(\cdot)$
GPS [1], GSS simple decrease [10]	in mesh (Ass. A.3)	integer (Ass. A.2)	$\bar{\rho}(\cdot) = 0$
MADS, non-smooth [2]	in mesh (Ass. A.3)	integer (Ass. A.2)	$\bar{\rho}(\cdot) = 0$
GSS suff. decrease, smooth [10]	any	any	forcing
suff. decrease, non-smooth [13]	any	any	forcing

Table 2: Summary of the remaining features of the various direct-search methods of directional type.

of new iterates, which can be simply achieved by selecting $\bar{\rho}(\cdot)$ as a forcing function in Algorithm 2.1.

Table 1 summarizes the properties of the set $\mathcal{D}$ of directions used by the different variants of this type of direct search. We include pointers to the literature and to the technical assumptions described in the appendix of this paper.

Note that the set of directions used for polling is not necessarily required to positively span $\mathbb{R}^n$ when this set of directions is (after normalization) dense in the unit sphere and one wants to derive results in a non-smooth context. In this setting, once we know that there exists a refining sequence, the convergence results are established along particular directions, called refining directions, which are already normalized. It is, in fact, the set of refining directions associated with the refining subsequence that is required to be dense in the unit sphere.

In Table 2 we continue the general description of the various variants of direct search in what concerns the search step, the step size update, and the type of decrease imposed to accept new iterates.

3 Complexity using sufficient decrease in the smooth case

It is relatively easy to derive complexity upper bounds on the number of successful and unsuccessful iterations for direct-search methods in the smooth case and when sufficient decrease is imposed (methods described in the third row of Tables 1 and 2). For this purpose, we need the following result, which is taken from [7, 10] (see also [6, Theorem 2.4 and Equation (7.14)]) and describes the relationship between the size of the gradient and the step size parameter at unsuccessful iterations. The results involves the cosine measure of a positive spanning set D_k (with nonzero vectors) which is defined by

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

Theorem 3.1 *Let D_k be a positive spanning set and $\alpha_k > 0$ be given. Assume that ∇f is Lipschitz continuous (with constant $\nu > 0$) in an open set containing all the poll points. If $f(x_k) \leq f(x_k + \alpha_k d)$, for all $d \in D_k$, then*

$$\|\nabla f(x_k)\| \;\leq\; \frac{\nu}{2}\,\mathrm{cm}(D_k)^{-1}\alpha_k\|d_k\| + \mathrm{cm}(D_k)^{-1}\frac{\rho(\alpha_k\|d_k\|)}{\alpha_k\|d_k\|}. \tag{1}$$

The class of methods under study in this section require $\mathcal{D}$ to satisfy the following assumption in order to achieve standard global convergence results (see [10]).

Assumption 3.1 *Any positive spanning set D_k in $\mathcal{D}$ satisfies $\mathrm{cm}(D_k) \geq \mathrm{cm}_{min} > 0$ and $0 < d_{min} \leq \|d\| \leq d_{max}$ for all $d \in D_k$.*

In the following theorem we provide a complexity upper bound on the number of successful iterations. Note that when $p = 2$, one has $p/\min(p-1,1) = 2$.

Theorem 3.2 *Consider the application of Algorithm 2.1 when $\bar{\rho}(t) = \rho(t) = ct^p$, $p > 1$, and D_k satisfies Assumption 3.1. Let f satisfy $f(x) \geq f_{low}$ for x in $L(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ and be continuously differentiable with Lipschitz continuous gradient on an open set containing $L(x_0)$.*

Given any $\epsilon \in (0,1)$, assume that $\|\nabla f(x_0)\| > \epsilon$ and let j_1 be the first iteration such that $\|\nabla f(x_{j_1+1})\| \leq \epsilon$. Then, to achieve $\|\nabla f(x_{j_1+1})\| \leq \epsilon$, Algorithm 2.1 takes at most

$$|S_{j_1}| \;\leq\; \left(\frac{f(x_0) - f_{low}}{L}\right)\epsilon^{-\frac{p}{\min(p-1,1)}}$$

successful iterations, where

$$L \;=\; c\, d_{min}^p \min\left(1, \beta_1^p \left(\frac{1}{\nu\, \mathrm{cm}_{min}^{-1}\, d_{max}/2 + \mathrm{cm}_{min}^{-1}\, c\, d_{max}^{p-1}}\right)^{\frac{p}{\min(p-1,1)}}\right).$$

Proof. Let us assume that $\|\nabla f(x_k)\| > \epsilon$, for $k = 0, \ldots, j_1$.

