Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 08–48

IMPLICITLY AND DENSELY DISCRETE BLACK-BOX OPTIMIZATION PROBLEMS

L. N. VICENTE

Abstract: This paper addresses derivative-free optimization problems where the variables lie implicitly in an unknown discrete closed set. The evaluation of the objective function follows a projection onto the discrete set, which is assumed dense rather than sparse. Such a mathematical setting is a rough representation of what is common in many real-life applications where, despite the continuous nature of the underlying models, a number of practical issues dictate rounding of values or projection to nearby feasible figures.

We discuss a definition of minimization for these implicitly discrete problems and outline a direct search algorithm framework for its solution. The main asymptotic properties of the algorithm are analyzed and numerically illustrated.

Keywords: Derivative-free optimization, (dense) discrete optimization, direct search, projection, rounding, location, grids.
AMS Subject Classification (2000): 90C56, 90B80 90B85, 90C10, 90C11.

1. The implicitly and densely discrete problem

It is known that many optimization problems are only apparently continuous. The practical nature of many applications involves an underlying discrete (most of the times unknown) structure which is not taken explicitely in the modeling or solution phases and only revealed later when the result determined by some optimization process is really implemented. In other application problems, which are of interest to us in this paper, the determination of the value of the objective function is made by first 'rounding' the values of the variables to allowable figures or by 'projecting' them to nearby values or grid points where it is possible or desirable to evaluate the real function. The underlying discrete structure is thus unknown to the optimizer and its manifestation is detected only when a function evaluation is demanded.

Date: September 26, 2008.
Support for this work was provided by FCT under grants POCI/MAT/59442/2004 and PTDC/MAT/64838/2006.

We can pose the problem under consideration as follows

$$\begin{array}{ll} \min_{x \in \mathcal{L}} & f(x) \\ \text{s.t.} & x \in \Omega, \end{array} \tag{1}$$

where $\Omega \subset \mathbb{R}^n$ is some feasible region and $\mathcal{L}$ is an unknown set in $\mathbb{R}^n$.

We ask $\mathcal{L}$ to meet the following requirements:

(L1) The set is discrete (i.e., for every $x \in \mathcal{L}$ there exists a neighborhood $\mathcal{N}$ such that $\mathcal{L} \cap \mathcal{N} = \{x\}$).

(L2) The set is closed (and thus $\mathcal{L}$ does not have any accumulation points or, less formally, points in $\mathcal{L}$ cannot become arbitrarily close).

(L3) The distance between a point in $\mathbb{R}^n$ (or, more specifically, in Ω) and the closest point (or closest points) in $\mathcal{L}$ cannot be arbitrarily large.

Because of L1 and L2 the intersection of $\mathcal{L}$ with a compact set must necessarily be finite. Integer lattices are examples of sets satisfying conditions L1–L3. Let $P_{\mathcal{L}} : 2^{\mathbb{R}^n} \to 2^{\mathcal{L}}$ denote a (idempotent) projection operator onto $\mathcal{L}$.

We say that $x_* \in \Omega \cap \mathcal{L}$ is an implicitly and densely discrete local minimizer if for some $\sigma_{out} > \sigma_{in} > 0$, the following conditions are satisfied:

$$f(x_*) \leq f(x) \quad \forall x \in R_{out}, \tag{2}$$

$$P_{\mathcal{L}}\left(\{y \in \mathbb{R}^n : \|y - x_*\| \leq \sigma_{in}\} \cap \Omega\right) \;=\; \{x_*\}, \tag{3}$$

$$f(x_*) \leq f(x) \quad \forall x \in R_{between}, \tag{4}$$

where

$$R_{out} \;=\; P_{\mathcal{L}}\left(\{y \in \mathbb{R}^n : \|y - x_*\| = \sigma_{out}\}\right) \cap \Omega$$

and

$$R_{between} \;=\; [(\Omega \cap \mathcal{L}) \cap \{y \in \mathbb{R}^n : \sigma_{in} < \|y - x_*\| < \sigma_{out}\}] \setminus R_{out}.$$

We depict an example of the sets R_{out} and $R_{between}$ in Figure 1.

The definition of implicitly and densely discrete local minimization implies a natural definition of local minimization (when $P_{\mathcal{L}}$ is orthogonal), in the sense that if $x_* \in \Omega \cap \mathcal{L}$ is an implicitly and densely discrete local minimizer there is a neighborhood $\mathcal{N}$ of x_* such that $f(x_*) \leq f(x)$ for all $x \in \Omega \cap \mathcal{L} \cap \mathcal{N}$ and $\Omega \cap \mathcal{L} \cap \mathcal{N} \neq \{x_*\}$.

Note also that the existence of a $\sigma_{in} > 0$ such that condition (3) is satisfied is trivially guaranteed given the properties of $\mathcal{L}$. However, our definition splits the main properties of local minimization into (2) and (4) and thus introduces $R_{between}$, which in turn depends on σ_{in}.

These definitions of local minimization might not be appropriate for sparsely discrete problems. According to our definition, for example, all the points in $\mathcal{L} \cap \Omega = \{-13, -8, -4, -1, 1, 4, 8, 13\}$ are local minimizers for $f(x) = -|x|$ when $P_{\mathcal{L}}$ is orthogonal. This example was pointed out to us by Audet [2] who suggested an alternative definition which does not break down for this example. However, in this paper we have in mind densely discrete optimization problems for which the definition above seems appropriate. Also, as will see in this paper, the definition of implicitly and densely discrete local minimization suits the convergence purposes of a vast class of direct search methods.

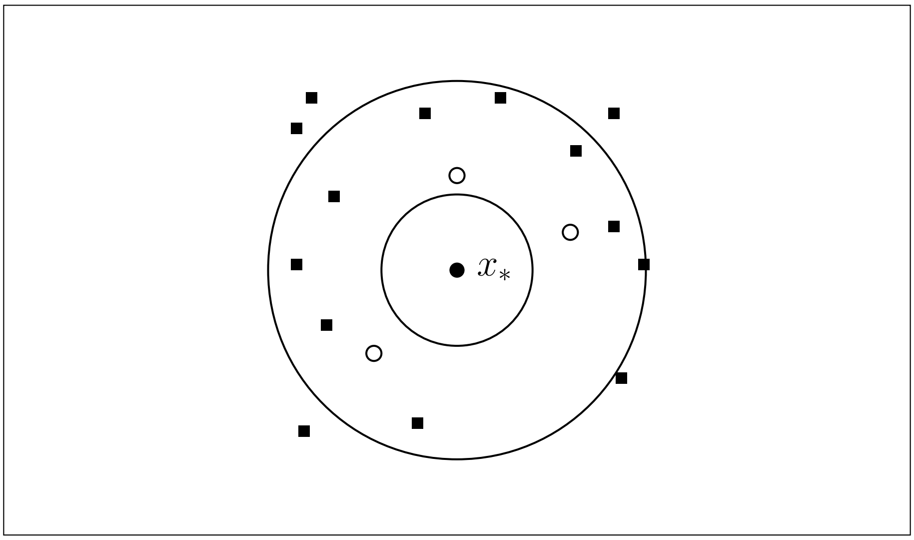

FIGURE 1. Example where the filled squares correspond to points in R_{out} and the empty circles to points in $R_{between}$. The circles have radii σ_{in} (inner) and σ_{out} (outer).

When dealing with an algorithmic context we assume that the projection operator returns always a singleton when operating on a single point and that ties, when occur, are broken in some application dependent criterion. In such situations, $P_{\mathcal{L}}(x) = y$ corresponds to $P_{\mathcal{L}}(\{x\}) = \{y\}$. For simplicity, we will consider that $P_{\mathcal{L}}$ is an orthogonal projector but this can be relaxed. In fact, we will see that all is needed is that when the distance between a point in $\mathbb{R}^n$ and its projection become arbitrarily large then Condition L3 is violated.

2. A direct search approach

Below we describe a general direct search type method for the solution of the implicitly and densely discrete optimization problem (1). We make use of the extreme barrier function:

$$f_\Omega(x) \;=\; \begin{cases} f(x) & \text{if } \; x \in \Omega, \\ +\infty & \text{otherwise.} \end{cases}$$

Algorithm 2.1 (**Direct search algorithm**).

Step 0 (initialization): Let $x_0 \in \Omega \cap \mathcal{L}$, $\alpha_0 > 0$, and $c \in (0,1)$.

Step 1 (polling): Select a set of directions D_k (not necessarily a positive spanning set).

For all $d \in D_k$:

- If $\|P_{\mathcal{L}}(x_k + \alpha_k d) - x_k\| \leq c\,\alpha_k$ then stop polling, set $x_{k+1} = x_k$ (and, for later presentation, set $v_k = d$), increase α_k, increment k by one unit, and return to the beginning of Step 1.
- Otherwise ($\|P_{\mathcal{L}}(x_k + \alpha_k d) - x_k\| > c\,\alpha_k$) then
  - If $f_\Omega(P_{\mathcal{L}}(x_k + \alpha_k d)) < f(x_k)$ then stop polling, set $x_{k+1} = P_{\mathcal{L}}(x_k + \alpha_k d)$ (and, for later presentation, set $d_k = d$), increase α_k or kept it constant, increase k by one unit, and return to the beginning of Step 1.
  - Otherwise continue the polling loop.