If $\alpha_k \geq 1$, then $\alpha_k \geq \epsilon$. Let us consider now the case where $\alpha_k < 1$. If k is the index of an unsuccessful iteration, one has from (1) that

$$\|\nabla f(x_k)\| \;\leq\; \frac{\nu}{2}\, \mathrm{cm}_{min}^{-1}\, d_{max}\alpha_k + \mathrm{cm}_{min}^{-1}\, c\, d_{max}^{p-1}\alpha_k^{p-1},$$

which then implies from $\alpha_k < 1$,

$$\epsilon \;\leq\; (\nu\, \mathrm{cm}_{min}^{-1}\, d_{max}/2 + \mathrm{cm}_{min}^{-1}\, c\, d_{max}^{p-1})\alpha_k^{\min(p-1,1)}.$$

Since at unsuccessful poll steps the step size is reduced by a factor of at most β_1 and it is not reduced at successful iterations, we can derive for any $k = 0, \ldots, j_1$, combining the two cases ($\alpha_k \geq 1$ and $\alpha_k < 1$) and considering that $\epsilon < 1$,

$$\alpha_k \;\geq\; \beta_1 \min\left(1, \left(\frac{1}{\nu\, \mathrm{cm}_{min}^{-1}\, d_{max}/2 + \mathrm{cm}_{min}^{-1}\, c\, d_{max}^{p-1}}\right)^{\frac{1}{\min(p-1,1)}}\right) \epsilon^{\frac{1}{\min(p-1,1)}}. \tag{2}$$

Let now k be the index of a successful iteration. From (2),

$$\begin{aligned} f(x_k) - f(x_{k+1}) \;&\geq\; c(\alpha_k \|d_k\|)^p \\ &\geq\; c\, d_{min}^p \min\left(1, \beta_1^p \left(\tfrac{1}{\nu\, \mathrm{cm}_{min}^{-1}\, d_{max}/2 + \mathrm{cm}_{min}^{-1}\, c\, d_{max}^{p-1}}\right)^{\frac{p}{\min(p-1,1)}}\right) \epsilon^{\frac{p}{\min(p-1,1)}}. \end{aligned}$$

We then obtain summing up for all successful iterations that

$$f(x_0) - f(x_{j_1}) \;\geq\; |S_{j_1}| L \epsilon^{\frac{p}{\min(p-1,1)}}$$

and the proof is completed. ■

Next, we bound the number of unsuccessful iterations.

Theorem 3.3 *Consider the application of Algorithm 2.1 when $\bar{\rho}(t) = \rho(t) = ct^p$, $p > 1$, and D_k satisfies Assumption 3.1. Let f satisfy $f(x) \geq f_{low}$ for x in $L(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ and be continuously differentiable with Lipschitz continuous gradient on an open set containing $L(x_0)$.*

Given any $\epsilon \in (0,1)$, assume that $\|\nabla f(x_0)\| > \epsilon$ and let j_1 be the first iteration such that $\|\nabla f(x_{j_1+1})\| \leq \epsilon$. Then, to achieve $\|\nabla f(x_{j_1+1})\| \leq \epsilon$, Algorithm 2.1 takes at most

$$|U_{j_1}| \;\leq\; L_1|S_{j_1}| + L_2 + \frac{\log\left(L_3 \epsilon^{\frac{1}{\min(p-1,1)}}\right)}{\log(\beta_2)}$$

unsuccessful iterations, where

$$L_1 \;=\; -\frac{\log(\gamma)}{\log(\beta_2)}, \qquad L_2 \;=\; -\frac{\log(\alpha_0)}{\log(\beta_2)},$$

and

$$L_3 \;=\; \beta_1 \min\left(1, \left(\frac{1}{\nu\,\mathrm{cm}_{min}^{-1}\, d_{max}/2 + \mathrm{cm}_{min}^{-1}\, c\, d_{max}^{p-1}}\right)^{\frac{1}{\min(p-1,1)}}\right).$$