Step 2 (unsuccessful polling): If for all $d \in D_k$ it happened that $f_\Omega(P_{\mathcal{L}}(x_k + \alpha_k d)) \geq f(x_k)$ and $\|P_{\mathcal{L}}(x_k + \alpha_k d) - x_k\| > c\,\alpha_k$ then set $x_{k+1} = x_k$, decrease α_k, increment k by one unit, and return to the beginning of Step 1.

Note that the sequence of iterates $\{x_k\}$ generated by this algorithm necessarily lies in $\Omega \cap \mathcal{L}$.

We ask the sets of directions D_k to satisfy the following assumptions:

(D1) $\|d\| \geq 1 \;\forall d \in D_k$, $\forall k$.
(D2) $\alpha_k d \to 0 \;\forall d \in D_k$ whenever $\alpha_k \to 0$ for some subsequence.
(D3) $\displaystyle\bigcup_{k \in K} D_k$ is dense in the unit sphere of $\mathbb{R}^n$ for all subsequences K for which $\alpha_k \to \alpha > 0$.

These conditions, in particular D2, are compatible with those imposed on positive spanning sets D_k by mesh adaptive direct search methods [3] for continuous problems.

Also, one can see that the only difference between the standard polling procedure of directional direct search methods and the polling scheme of Algorithm 2.1 is the introduction of the test $\|P_{\mathcal{L}}(x_k + \alpha_k d) - x_k\| \leq c\,\alpha_k$ based on which we detect whether the algorithmic mesh is too fine when compared to $\mathcal{L}$.

3. Analysis of the direct search method

Now we analyze the convergence properties of Algorithm 2.1. Essentially we will show that part of the conditions of implicitly and densely discrete local minimization are asymptotically satisfied.

Theorem 3.1. *Let $\{x_k\}$ be a sequence of iterates generated by Algorithm 2.1 where the sets of polling directions satisfy Assumptions D1–D3.*

If $L(x_0) = \{x \in \mathbb{R}^n : \ f(x) \leq f(x_0)\}$ is bounded then there exist $x_* \in \Omega \cap \mathcal{L}$ and subsequences K_{in} and K_{out} and positive numbers $0 < \alpha_{in} < \alpha_{out}$ such that $\alpha_k \to \alpha_{out}$ when $k \in K_{out}$ and

$$f_\Omega(P_{\mathcal{L}}(x_* + \alpha_k d)) \ \geq \ f(x_*) \quad \forall d \in D_k,\ \forall k \in K_{out}, \tag{5}$$

where $\cup_{k \in K_{out}} D_k$ is dense in the unit sphere in $\mathbb{R}^n$, and $\alpha_k \to \alpha_{in}$ when $k \in K_{in}$ and

$$\exists_{v_k \in D_k, \|v_k\| \geq 1} : \ \|P_{\mathcal{L}}(x_* + \alpha_k v_k) - x_*\| \ \leq \ c\,\alpha_{in} \quad \forall k \in K_{in}, \tag{6}$$

with $\|v_k\| \to 1$ in K_{in}.

Proof: First we note that because $L(x_0)$ is bounded the set $\mathcal{L} \cap L(x_0)$ is finite. Since the algorithm only moves to a new point in $\mathcal{L}$ when a decrease is found there must exists a $\bar{k}$ such that

$$x_k \ = \ x_{\bar{k}} \ = \ x_* \quad \forall k \geq \bar{k}.$$

Let us first prove that $\{\alpha_k\}$ is a bounded sequence. For this purpose let us assume that there exists a subsequence driving α_k to $+\infty$. If that is the case, there must exists another subsequence denoted by K where α_k is increased and $\alpha_k \to +\infty$ for $k \in K$. From the algorithm we have $\|P_{\mathcal{L}}(x_* + \alpha_k v_k) - x_*\| \leq c\,\alpha_k$ for k sufficiently large in K, and it can be easily proved for such values of k that

$$\|(x_* + \alpha_k v_k) - P_{\mathcal{L}}(x_* + \alpha_k v_k)\| \ \geq \ (\|v_k\| - c)\,\alpha_k.$$

Using Assumption D1 and taking $\alpha_k \to +\infty$ for $k \in K$ in both sides of this inequality contradict Condition L3 of the definition of $\mathcal{L}$ (note that it is here that one needs to qualify the projection operator).