Proof. Since either $\alpha_{k+1} \leq \beta_2 \alpha_k$ or $\alpha_{k+1} \leq \gamma \alpha_k$, we obtain by induction

$$\alpha_{j_1} \;\leq\; \alpha_0 \gamma^{|S_{j_1}|} \beta_2^{|U_{j_1}|},$$

which in turn implies from $\log(\beta_2) < 0$

$$|U_{j_1}| \;\leq\; -\frac{\log(\gamma)}{\log(\beta_2)}|S_{j_1}| - \frac{\log(\alpha_0)}{\log(\beta_2)} + \frac{\log(\alpha_{j_1})}{\log(\beta_2)}.$$

Thus, from $\log(\beta_2) < 0$ and the lower bound (2) on α_k, we obtain the desired result. ■

Combining Theorems 3.2 and 3.3, we can finally state, informally, the following corollary. Note, again, that $p/\min(p-1,1) = 2$ when $p = 2$.

Corollary 3.1 *Under the assumptions of Theorems 3.2 and 3.3, Algorithm 2.1 takes at most $\mathcal{O}\left(\epsilon^{-\frac{p}{\min(p-1,1)}}\right)$ iterations to reduce the gradient below $\epsilon \in (0,1)$.*

It has been proved in [4] that the upper bound $\mathcal{O}(\epsilon^{-2})$ for steepest descent is sharp by providing an example where a steepest descent method (with a Goldstein-Armijo linesearch) requires, for any $\epsilon \in (0,1)$, at least $\mathcal{O}(\epsilon^{-2+\tau})$ iterations to reduce the norm of the gradient below ϵ, where $\tau > 0$ is arbitrarily small. The example constructed in [4] was given for $n = 1$.

It turns out that in the unidimensional case, a direct-search method of the type given mentioned in the third row of Tables 1 and 2 (where sufficient decrease is imposed) can be casted as a steepest descent method with

Goldstein-Armijo linesearch, when the objective function is monotone decreasing (which happens to be the case in the example in [4]) and one considers the case $p = 2$. In fact, when $n = 1$, and up to normalization, there is essentially one positive basis, $\{-1, 1\}$. Thus, unsuccessful steps are nothing else than reductions of step size along the negative gradient direction. Also, since at unsuccessful iterations (see Theorem 3.1) one has $\|g_k\| \leq C_1 \alpha_k$ for some positive constant C_1 and $g_k = \nabla f(x_k)$, and since successful iterations do not decrease the step size, one obtains $\alpha_k \geq C_2 \|g_k\|$ for some constant $C_2 \in (0, 1)$. By setting $\gamma_k = \alpha_k / \|g_k\|$, one can then see that successful iterations take the form $x_{k+1} = x_k - \gamma_k g_k$ with $f(x_{k+1}) \leq f(x_k) - c\, C_2 \gamma_k \|g_k\|^2$.

4 Complexity using integer lattices in the smooth case

Let us recall an auxiliary result [1] (see also [6, Lemma 7.6]) about the globalization of direct-search methods using integer lattices which says that the minimum distance between any two distinct points in the mesh M_k (see Assumption A.3) is bounded below by a multiple of the step size parameter α_k.

Lemma 4.1 *Let Assumption A.1 hold. For any integer $k \geq 0$, one has that*

$$\min_{\substack{y, w \in M_k \\ y \neq w}} \|y - w\| \;\geq\; \frac{\alpha_k}{\|G^{-1}\|}.$$

Thus, if the objective function exhibits in successful iterations a decrease of the form

$$f(x_k) - f(x_{k+1}) \;\geq\; \theta \|x_k - x_{k+1}\|^p, \tag{3}$$

for some $\theta > 0$, then, from Lemma 4.1, one would get

$$f(x_k) - f(x_{k+1}) \;\geq\; (\theta \|G^{-1}\|^{-p})\, \alpha_k^p,$$

and one would be able to prove results of the form of Theorems 3.2 and 3.3 for the case where $\bar{\rho}(\cdot) = 0$ and the globalization is guaranteed by integer lattices (methods mentioned in the first two rows of Tables 1 and 2). Without a condition of the form (3), it seems unlikely to derive a worst complexity bound for the simple decrease case.