It can also be proved that α_k is uniformly bounded away from zero. In fact, if there was a subsequence driving α_k to zero then there would exist another subsequence denoted by J for which α_k is decreased (which then means that $\|P_{\mathcal{L}}(x_* + \alpha_k d) - x_*\| > c\,\alpha_k > 0$ for k sufficiently large) and $\alpha_k \to 0$ for $k \in J$. From Assumption D2 we obtain $\alpha_k d \to 0$ in J for some $d \in D_k$. Thus, we derive that

$$x_* + \alpha_k d \;\to\; x_* \in \mathcal{L} \quad \text{and} \quad P_{\mathcal{L}}(x_* + \alpha_k d) \;\neq\; x_* \;\; \forall k \in J,$$

which contradicts conditions L1–L2 of the definition of $\mathcal{L}$.

Since α_k is uniformly bounded from above and away from zero, there must exist subsequences K_{in} and K_{out} and positive numbers $0 < \alpha_{in} < \alpha_{out}$ such that $\alpha_k \to \alpha_{out}$ and α_k is decreased for all $k \in K_{out}$ and $\alpha_k \to \alpha_{in}$ and α_k is increased for all $k \in K_{in}$. Thus,

$$f_\Omega(P_{\mathcal{L}}(x_* + \alpha_k d)) \;\geq\; f(x_*) \quad \forall d \in D_k,\; \forall k \in K_{out},$$

and

$$\|P_{\mathcal{L}}(x_* + \alpha_k v_k) - x_*\| \;\leq\; c\,\alpha_k \quad \text{for some } v_k \in D_k,\; \forall k \in K_{in}.$$

The proof is completed by using Assumptions D1 and D3. ■

It is clear that the point x_* identified in this theorem satisfies a condition (see (5)) which is practically the same as (2), with $\sigma_{out} = \alpha_{out}$.

Condition (3) is roughly approximated by (6) if the constant c is small and $\{v_k, k \in K_{in}\}$ is dense in the unit sphere. The absolute satisfaction of (3) would require some form of dense sampling close to x_*.

What is clearly missing in the result of Theorem 3.1 is a condition of type (4). However, to capture the points in $R_{between}$ seems a rather difficult task for any reasonable algorithmic framework.

4. A numerical illustration

We made some experiments in MATLAB [1] to observe and possibly confirm our theoretical findings. The implementation chosen for Algorithm 2.1 was rather simple. The set D_k is set to the positive basis $[Q_k \;\; -Q_k]$ where Q_k is an orthonormal matrix computed by first randomly generating the first column (independently of α_k). We report here some results obtained for the minimization in $\mathcal{L}$ of a quadratic function perturbed by oscillatory noise:

$$\begin{aligned}\min_{(x_1,x_2)\in\mathcal{L}} \quad & 10(x_1^2)(1+0.75\cos(70x_1)/12)+\cos(100x_1)^2/24+\\ & 2(x_2^2)(1+0.75\cos(70x_2)/12)+\cos(100x_2)^2/24+4x_1x_2.\end{aligned}$$

The unique minimizer of this problem is $x_* = (0,0)$. The set $\mathcal{L}$ has been chosen as the integer lattice $\{\gamma z : z \in \mathbb{Z}^n\}$, with $\gamma \neq 0$. In Figure 2, we plot the behavior of the mesh or step size parameter α_k. One can easily observe in both plots that α_k oscillates between an upper and a lower value as predicted by Theorem 3.1. For these experiments we ran the algorithm for a specified number of iterations (500 in the first case and 1000 in the second one). Introducing a stopping criterion for the algorithm would require some practical rules to approximate the inferior and superior limits of the sequence $\{\alpha_k\}$.