5 Complexity using sufficient decrease in the non-smooth case

Conceptually, a derivation of a complexity upper bound in the non-smooth case seems also possible but under specific strong assumptions to overcome some inherent technical difficulties. In the non-smooth case, the stopping condition $\|\nabla f(x_{j_1+1})\| \leq \epsilon$ can be replaced by $f^\circ(x_{j_1+1}; v) \geq -\epsilon$ for all directions $v \in \mathbb{R}^n$, where $f^\circ(x; d)$ represents the Clarke generalized directional derivative at x along d (which is guaranteed to exists if f is Lipschitz continuous near x). The first step in the analysis is to extend (1) from the smooth to the non-smooth case. However, ultimately, what we need to show is that when k is the index of an unsuccessful iteration and $f^\circ(x_k; v_k) \leq -\epsilon$ for some direction $v_k \in \mathbb{R}^n$, the step size α_k is bounded below by a constant times an appropriate power of ϵ.

Towards this goal, since the iteration k under consideration is unsuccessful,

$$f(x_k + \alpha_k d_k) \;\geq\; f(x_k) - \rho(\alpha_k \|d_k\|),$$

for all $d_k \in D_k$. We are interested in the direction d_k closest to v_k. By applying the Lebourg Theorem [5, Theorem 2.3.7],

$$\langle g(x_k + t_k(\alpha_k d_k)), \alpha_k d_k \rangle \;\geq\; -\rho(\alpha_k \|d_k\|),$$

where $g(x_k + t_k(\alpha_k d_k)) \in \partial f(x_k + t_k(\alpha_k d_k))$ and $t_k \in (0, 1)$. Now, let us assume the existence of $h(x_k) \in \partial f(x_k)$ such that $\|g(x_k + t_k(\alpha_k d_k)) - h(x_k)\| \leq \nu_1 \|t_k(\alpha_k d_k)\|$. Under this condition,

$$\begin{aligned} \nu_1 \|\alpha_k d_k\|^2 &\geq -\langle h(x_k), \alpha_k d_k \rangle - \rho(\alpha_k \|d_k\|) \\ &= -\|\alpha_k d_k\| \langle h(x_k), d_k / \|d_k\| \rangle - \rho(\alpha_k \|d_k\|). \end{aligned}$$

Thus, we obtain

$$-f^\circ(x_k; d_k/\|d_k\|) \;\leq\; \nu_1 \|\alpha_k d_k\| + \frac{\rho(\alpha_k \|d_k\|)}{\alpha_k \|d_k\|}.$$

We have thus bounded the Clarke derivative but along $d_k/\|d_k\|$ not v_k.

Assuming that v_k satisfies $\|v_k - d_k/\|d_k\|\| \leq r_k$, from the properties of the Clarke generalized directional derivative,

$$-\nu_2 r_k - f^\circ(x_k; v_k) \;\leq\; \nu_1 \|\alpha_k d_k\| + \frac{\rho(\alpha_k \|d_k\|)}{\alpha_k \|d_k\|},$$

where $\nu_2 > 0$ is the Lipschitz constant of f around x_k. So, we can see that it is possible to derive, under the condition stated above, a complexity upper bound of the order of

$$(\epsilon - \nu_2 \max_{k \in \{0,\ldots,j_1\}} r_k)^{-\frac{p}{\min(p-1,1)}}.$$

(Note that this result could have been derived assuming instead $|f^\circ(x_k + t_k(\alpha_k d_k); \alpha_k d_k) - f^\circ(x_k; \alpha_k d_k)| \leq \nu_1 \|\alpha_k d_k\|^2$.) It is then obvious that one must have $\nu_2 \max_{k \in \{0,\ldots,j_1\}} r_k < \epsilon$ which is not likely to happen.

6 Final remarks

The study of worst case complexity of direct search (of directional type) brings new insights about the differences and similarities of the various methods and their theoretical limitations. Without new 'proof technology', it seems only possible to establish a worst case complexity bound for those direct-search methods based on the acceptance of new iterates by a sufficient decrease condition (using a forcing function of the step size parameter) and when applied to smooth functions.

It is interesting to note that direct-search methods which poll using a set of directions dense in the unit sphere allow the derivation of global convergence results for non-smooth functions without specifically requiring the use of a positive spanning set for polling (see [2, 13]). However, they seem not capable of offering a complexity bound (even for the smooth case) without polling at a positive spanning set.

A Appendix

Generalized pattern search (GPS) makes use of a finite set of directions $D = \mathcal{D}$ which satisfy appropriate integrality requirements for globalization by integer lattices.

Assumption A.1 *The set D of positive spanning sets is finite and the elements of D are of the form $G\bar{z}_j$, $j = 1, \ldots, |D|$, where $G \in \mathbb{R}^{n \times n}$ is a nonsingular matrix and each $\bar{z}_j$ is a vector in $\mathbb{Z}^n$.*

In addition, the update of the step size parameter must conform to some form of integrality.

Assumption A.2 *The step size parameter is updated as follows: Choose a rational number* $\tau > 1$*, a nonnegative integer* $m^+ \geq 0$*, and a negative integer* $m^- \leq -1$*. If the iteration is successful, the step size parameter is maintained or increased by taking* $\alpha_{k+1} = \tau^{m_k^+}\alpha_k$*, with* $m_k^+ \in \{0, \ldots, m^+\}$*. Otherwise, the step size parameter is decreased by setting* $\alpha_{k+1} = \tau^{m_k^-}\alpha_k$*, with* $m_k^- \in \{m^-, \ldots, -1\}$*.*

Note that these rules respect those of Algorithm 2.1 by setting $\beta_1 = \tau^{m^-}$, $\beta_2 = \tau^{-1}$, and $\gamma = \tau^{m^+}$.

Finally, the search step is restricted to points in a previously (implicitly defined) mesh or grid.

Assumption A.3 *The search step in Algorithm 2.1 only evaluates points in*

$$M_k \;=\; \bigcup_{x \in S_k} \{x + \alpha_k D z : z \in \mathbb{N}_0^{|D|}\},$$

where S_k *is the set of all the points evaluated by the algorithm previously to iteration* k*.*

Note that poll points must also lie on the mesh, but this requirement is trivially satisfied from the definition of the mesh M_k given below (i.e., one trivially has $P_k \subset M_k$).

Generating set search (GSS) when based on globalization by integer lattices requires a similar assumptions as the three above.

For non-smooth and discontinuous objective functions, ones needs to make use of an infinite set of directions $\mathcal{D}$ dense (after normalization) in the unit sphere. MADS makes use of such a set of directions but, since it is also based on globalization by integer lattices, the set $\mathcal{D}$ must then be generated from a finite set D satisfying Assumption A.1 (which will be guaranteed by the first requirement of the next assumption).

Assumption A.4 *Let* D *represent a finite set of positive spanning sets satisfying Assumption A.1.*

The set $\mathcal{D}$ *is so that the elements* $d_k \in D_k \subseteq \mathcal{D}$ *satisfy the following conditions:*

1. d_k *is a nonnegative integer combination of the columns of* D*.*
2. *The distance between* x_k *and the point* $x_k + \alpha_k d_k$ *tends to zero if and only if* α_k *does:*

$$\lim_{k \in K} \alpha_k \|d_k\| \;=\; 0 \iff \lim_{k \in K} \alpha_k \;=\; 0,$$

   for any infinite subsequence K*.*

3. *The limits of all convergent subsequences of* $\bar{D}_k = \{d_k/\|d_k\| : \; d_k \in D_k\}$ *are positive spanning sets for* $\mathbb{R}^n$*.*

Acknowledgments

The author would like to thank Michael L. Overton and Mert Gurbuzbalaban (New York University) for interesting discussions on the topic of the paper.

References

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

[2] C. Audet and J. E. Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.*, 17:188–217, 2006.

[3] N. I. M. Gould C. Cartis and Ph. L. Toint. Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity. *Math. Program.*, 2010, to appear.

[4] C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization. Technical Report 08/14, Dept. of Mathematics, FUNDP, Namur, 2009.

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

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

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

[8] S. Gratton, M. Mouffe, Ph. L. Toint, and M. Weber-Mendonca. A recursive trust-region method in infinity norm for bound-constrained nonlinear optimization. *IMA J. Numer. Anal.*, 28:827–861, 2008.

[9] S. Gratton, A. Sartenaer, and Ph. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM J. Optim.*, 19:414–444, 2008.

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

[11] Yu. Nesterov. *Introductory Lectures on Convex Optimization.* Kluwer Academic Publishers, Dordrecht, 2004.

[12] Yu. Nesterov and B. T. Polyak. Cubic regularization of Newton's method and its global performance. *Math. Program.*, 108:177–205, 2006.

[13] L. N. Vicente and A. L. Custódio. Analysis of direct searches for discontinuous functions. Technical Report 09-38, Departamento de Matemática, Universidade de Coimbra, Portugal, 2009.