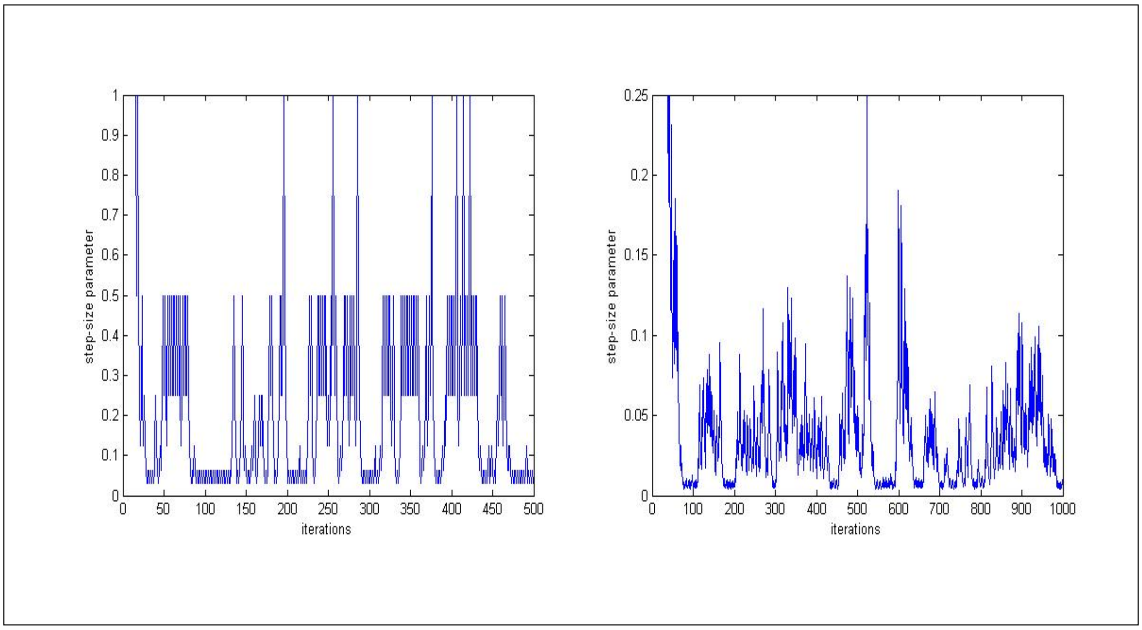

FIGURE 2. Illustration of the behavior of the mesh or step size parameter α_k in Algorithm 2.1. For the run reported on the left we chose $\gamma = 0.1$ and $c = 0.95$, increased α_k by a factor of 2 (in both occurrences of the algorithm), and decreased it by a factor of 0.5 in unsuccessful polling steps. The plot on the right corresponds to a finer lattice ($\gamma = 0.01$) and a decreasing factor of 0.75. Both runs started from $x_0 = (5,-10)$ and $\alpha_0 = 1$.

5. Discussion and open issues

This paper is a first attempt to shed some light on the solution of implicitly discrete black-box optimization problems. These problems are of the derivative-free type (see [4]) but characterized by the existence of an implicit and unknown discrete set (assumed relatively dense) where optimization points are first projected before the objective function is evaluated. In this paper we suggested a reasonable working definition of local minimization which suits the needs of direct search methods. The proposed direct search method is directional and resembles mesh adaptive direct search methods [3] when '$\mathcal{L} \to \mathbb{R}^n$'. The denser the set $\mathcal{L}$ is in $\mathbb{R}^n$ the smaller the value of σ_{in} becomes. When one suspects that $\mathcal{L}$ is 'sufficiently dense' in $\mathbb{R}^n$, the set of directions must also satisfy the property that for all subsequences driving α_k to zero the union of the D_k's is dense in $\mathbb{R}^n$.

A number of issues are of interest and have not been fully addressed here. The following is an attempt to enumerate some of them:

- Can we incorporate other mechanisms in the suggested direct search framework to look for points in $R_{between}$? Could we, for instance, target different layers instead of one (the one now corresponding to σ_{out})? We doubt, however, that it would be possible to capture all the points in $R_{between}$ without some form of dense layers in the unit ball.
- Would it be possible to develop stopping criteria which can be satisfied asymptotically without approximating the inferior and superior limits of the sequence $\{\alpha_k\}$?
- It is known that pattern search algorithms [4, 5] generate points in an integer lattice. Would it be possible for such type of direct search algorithms to 'align' their integer lattices with the application discrete structure? Under what conditions would that be achievable?
- We wanted the current approach to be as general as possible. We have not taken advantage, for instance, of the size of $\|P_{\mathcal{L}}(x_k + \alpha_k d) - (x_k + \alpha_k d)\|$ in the algorithm when compared to α_k.
- Is it possible to derive other definitions of local minimization for implicitly and densely discrete problems both capable of giving a satisfactory answer for sparsely discrete problems and of fitting the convergence needs of direct search methods?

References

[1] *MATLAB, The MathWorks Inc.* http://www.mathworks.com.

[2] C. Audet, *Private communication.*
[3] C. Audet and J. E. Dennis Jr., *Mesh adaptive direct search algorithms for constrained optimization*, SIAM J. Optim., 17 (2006), pp. 188–217.
[4] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.
[5] 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.

L. N. Vicente
CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt).