Optimization and Engineering, 5, 101–122, 2004

Convergence Results for Generalized Pattern Search Algorithms are Tight*

CHARLES AUDET
GERAD and École Polytechnique de Montréal, Département de Mathématiques et de Génie Industriel, C.P. 6079, Succ. Centre-ville, Montréal (Québec), H3C 3A7, Canada
email: charles.audet@gerad.ca

Received October 29, 2002; Revised March 12, 2003

Abstract. The convergence theory of generalized pattern search algorithms for unconstrained optimization guarantees under mild conditions that the method produces a limit point satisfying first order optimality conditions related to the local differentiability of the objective function. By exploiting the flexibility allowed by the algorithm, we derive six small dimensional examples showing that the convergence results are tight in the sense that they cannot be strengthened without additional assumptions, i.e., that certain requirement imposed on pattern search algorithms are not merely artifacts of the proofs.

In particular, we first show the necessity of the requirement that some algorithmic parameters are rational. We then show that, even for continuously differentiable functions, the method may generate infinitely many limit points, some of which may have non-zero gradients. Finally, we consider functions that are not strictly differentiable. We show that even when a single limit point is generated, the gradient may be non-zero, and zero may be excluded from the generalized gradient, therefore, the method does not necessarily produce a Clarke stationary point.

Keywords: pattern search algorithms, convergence analysis, unconstrained optimization, non-smooth analysis, Clarke derivatives

1. Introduction

Consider the unconstrained optimization problem of minimizing a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, without any knowledge of its derivatives, and without any means of approximating them accurately. Torczon (1997) observes that several existing search methods for this problem share a common structure, and they can be subsumed into a more general one. The pattern search algorithm defined there, and slightly generalized in Lewis and Torczon (1996), encompasses a wide class of algorithms, and still yields strong convergence results under the assumption of continuous differentiability of f. The algorithm produces a sequence of iterates $x_0, x_1, \ldots$ in $\mathbb{R}^n$ with monotone non-increasing objective function value. Torczon (1997) shows that when f is continuously differentiable then

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{1}$$

*This work was supported by FCAR grant NC72792, NSERC grant 239436-01 and AFOSR F49620-01-1-0013.

Booker et al. (1999) present an equivalent formulation of the algorithm, that they call the *Generalized Pattern Search* (GPS) algorithm. Audet and Dennis (2003) use Clarke's calculus (Clarke, 1990) to show a hierarchy of convergence behavior for GPS. They show that any limit point $\hat{x}$ (of any refining subsequence, see Section 2.1 below where GPS is fully described) satisfies some optimality condition that depends on local differentiability of the function f. Their main result is a necessary condition for optimality, namely that if f is Lipschitz near $\hat{x}$, then the generalized directional derivative

$$f^\circ(\hat{x}; d) = \limsup_{y \to \hat{x},\ t \downarrow 0} \frac{f(y + tv) - f(y)}{t}$$

is non-negative for some directions d forming a positive spanning set (i.e., non-negative linear combinations of the directions d span $\mathbb{R}^n$). This result is weaker than showing that $\hat{x}$ is a Clarke stationary point, i.e., that zero belongs to the generalized gradient

$$\partial f(\hat{x}) = \{s \in \Re^n : f^\circ(\hat{x}; v) \geq v^T s \text{ for all } v \in \Re^n\}.$$

Clarke stationarity is shown to be false in general for GPS in Section 5.2. For a Lipschitz function near $\hat{x}$, $0 \in \partial f(\hat{x})$ is equivalent to $f^\circ(\hat{x}; d) \geq 0$ for all directions $d \in \mathbb{R}^n$.

It is also shown in Audet and Dennis (2003) that if f is strictly differentiable at $\hat{x}$ (i.e., $\lim_{y \to \hat{x}, t \downarrow 0} \frac{f(y+tv)-f(y)}{t} = \nabla f(\hat{x})^T v$ for all $v \in \mathbb{R}^n$—therefore continuous differentiability implies strict differentiability), then $\nabla f(\hat{x}) = 0$. This last result is stronger than (1) since it only requires local strict differentiability instead of global continuous differentiability.

The contribution of this paper is to show through six examples in $\mathbb{R}^1$ or $\mathbb{R}^2$ that the theoretical convergence results presented in Torczon (1997) and Audet and Dennis (2003) cannot be strengthened without introducing additional assumptions on the objective function f or additional requirements on the algorithm.

These examples are constructed in a way that exploits the flexibility allowed by the GPS exploration of the space of variables. The fact that the flexibility allowed by the method limits the convergence results should not be thought as of a weakness of GPS, since the same flexibility is what often makes the algorithm work well in practice on some real problems (e.g., see Booker et al., 1999 where GPS is used jointly with surrogate functions to optimize the design of an helicopter rotor blade).

The paper is divided as follows. Section 2 briefly describes the GPS class of algorithms, and then highlights the key features of the examples presented in Sections 3, 4 and 5. Concluding remarks are presented in Section 6.

Notation. $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$ and $\mathbb{N}$ respectively denote the sets of real numbers, rational numbers, integers and non-negative integers. For a matrix D, the notation $d \in D$ indicates that d is a column of D.

2. Generalized pattern search

The reader is referred to Torczon (1997) and Lewis and Torczon (1996) for the original description of pattern search algorithm and its parameters. We adopt the notation for

GPS presented in Audet and Dennis (2003), and present a very brief description of the algorithm.

2.1. *Generalized pattern search algorithms*

Any instance of a GPS algorithm requires the following parameters:

– Initial point $x_0 \in \mathbb{R}^n$ (with finite value $f(x_0)$) and mesh size parameter $\Delta_0 > 0$ in $\mathbb{R}$.
– Non-singular generating matrix $G \in \mathbb{R}^{n\times n}$ and a $n \times m$ matrix $Z \subset \mathbb{Z}^{n\times m}$ whose columns form a positive spanning set.
– Mesh update parameters $\tau \in \mathbb{Q}$ and $w^-, w^+ \in \mathbb{Z}$ with $\tau > 1$, $w^- \leq -1$ and $w^+ \geq 0$.

Generalized pattern search algorithms generate a sequence of iterates $\{x_k\}$ in $\mathbb{R}^n$ with non-increasing objective function value. Each trial point where the algorithm evaluates f must lie on the current mesh, defined by

$$M_k = \{x_k + \Delta_k Dz : z \in \mathbb{N}^m\}, \quad \text{where } D = GZ, \tag{2}$$

and where $\Delta_k > 0$ is the mesh size parameter at iteration k (the rules for constructing it from Δ_{k-1} are presented below).

At each iteration, trial points can be generated in two steps. In the first step, called the SEARCH, any finite strategy can be used to find a mesh point that yields a lower objective function value than the incumbent as long as only finitely many trial points are selected. When a trial point $x_{k+1} \in M_k$ satisfying $f(x_{k+1}) < f(x_k)$ is found, then x_{k+1} is said to be an *improved mesh point*, and the SEARCH step can be halted.

The SEARCH step is optional: An empty SEARCH step consists in considering no trial points. The SEARCH is where most (but not all) of the GPS flexibility is located; it is there, for example, that the user knowledge of the optimization problem can be used and where surrogate functions (Booker et al., 1999) can be used to identify potentially good trial points. What we call a SEARCH point is what (Torczon, 1997) refers to as an exploratory move with step in $\Delta_k BL_k$.

Whenever the SEARCH step fails to generate an improved mesh point, a second step, called the POLL, is invoked before the iteration is completed. This second step consists of evaluating the objective function at the neighboring mesh points to see if a lower value can be found there. The set of neighboring mesh points (called the poll set) is constructed using a positive spanning matrix D_k composed of columns of the finite set D (defined in Eq. (2)):

$$\text{Poll set:} \quad \{x_k + \Delta_k d : d \in D_k\} \ \subset \ M_k. \tag{3}$$

If the poll set contains a trial point $x_{k+1} \in M_k$ such that $f(x_{k+1}) < f(x_k)$, then, as in the SEARCH step x_{k+1} is said to be an *improved mesh point*, and the POLL step can be halted.

If both the SEARCH and the POLL steps fail to generate an improved mesh point, then the current incumbent solution x_k is said to be a *mesh local optimizer* (i.e., its objective function value is less than or equal to that of all points from the poll set). The algorithm then refines the mesh by setting the mesh size parameter

$$\Delta_{k+1} = \tau^{w_k} \Delta_k \quad \text{where } w_k \in \mathbb{Z} \quad \text{and} \quad w^- \leq w_k \leq -1 \tag{4}$$

and therefore $0 < \tau^{w_k} < 1$. If either the SEARCH or POLL step produces an improved mesh point, then the new point $x_{k+1} \neq x_k$ has a strictly lower objective function value, the mesh size parameter is kept the same or is increased, and the process is reiterated. It follows that if x_{k+1} is an improved mesh point, than later iterates can never return to previous iterates: $x_j \neq x_\ell$ and $f(x_j) > f(x_\ell)$ for any $j \leq k$ and $\ell > k$. The coarsening of the mesh follows the rule

$$\Delta_{k+1} = \tau^{w_k} \Delta_k \quad \text{where } w_k \in \mathbb{N} \quad \text{and} \quad 0 \leq w_k \leq w^+ \tag{5}$$

and therefore $\tau^{w_k} \geq 1$. The algorithm is stated formally.

A basic GPS algorithm

- INITIALIZATION: Let x_0, Δ_0, G, Z, τ, w^- and w^+ satisfy the requirements given above. Set the iteration counter k to 0.
- SEARCH AND POLL STEP: Perform the SEARCH and possibly the POLL steps (or only part of them if an improved mesh point x_{k+1} is found on the mesh M_k defined by Eq. (2)).
  - OPTIONAL SEARCH: Evaluate f on a finite subset of trial points on the mesh M_k.
  - LOCAL POLL: Evaluate f on the poll set defined in Eq. (3).
- PARAMETER UPDATE: If the SEARCH or POLL step produced an improved mesh point, i.e., an $x_{k+1} \in M_k$ for which $f(x_{k+1}) < f(x_k)$, then update $\Delta_{k+1} \geq \Delta_k$ according to rule (5).
  Otherwise, $f(x_k) \leq f(x_k + \Delta_k d)$ for all $d \in D_k$ and so x_k is a mesh local optimizer; set $x_{k+1} = x_k$, update $\Delta_{k+1} < \Delta_k$ according to rule (4).
  Increase $k \leftarrow k + 1$ and go back to the SEARCH and POLL step.

The convergence results of Audet and Dennis (2003) hold for the limits of subsequences of mesh local optimizers located on meshes that get infinitely fine. These convergent subsequences are called *refining subsequences*. The POLL step is what ensures the convergence results (the SEARCH step cannot be used in these proofs since is free of any rules, except that it must be finite and on the mesh).

2.2. *Summary of the examples*

GPS convergence analyses that appear in the literature are divided into two phases. First, relying on the assumption that all iterates belong to a compact set, the limit inferior of the mesh size parameter is shown to converge to zero. The proof of this first phase is independent of the differentiability or continuity of f. In Torczon (1997) and Lewis and Torczon (1996) it is assumed that the level sets of f are bounded. This assumption is more restrictive than only assuming that all iterates belong to a compact set.

Then, local properties of the objective function f are used to establish appropriate convergence results. The examples presented in this paper validate the necessity of including certain assumptions in both these phases. None of these examples contradict any of the theoretical convergence results.

The first two examples show that lifting rationality restrictions on two of the GPS parameters can result in a mesh size parameter bounded away from zero, thereby nullifying the entire convergence theory. The first one, Example A, shows the necessity of ensuring that the positive spanning set D is the product of a non-singular real matrix and an integer matrix, while Example B shows that τ must be rational.

The two next examples illustrate how GPS can produce multiple limit points of the sequence of iterates, even when f is continuously differentiable. The first of these, Example C, shows that the gradient at one such limit point may be non-zero. Therefore, the main result of Torczon (1997) cannot be strengthened to

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{6}$$

Under stronger algorithmic requirements (restated here in Section 4.2), Torczon (1997) shows that (1) may be improved to (6). Example D shows that even under these additional restrictions, there still could be an infinite number of limit points.

Finally, in the last two examples, the sequence of iterates produced by GPS is convergent. Moreover, in both examples, the objective function is Lipschitz, differentiable and has a non-zero gradient at the limit point $\hat{x}$, but is not globally strictly differentiable (thus not globally continuously differentiable). Therefore, the convergence analysis of Audet and Dennis (2003) must be applied since that of Torczon (1997) requires global continuous differentiability. Furthermore, the function f in Example E is in $\mathbb{R}^1$ and therefore the Clarke generalized directional derivatives are non-negative in all directions in $\mathbb{R}^1$. In other words, the sequence of iterates converge to a Clarke stationary point where the gradient exists and is non-zero. Example F shows that zero does not necessarily lie in the generalized gradient at $\hat{x}$. This confirms that without strict differentiability, GPS cannot be guaranteed to produce a Clarke stationary point.

All the examples are constructed using the notation of Audet and Dennis (2003). However, each of them can be reconstructed using the notation in Torczon (1997) and the same behavior would be observed.

3. Rationality requirements

Rationality or integrality requirements appear twice in the description of GPS algorithms. The two examples of this section show that removing either of these requirements compromises the existence of refining subsequences.

3.1. *Example A: The positive spanning set D*

Each direction d_j in the positive spanning set D (for $j = 1, 2, \ldots, m$) is the product Gz_j of the non-singular generating matrix $G \in \mathbb{R}^{n \times n}$ by an integer vector $z_j \in \mathbb{Z}^n$ (note that this is equivalent to requiring that $z_j \in \mathbb{Q}^n$ since G is real). An example briefly outlined in Audet and Dennis (2003) validates this requirement by showing that without it, the sequence of mesh size parameters $\{\Delta_k\}$ can remain bounded away from zero; in fact Δ_k remains constant for all $k \in \mathbb{N}$. For completeness, we present the example here in full.

Consider the optimization problem in $\mathbb{R}^1$ with the continuously differentiable function $f(x) = x$, and parameters $x_0 = \pi - 3$, $\Delta_0 = 1$, $G = 1$, $Z = D = [-1, \ +\pi]$, $\tau = 2$, $w^- = -1$ and $w^+ = 0$. At iteration k, if $\Delta_k = 1$ and if there exist positive integers a_k and b_k such that $x_k = b_k\pi - a_k > 0$, then the SEARCH considers the trial point

$$t_k = x_k + \Delta_k((q_k - 1)b_k\pi - ((q_k - 1)a_k + 1)) \ = \ q_k x_k - 1 \tag{7}$$

where $q_k = \lceil \frac{1}{x_k} \rceil \in \mathbb{N} \setminus \{0\}$, where $\lceil \frac{1}{x_k} \rceil$ denotes the smallest integer larger than $\frac{1}{x_k}$. This trial search point t_k belongs to the current mesh M_k since it is constructed through non-negative integer combinations of the direction matrix $D_k = D : (q_k - 1)b_k \in \mathbb{N}$ times the direction $d = \pi$ and $((q_k - 1)a_k + 1)) \in \mathbb{N}$ times the direction $d = -1$. Note that with these parameters, the algorithm is not a GPS instance since D cannot be written as a scalar multiple of a 1×2 integer matrix.

Theorem 3.1. *Let $\{x_k\}$ and $\{\Delta_k\}$ be the sequences of iterates and mesh size parameters generated by the above algorithm with the non-rational set of directions $D = [-1, \ +\pi]$. Then $\Delta_k = 1$ for all $k \in \mathbb{N}$ and $1 > x_0 > x_1 > x_2 > \cdots > 0$.*

Proof: Suppose that $\Delta_k = 1$ and that $x_k = b_k\pi - a_k > 0$ for some $a_k, b_k \in \mathbb{N} \setminus \{0\}$. Therefore, the trial SEARCH point t_k in Eq. (7) satisfies

$$0 \ = \ \frac{1}{x_k}x_k - 1 \ < \ t_k = \left\lceil \frac{1}{x_k} \right\rceil x_k - 1 \ < \ \left(\frac{1}{x_k} + 1\right)x_k - 1 \ = \ x_k.$$

It follows that the trial point belongs to the open interval $]0, x_k[$ and therefore $f(t_k) < f(x_k)$. Thus $x_{k+1} = t_k$ is an improved mesh point and $\Delta_{k+1} = \Delta_k = 1$. The proof follows by induction on k. □

Therefore, the sequence of iterates is monotone decreasing and belongs to the compact set $[0, x_0]$. The mesh size parameter remains constant at the value 1 since the SEARCH step always generates an improved mesh point. In conclusion, even for continuously differentiable functions (the function above is linear), rationality of the set of directions Z is necessary to ensure that for a fixed mesh size, the mesh is not a dense subset of $\mathbb{R}^n$.

3.2. Example B: The mesh update parameter τ

In GPS methods, the mesh update parameter $\tau > 1$ of Eqs. (4) and (5) must be rational. To show that this requirement cannot be relaxed, we assume in this section that τ is irrational, and we present a GPS instance that generates the following sequence of iterates:

$$\begin{aligned} u_0 &= 0, & u_{3j} &= u_{3j-1} - \lfloor u_{3j-1} \rfloor, \\ u_2 &= u_1 = \tau, & u_{3j+2} &= u_{3j+1} = u_{3j} + \tau \end{aligned} \tag{8}$$

for $j = 1, 2, \ldots$, and where $\lfloor u_{3j-1} \rfloor$ denotes the largest integer less than or equal to u_{3j-1}. One can easily show by induction that for $j = 0, 1, \ldots$

$$u_{3j} = j\tau - \lfloor j\tau \rfloor \in [0, 1[\quad \text{and} \quad u_{3j+2} = u_{3j+1} = (j+1)\tau - \lfloor j\tau \rfloor \in [\tau, \tau + 1[.$$

Also, since τ is irrational, the sequence has the properties that if i and j are natural numbers then $u_{3i} \neq u_{3j+1} = u_{3j+2}$ and if in addition $i \neq j$ then $u_{3i} \neq u_{3j}$ and $u_{3i+1} = u_{3i+2} \neq u_{3j+1} = u_{3j+2}$. We will show that the sequence of iterates has the property that the mesh size parameter Δ_k is bounded away from zero when the algorithm is applied to the following well-defined function $f : \mathbb{R} \to \mathbb{R}$:

$$f(x) = \begin{cases} 3^{-\ell} & \text{if } x = \ell\tau - \lfloor \ell\tau \rfloor \text{ for some } \ell \in \mathbb{N} \\ 2 \times 3^{-(\ell+1)} & \text{if } x = (\ell+1)\tau - \lfloor \ell\tau \rfloor \text{ for some } \ell \in \mathbb{N} \\ 1 & \text{otherwise.} \end{cases}$$

This function is discontinuous, so the part of the GPS convergence analysis requiring local Lipschitz continuity can not be applied. However, the part of the GPS analysis showing that the mesh size parameter gets infinitely small can be applied since the analysis is independent of the continuity of f.

Consider the parameters $x_0 = 0$, $\Delta_0 = \tau \notin \mathbb{Q}$, $G = 1$, $Z = D = [+1;\ -1]$, $w^- = -1$ and $w^+ = 1$. The mesh refining rule (4) is $\Delta_{k+1} = \tau^{-1}\Delta_k$ and define the mesh coarsening rule (5) as $\Delta_{k+1} = \Delta_k$ when k is a multiple of three, and as $\Delta_{k+1} = \tau\Delta_k$ otherwise. Define the SEARCH step as follows. When the iteration number k can be written as $3j + 2$ for some $j \in \mathbb{N}$, then the trial point $x_k - \lfloor x_k \rfloor \Delta_k$ is considered. When the POLL step is invoked, it evaluates f at $x_k + \Delta_k$, and if it did not produce an improved mesh point, it considers $x_k - \Delta_k$. These parameters do not define a GPS instance since τ is irrational.

Theorem 3.2. *Let $\{x_k\}$ be the sequence of iterates generated by the above algorithm with irrational mesh size update parameter $\tau > 1$. Then $x_k = u_k$ for all $k \in \mathbb{N}$ (where u_k is defined in Eq.* (8)*), and the sequence of mesh size parameters satisfies $\Delta_{3j} = \Delta_{3j+1} = \tau$ and $\Delta_{3j+2} = 1$ for $j \in \mathbb{N}$.*

Proof: Let us proceed by induction on j. We first show that the result holds when $k = 3j = 0$.

At iteration 0, $x_0 = 0 = u_0$, and $\Delta_0 = \tau$, and no SEARCH is performed. The first trial poll point improves the incumbent: $f(x_0+\Delta_0) = f(\tau) = \frac{2}{3} < f(x_0) = 1$. The iteration number $k = 0$ is a multiple of three, so $x_1 = \tau$ and $\Delta_1 = \tau$. No SEARCH is performed at iteration 1 and the POLL step shows that x_1 is a mesh local optimizer since $f(x_1+\Delta_1) = f(2\tau) = 1 > f(x_1)$ and $f(x_1 - \Delta_1) = f(0) = 1 > f(x_1)$. Therefore $x_2 = \tau$ and $\Delta_2 = 1$. The SEARCH step at iteration 2 improves the incumbent: $f(x_2 - \lfloor x_2 \rfloor \Delta_2) = f(\tau - \lfloor \tau \rfloor) = \frac{1}{3} < f(x_2)$. The iteration number $k = 2$ is not a multiple of three, so $x_3 = \tau - \lfloor \tau \rfloor = u_3$ and $\Delta_3 = \tau$.

Assuming that the result is true for $k \leq 3j$ for some $j \in \mathbb{N}$ we will show that the result is also true for $k = 3j + 1,\ 3j + 2$ and $3(j + 1)$.

– Iteration $k = 3j$: $x_{3j} = u_{3j} = j\tau - \lfloor j\tau \rfloor$, $\Delta_{3j} = \tau$ and $f(x_{3j}) = 3^{-j}$. No SEARCH is performed and the first trial poll point improves the incumbent: $f(x_k + \Delta_k) = f((j + 1)\tau - \lfloor j\tau \rfloor) = 2 \times 3^{-(j+1)} < f(x_k)$. The iteration number k is a multiple of three, so the mesh size parameter is kept constant.
– Iteration $k = 3j + 1$: $x_k = (j + 1)\tau - \lfloor j\tau \rfloor = u_k$, $\Delta_k = \tau$ and $f(x_k) = 2 \times 3^{-(j+1)}$. No SEARCH is performed and the POLL step shows that x_k is a mesh local optimizer since $f(x_k + \Delta_k) = f((j + 2)\tau - \lfloor j\tau \rfloor) = 1 > f(x_k)$ (indeed, the fact that τ is irrational implies that there are no $\ell \in \mathbb{N}$ such that $(j + 2)\tau - \lfloor j\tau \rfloor = \ell\tau - \lfloor \ell\tau \rfloor$ or $(j + 2)\tau - \lfloor j\tau \rfloor = (\ell + 1)\tau - \lfloor \ell\tau \rfloor$) and $f(x_k - \Delta_k) = f(x_{k-1}) = 3^{-j} > f(x_k)$. The mesh size parameter is reduced.
– Iteration $k = 3j + 2$: $x_k = (j + 1)\tau - \lfloor j\tau \rfloor = u_k$, $\Delta_k = 1$ and $f(x_k) = 2 \times 3^{-(j+1)}$. The SEARCH step improves the incumbent: $f(x_k - \lfloor x_k \rfloor \Delta_k) = f((j + 1)\tau - \lfloor (j + 1)\tau \rfloor) = 3^{-(j+1)} < f(x_k)$. The iteration number k is not a multiple of three, so the mesh size parameter becomes $\Delta_{3(j+1)} = \tau$ and iteration $3(j + 1)$ is initiated from $x_{3(j+1)} = (j + 1)\tau - \lfloor (j + 1)\tau \rfloor) = u_{3(j+1)}$.

$\square$

The first phase of the GPS convergence analysis uses the rationality of τ to show that there is a subsequence of mesh size parameters that converges to zero, and its proof is independent of the smoothness of f. The approach presented in Coope and Price (2001) is related to GPS, and allows trial points located off the mesh, but explicitly forces the mesh size parameter to converge to zero. By removing the GPS rationality requirement on τ the above example shows that a non-smooth function f can violate this result: the sequence of mesh size parameters is bounded away from zero since it takes only two strictly positive values. This suggests the following open question: *Is there an example with an irrational value of τ but with a continuous function f such that the mesh size parameter is bounded away from zero?*

4. Infinitely many limit points

In this section we present two examples of continuously differentiable functions in $\mathbb{R}^2$ for which a GPS instance produces infinitely many limit points. The differentiability of

f ensures that in addition to the convergence analysis of Audet and Dennis (2003), that of Lewis and Torczon (1996) and Torczon (1997) can also be applied.

In Example C, the gradient at a limit point is non-zero (and thus the convergence result (1) cannot be improved to (6)). Example D shows that the algorithm can still generate infinitely many limit points even under the stronger algorithmic requirements explicitly presented in the example.

4.1. *Example C: Non-zero gradient at a limit point*

Consider the function f defined as follows,

$$f(a,b) = \begin{cases} f_1(a,b) = -26a^3 - 32a^2b + 7|b|^3 & \text{if } a < 0 \\ f_2(a,b) = (7-8a^2)|b|^3 & \text{if } 0 \le a < \dfrac{1}{2} \\ f_3(a,b) = f_2(a,b) + 8\left(a - \dfrac{1}{2}\right)^2 (b^3 + b + a - 1) & \text{if } \dfrac{1}{2} \le a. \end{cases}$$

It can be shown that f is continuously differentiable everywhere on $\mathbb{R}^2$: For any $b \in [\frac{-1}{2}, \frac{1}{2}]$, the values and gradients at the transition points of f are

$$f(0,b) = f_1(0,b) = 7|b|^3, \quad f\left(\frac{1}{2}, b\right) = f_2\left(\frac{1}{2}, b\right) = 5|b|^3$$
$$\nabla f(0,b) = \begin{bmatrix} 0 \\ 21b|b| \end{bmatrix}, \quad \nabla f\left(\frac{1}{2}, b\right) = \nabla f_2\left(\frac{1}{2}, b\right).$$

The GPS parameters are the following: $x_0 = (a_0, b_0)^T = (-1, 1)$, $\Delta_0 = \frac{1}{4}$, $G = I$ (the identity matrix),

$$Z = D = \begin{bmatrix} 1 & -1 & 0 & 0 & 4 \\ 0 & 0 & 1 & -1 & -2 \end{bmatrix}, \quad \tau = 2, w^- = -1 \quad \text{and} \quad w^+ = 1.$$

The SEARCH step is empty. The POLL step stops as soon as an improved mesh point is detected and the mesh size parameter is doubled. The positive spanning set D_k used at iteration k with $x_k = (a_k, b_k)^T$ varies as follows (where d_j is the jth column of D)

$$D_k = \begin{cases} [d_5; d_2; d_3] & \text{if } a_k = -4\Delta_k \\ [d_1; d_2; d_3; d_4] & \text{if } -4\Delta_k \ne a_k < \dfrac{1}{2} \\ [d_2; d_1; d_3; d_4] & \text{if } a_k \ge \dfrac{1}{2}. \end{cases}$$

Figure 1(a) displays part of the domain of function f and plots the first few iterates. Figure 1(b) shows the shape of the function f when a varies from $-b$ to $1 - b$, and b is

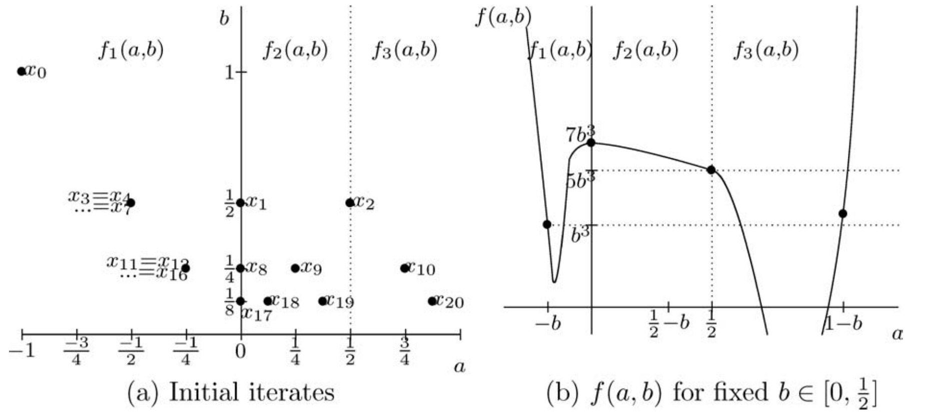

Figure 1. Initial iterates of GPS and objective function.

fixed to a value between 0 and $\frac{1}{2}$. We will show that a subsequence of iterates converges to $(1, 0)^T$ with $\nabla f(1, 0) \neq 0$. Table 1 details the search strategy of the first eight iterations. A new cycle is initiated at iteration 7. The trial points that improve the incumbent appear in shaded boxes.

The key structure of this example is that for any fixed value of b on $[0, \frac{1}{2}]$, the function value strictly decreases from $7b^3$ to $5b^3$ when a varies from 0 to $\frac{1}{2}$, as illustrated in Figure 1(b). From the iterate $(\frac{1}{2} - b, b)^T$, a mesh size parameter of $\frac{1}{2}$ yields the improved mesh point $(1 - b, b)^T$, at which the function value is $(1 + 8b)b^3 \in]b^3, 5b^3]$. From there, a mesh size parameter of 1 leads to $(-b, b)^T$ with function value b^3.

We define the ℓth cycle (for an $\ell \in \mathbb{N}$) as the iterations beginning at $k = \ell^2 + 6\ell$ and ending and including $k = (\ell + 1)^2 + 6(\ell + 1) - 1$. As shown below, a cycle starts at the iterate $x_k = (\frac{-1}{2^\ell}, \frac{1}{2^\ell})^T$ with $\Delta_k = \frac{1}{2^{\ell+2}}$ and function value $\frac{1}{8^\ell}$. The cycle goes through $\ell + 3$ successful iterations, increasing the mesh size parameter to 2, then, goes through $\ell + 4$ unsuccessful iterations leading to the start of cycle $\ell + 1$. For example, cycle 0 starts

Table 1. Initial eight iterations of GPS with $f(x_0) = 1$.

k	x_k	Δ_k	Poll			
0	$(-1, 1)$	$\frac{1}{4}$	$f_2(0, \frac{1}{2}) = \frac{7}{8}$			
1	$(0, \frac{1}{2})$	$\frac{1}{2}$	$f_3(\frac{1}{2}, \frac{1}{2}) = \frac{5}{8}$			
2	$(\frac{1}{2}, \frac{1}{2})$	1	$f_1(\frac{-1}{2}, \frac{1}{2}) = \frac{1}{8}$			
3	$(\frac{-1}{2}, \frac{1}{2})$	2	$f_3(\frac{3}{2}, \frac{1}{2}) = \frac{61}{8}$	$f_1(\frac{-5}{2}, \frac{1}{2}) = \frac{2457}{8}$	$f_1(\frac{-1}{2}, \frac{5}{2}) = \frac{741}{8}$	$f_1(\frac{-1}{2}, \frac{-3}{2}) = \frac{311}{8}$
4	x_3	1	$f_2(\frac{1}{2}, \frac{1}{2}) = \frac{5}{8}$	$f_1(\frac{-3}{2}, \frac{1}{2}) = \frac{421}{8}$	$f_1(\frac{-1}{2}, \frac{3}{2}) = \frac{119}{8}$	$f_1(\frac{-1}{2}, \frac{-1}{2}) = \frac{65}{8}$
5	x_3	$\frac{1}{2}$	$f_1(0, \frac{1}{2}) = \frac{7}{8}$	$f_1(-1, \frac{1}{2}) = \frac{87}{8}$	$f_1(\frac{-1}{2}, 1) = \frac{18}{8}$	$f_1(\frac{-1}{2}, 0) = \frac{26}{8}$
6	x_3	$\frac{1}{4}$	$f_1(\frac{-1}{4}, \frac{1}{2}) = \frac{9}{32}$	$f_1(\frac{-3}{4}, \frac{1}{2}) = \frac{91}{32}$	$f_1(\frac{-1}{2}, \frac{3}{4}) = \frac{13}{64}$	$f_1(\frac{-1}{2}, \frac{1}{4}) = \frac{87}{64}$
7	x_3	$\frac{1}{8}$	$f_2(0, \frac{1}{4}) = \frac{7}{8^2}$			

with $(a_0, b_0) = (-1, 1)$ and $\Delta_0 = \frac{1}{4}$. Three successful iterations go through the points $(0, \frac{1}{2})^T, (\frac{1}{2}, \frac{1}{2})^T$ and finally $(\frac{-1}{2}, \frac{1}{2})^T$, where $\Delta_3 = 2$. Then four unsuccessful iterations reduce the mesh size parameter to $\Delta_7 = \frac{1}{8}$. This initiates cycle 1 at $k = 7$.

We show that the sequence of iterates possesses an infinite number of limit points, namely all points of the set $\{(\frac{1}{2^\ell}, 0)^T : \ell = 0, 1, \ldots\} \cup \{(0, 0)^T\}$. Moreover, the gradient norm of all these points is zero, except for the point $(1, 0)^T$.

Theorem 4.1. *For any integer $\ell \geq 0$, the iterates and mesh size parameters of cycle ℓ generated by the above instance of GPS are*

$$(a_k, b_k), \Delta_k = \begin{cases} \left(\dfrac{-1}{2^\ell}, \dfrac{1}{2^\ell}\right), & \dfrac{1}{2^{\ell+2}} \quad \text{if } k = \ell^2 + 6\ell \\ \left(\dfrac{2^{j-1}-1}{2^{\ell+1}}, \dfrac{1}{2^{\ell+1}}\right) & \dfrac{1}{2^{\ell-j+2}} \quad \text{if } k = \ell^2 + 6\ell + j, \\ & \qquad \text{with } j \in \{1, 2, \ldots, 2+\ell\}, \\ \left(\dfrac{-1}{2^{\ell+1}}, \dfrac{1}{2^{\ell+1}}\right), & \dfrac{1}{2^{j-2}} \quad \text{if } k = \ell^2 + 7\ell + 2 + j \\ & \qquad \text{with } j \in \{1, 2, \ldots, 4+\ell\}. \end{cases}$$

Proof: The proof is done by induction on ℓ. We already verified in Table 1 that the result holds for the cycle $\ell = 0$, that is, from iterations 0 to 6.

Suppose that cycle ℓ is initiated at iterate $x_k = (\frac{-1}{2^\ell}, \frac{1}{2^\ell})^T$ where $k = \ell^2 + 6\ell$, and that $\Delta_k = \frac{1}{2^{\ell+2}}$. The objective function value is $f(x_k) = f_1(\frac{-1}{2^\ell}, \frac{1}{2^\ell}) = \frac{1}{8^\ell}$. The top half of Table 2 details all iterations of cycle ℓ. The function values for the iterations $k = \ell^2 + 7\ell + 2 + j$ for $1 \leq j \leq 4 + \ell$ appear in the bottom half of the table. The objective function value of all four trial points (labelled A_+, A_-, B_+ and B_-) are greater than the incumbent value $f(x_k) = f_1(-u, u) = u^3$, where $u = \frac{1}{2^{\ell+1}}$, and therefore all these iterations are unsuccessful in improving the incumbent solution.

The iteration number following the last iteration presented in Table 2 is $k = (\ell^2 + 7\ell) + 2 + (4 + \ell) + 1 = (\ell + 1)^2 + 6(\ell + 1)$. The incumbent solution is not improved, therefore, cycle $\ell + 1$ is initiated at $x_k = (\frac{-1}{2^{\ell+1}}, \frac{1}{2^{\ell+1}})^T$ with mesh size parameter $\Delta_k = \frac{1}{2^{\ell+3}}$. This completes the proof. □

The previous theorem details all iterates generated by the algorithm. We now discuss some properties of the limit points of the sequence of iterates. All iterates and trial points generated by the algorithm belong to a compact set. Consider the sequence of iterations $k = \ell^2 + 7\ell + 2$ for $\ell \geq 0$: all corresponding iterates $x_k = (1 - \frac{1}{2^{\ell+1}}, \frac{1}{2^{\ell+1}})^T$ are successful and converge to the point $(1, 0)^T$ at which the gradient is $\nabla f(1, 0) = \nabla f_3(1, 0) = (2, 2)^T$. All other limit points are of the form $(\frac{1}{2^\ell}, 0)^T$ for $\ell > 0$, and they have zero gradient. Moreover, only the sequence of iterates that converges to $(0, 0)^T$ contains mesh local optimizers.

In summary, even if the objective function is continuously differentiable everywhere, there can be a subsequence of iterates converging to a solution whose gradient is non-zero. In the example presented above, that point is the limit of iterates that were successful in finding improved mesh points with mesh size parameters all equal to one. The large mesh

Table 2. Objective function values of cycle ℓ.

k	x_k Δ_k	Poll
$\ell^2+6\ell$	$(\frac{-1}{2^\ell},\frac{1}{2^\ell})$ $\frac{1}{2^{\ell+2}}$	$f(x_k+(4,-2)\Delta_k)=f_2(0,\frac{1}{2^{\ell+1}})=\frac{7}{8^{\ell+1}}$
$\ell^2+6\ell+j$ $1\le j\le\ell$	$(\frac{2^{j-1}-1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})$ $\frac{1}{2^{\ell-j+2}}$	$f(x_k+(1,0)\Delta_k)=f_2(\frac{2^j-1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})=\frac{7-8(\frac{2^j-1}{2^{\ell+1}})^2}{8^{\ell+1}}$
$\ell^2+7\ell+1$	$(\frac12-\frac{1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})$ $\frac12$	$f(x_k+(1,0)\Delta_k)=f_3(1-\frac{1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})=\frac{1+\frac{1}{2^{\ell-2}}}{8^{\ell+1}}$
$\ell^2+7\ell+2$	$(1-\frac{1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})$ 1	$f(x_k+(-1,0)\Delta_k)=f_1(\frac{-1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})=\frac{1}{8^{\ell+1}}$
$\ell^2+7\ell+2+j$ $1\le j\le 4+\ell$	$(\frac{-1}{2^{\ell+1}},\frac{1}{2^{\ell+1}})$ $\frac{1}{2^{j-2}}$	$A_\pm:\ f(x_k\pm(1,0)\Delta_k)=f(\frac{-1}{2^{\ell+1}}\pm\frac{1}{2^{j-2}},\frac{1}{2^{\ell+1}})$ $B_\pm:\ f(x_k\pm(0,1)\Delta_k)=f_1(\frac{-1}{2^{\ell+1}},\frac{1}{2^{\ell+1}}\pm\frac{1}{2^{j-2}})$

$A_+: f(x_k+(1,0)\Delta_k)$	$=\ f(\frac{1}{2^{j-2}}-u,u)$ where $u=\frac{1}{2^{\ell+1}}$
$j=1$	$f_3(2-u,u)=18-24u+8u^2-7u^3+8u^4$
$j=2$	$f_3(1-u,u)=u^3+8u^4$
$3\le j\le 3+\ell$	$a=\frac{1}{2^{j-2}}-u\in[0,\frac12]\Rightarrow f_2(a,u)\in[5u^3,7u^3]$ (see Figure 1(b))
$j=4+\ell$	$f_1(\frac{-u}{2},u)=\frac94u^3$
$A_-:\quad f(x_k-(1,0)\Delta_k)$	$=\ f(\frac{-1}{2^{j-2}}-u,u)$ where $u=\frac{1}{2^{\ell+1}}$
$1\le j\le 4+\ell$	$a=-u-\frac{1}{2^{j-2}}\le\frac{-3u}{2}$ $\Rightarrow f_1(a,u)=-26a^3-32a^2u+7u^3$ $\ge\frac{26\times3}{2}a^2u-32a^2u+7u^3=7(a^2u+u^3)$
$B_+: f(x_k+(0,1)\Delta_k)$	$=\ f(-u,u+\frac{1}{2^{j-2}})$ where $u=\frac{1}{2^{\ell+1}}$
$1\le j\le 2+\ell$	$b=u+\frac{1}{2^{j-2}}\ge u+\frac{1}{2^\ell}=3u$ $\Rightarrow f_1(-u,b)=26u^3-32u^2b+7b^3$ $\ge 26u^3-32u^2b+63u^2b=26u^3+31u^2b$
$j=3+\ell$	$f_1(-u,2u)=18u^3$
$j=4+\ell$	$f_1(-u,\frac{3u}{2})=\frac{13}{8}u^3$
$B_-: f(x_k-(0,1)\Delta_k)$	$=\ f(-u,u-\frac{1}{2^{j-2}})$ where $u=\frac{1}{2^{\ell+1}}$
$1\le j\le 3+\ell$	$b=u-\frac{1}{2^{j-2}}\le u-\frac{1}{2^{\ell+1}}=0$ $\Rightarrow f_1(-u,b)=26u^3-32u^2b-7b^3\ge 26u^3$
$j=4+\ell$	$f_1(-u,\frac{u}{2})=\frac{87}{8}u^3$

size parameter allowed the POLL step to move away from the region containing the non-stationary point $(1,0)^T$. Of course, as predicted by the theory (Audet and Dennis, 2003), any limit point of a refining subsequence (i.e., subsequences of mesh local optimizers located on meshes that get infinitely fine) has a zero gradient limit point. In the example there is only one such limit point: the origin. Moreover, since the function is lower-semi continuous at all limit points, they all share the same objective function value (Audet and Dennis, 2003). Observe that the limit point $(1, 0)^T$ where the gradient is non-zero is the more interesting one since any local exploration around this point would identify a descent direction.

4.2. *Example D: Infinitely many limit points when under strong requirements*

The example presented in this section shows that even under the stronger requirements presented in Torczon (1997) for continuously differentiable functions (to strengthen the convergence result from (1) to (6)), there still can be infinitely many limit points produced by a GPS algorithm. These additional requirements are

(i) Any trial point $x_k + \Delta_k Dz$ for some $z \in \mathbb{N}^m$ has to be such that z is bounded in norm.
(ii) $\lim_{k\to\infty} \Delta_k = 0$.
(iii) $f(x_{k+1}) \leq f(x_k + \Delta_k d)$ for all $d \in D_k$.

Requirement (i) ensures that the distance from the trial points visited by SEARCH step to the incumbent solution x_k is bounded above by a multiple of the mesh size parameter. And so, as k increases, requirement (ii) ensures that the SEARCH is conducted locally around x_k. Requirement (iii) ensures that the iteration does not stop at the first trial point that deceases the objective function value f. All other examples in this document violate at least one of these requirements.

Consider the convex continuously differentiable function in $\mathbb{R}^2$:

$$f(a, b) = \begin{cases} (a^2+1)b^2 & \text{if } b \geq 0 \\ ((1-a)^2+1)b^2 & \text{if } b < 0. \end{cases}$$

For any $b > 0$ the function monotonically decreases when a varies from 1 to 0, and for any $b < 0$ the function monotonically decreases when a varies from 0 to 1. It achieves its minimum value 0 whenever $b = 0$.

The GPS parameters are $x_0^T = (0, \frac{2}{3})$, $\Delta_0 = 1$, $G = I$ (the identity matrix), $Z = D = [I; -I]$, $\tau = 2$, $w^- = -1$ and $w^+ = 0$. The SEARCH is empty, therefore the requirement (i) is satisfied. Requirement (ii) is also satisfied since all level sets of f are bounded and the mesh size parameter is never increased (Torczon, 1997). All four trial points are evaluated in the POLL step. When there is improvement, the best one is chosen to be the next iterate; therefore requirement (iii) is respected.

Figure 2 and Table 3 detail the first iterates generated by the algorithm. The trial points that improve the incumbent appear in shaded boxes in the table. We will show that there are iterates that go infinitely close to every point on the line segment from the origin to $(1, 0)^T$.

We define an even cycle to be the successful iterations starting at $(0, b)^T$ where $b > 0$, then going to $(0, \frac{-b}{2})^T$ and ending at the unsuccessful one at $(1, \frac{-b}{2})^T$. The following odd cycle is composed of the successful iterations starting at $(1, \frac{-b}{2})^T$, then going to $(1, \frac{b}{4})^T$ and ending at the unsuccessful one at $(0, \frac{b}{4})^T$. We now explicitly write the values of all iterates.

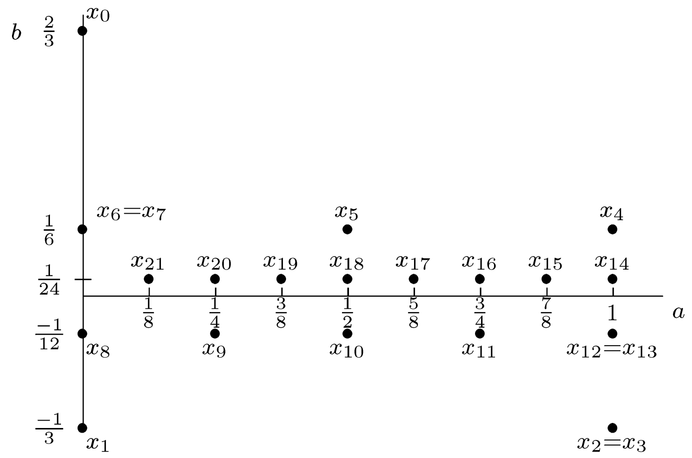

Figure 2. Initial iterates of GPS.

Theorem 4.2. *For any integer $\ell \geq 0$, the iterates of cycle ℓ are*

$$
\begin{aligned}
x_{2^\ell+2\ell-1} &= \left(0, \frac{2\Delta_k}{3}\right), \quad \textit{when } k = 2^\ell + 2\ell - 1,\\
x_{2^\ell+2\ell+j} &= \left(j\Delta_k, \frac{-\Delta_k}{3}\right), \quad j \in \{0, 1, \ldots, 2^\ell\} \quad \textit{and} \quad \ell \textit{ is even},\\
x_{2^\ell+2\ell-1} &= \left(1, \frac{-2\Delta_k}{3}\right), \quad \textit{when } k = 2^\ell + 2\ell - 1,\\
x_{2^\ell+2\ell+j} &= \left(1 - j\Delta_k, \frac{-\Delta_k}{3}\right), \quad j \in \{0, 1, \ldots, 2^\ell\} \quad \textit{and} \quad \ell \textit{ is odd}
\end{aligned}
$$

and the mesh size parameters are all equal to $\Delta_k = \frac{1}{2^\ell}$.

Table 3. Initial eight iterations of GPS with $f(x_0) = \frac{4}{9}$.

Cycle	k	x_k	Δ_k	$f(a_k + \Delta_k, b_k)$	$f(a_k, b_k + \Delta_k)$	$f(a_k - \Delta_k, b_k)$	$f(a_k, b_k - \Delta_k)$
0	0	$(0, \frac{2}{3})$	1	$f(1, \frac{2}{3}) = \frac{8}{3^2}$	$f(0, \frac{5}{3}) = \frac{25}{3^2}$	$f(-1, \frac{2}{3}) = \frac{8}{3^2}$	$f(0, \frac{-1}{3}) = \frac{2}{3^2}$
	1	$(0, \frac{-1}{3})$	1	$f(1, \frac{-1}{3}) = \frac{1}{3^2}$	$f(0, \frac{2}{3}) = \frac{4}{3^2}$	$f(-1, \frac{-1}{3}) = \frac{5}{3^2}$	$f(0, \frac{-4}{3}) = \frac{32}{3^2}$
	2	$(1, \frac{-1}{3})$	1	$f(2, \frac{-1}{3}) = \frac{2}{3^2}$	$f(1, \frac{2}{3}) = \frac{8}{3^2}$	$f(0, \frac{-1}{3}) = \frac{2}{3^2}$	$f(1, \frac{-4}{3}) = \frac{16}{3^2}$
1	3	x_2	$\frac{1}{2}$	$f(\frac{3}{2}, \frac{-1}{3}) = \frac{5}{6^2}$	$f(1, \frac{1}{6}) = \frac{2}{6^2}$	$f(\frac{1}{2}, \frac{-1}{3}) = \frac{5}{6^2}$	$f(1, \frac{-4}{6}) = \frac{4}{3^2}$
	4	$(1, \frac{1}{6})$	$\frac{1}{2}$	$f(\frac{3}{2}, \frac{1}{6}) = \frac{13/4}{6^2}$	$f(1, \frac{4}{6}) = \frac{32}{6^2}$	$f(\frac{1}{2}, \frac{1}{6}) = \frac{5/4}{6^2}$	$f(1, \frac{-1}{3}) = \frac{4}{6^2}$
	5	$(\frac{1}{2}, \frac{1}{6})$	$\frac{1}{2}$	$f(1, \frac{1}{6}) = \frac{2}{6^2}$	$f(\frac{1}{2}, \frac{2}{3}) = \frac{20}{6^2}$	$f(0, \frac{1}{6}) = \frac{1}{6^2}$	$f(\frac{1}{2}, \frac{-1}{3}) = \frac{5}{6^2}$
	6	$(0, \frac{1}{6})$	$\frac{1}{2}$	$f(\frac{1}{2}, \frac{1}{6}) = \frac{5/4}{6^2}$	$f(0, \frac{2}{3}) = \frac{16}{6^2}$	$f(\frac{-1}{2}, \frac{1}{6}) = \frac{5/4}{6^2}$	$f(0, \frac{-1}{3}) = \frac{8}{6^2}$
2	7	x_6	$\frac{1}{4}$	$f(\frac{1}{4}, \frac{1}{6}) = \frac{17/4}{12^2}$	$f(0, \frac{5}{12}) = \frac{25}{12^2}$	$f(\frac{-1}{4}, \frac{1}{6}) = \frac{17/4}{12^2}$	$f(0, \frac{-1}{12}) = \frac{2}{12^2}$

Table 4. Objective function values of the trial points of the even cycle ℓ.

k	x_k	$f(a_k + \Delta_k, b_k)$ $f(a_k - \Delta_k, b_k)$	$f(a_k, b_k + \Delta_k)$ $f(a_k, b_k - \Delta_k)$
$2^\ell + 2\ell - 1$	$(0, \frac{2\Delta_k}{3})$	$(\Delta_k^2 + 1)\frac{4\Delta_k^2}{9}$	$\frac{25\Delta_k^2}{9}$
		$(\Delta_k^2 + 1)\frac{4\Delta_k^2}{9}$	$\boxed{\frac{2\Delta_k^2}{9}}$
$2^\ell + 2\ell + j$	$(j\Delta_k, \frac{-\Delta_k}{3})$	$\boxed{((1 - (j+1)\Delta_k)^2 + 1)\frac{\Delta_k^2}{9}}$	$(j^2\Delta_k^2 + 1)\frac{4\Delta_k^2}{9}$
$0 \le j \le 2^\ell - 1$		$((1 - (j-1)\Delta_k)^2 + 1)\frac{\Delta_k^2}{9}$	$((1 - j\Delta_k)^2 + 1)\frac{16\Delta_k^2}{9}$
$2^{\ell+1} + 2\ell$	$(1, \frac{-\Delta_k}{3})$	$(\Delta_k^2 + 1)\frac{\Delta_k^2}{9}$ $(\Delta_k^2 + 1)\frac{\Delta_k^2}{9}$	$\frac{8\Delta_k^2}{9}$ $\frac{16\Delta_k^2}{9}$

Proof: The proof is done by induction on ℓ. We already verified in Table 3 that the result holds for the even cycle $\ell = 0$, and for the odd cycle $\ell = 1$.

Suppose that the even cycle ℓ is initiated with $\Delta_k = \frac{1}{2^\ell}$ and $x_k = (0, \frac{2\Delta_k}{3})^T$ for $k = 2^\ell + 2\ell - 1$. The current objective function value is $f(x_k) = \frac{4\Delta_k^2}{9}$. Table 4 details the objective function values at the poll points for the even cycle ℓ. The trial points that improve the incumbent appear in shaded boxes. All iterations but the last improve the incumbent, therefore the mesh size parameter remains constant through the cycle.

The odd cycle $\ell + 1$ starts at iteration $k' = 2^{(\ell+1)} + 2(\ell + 1) - 1$. The last iteration appearing in Table 4 does not improve the incumbent, and therefore $\Delta_{k'} = \frac{\Delta_k}{2} = \frac{1}{2^{\ell+1}}$ and $x_{k'} = (1, \frac{-2\Delta_{k'}}{3})^T$. We have shown that the result is true for even values of ℓ. The proof for odd cycles is similar, and is omitted. $\Box$

This example shows that even under stronger restrictions there still can be infinitely many limit points of the sequence of iterates. Every point of the of the line segment whose endpoints are $(0, 0)^T$ and $(1, 0)^T$ is a limit point and the gradient at these points is zero (Torczon, 1997). The function is lower-semi continuous at these points, and so they all share the same objective function value (Audet and Dennis, 2003).

This example, and its conclusion that infinitely many limit points can be generated, also applies to the framework of Coope and Price (2001) since the mesh size parameter converges to zero.

5. Single limit point with non-zero gradient

The examples presented in this section apply to the convergence analysis of Audet and Dennis (2003) for functions that are not continuously differentiable. They show that the following result is tight.

Proposition 5.1. *Let $f : \mathbb{R}^1 \to \mathbb{R}$ be Lipschitz near a limit point $\hat{x}$ of a refining subsequence produced by a GPS algorithm. Then $0 \in \partial f(\hat{x})$.*

Proof: By a result in Audet and Dennis (2003), $f^\circ(\hat{x}; d_1) \geq 0 \leq f^\circ(\hat{x}; d_2)$ for some directions $d_1 < 0$ and $d_2 > 0$. Moreover, since $f^\circ(\hat{x}; \lambda v) = \lambda f^\circ(\hat{x}; v)$ for any $\lambda \geq 0$ and $v \in \mathbb{R}^1$, it follows that $f^\circ(\hat{x}; v) \geq 0$ for any direction $v \in \mathbb{R}^1$ and the generalized gradient of f at $\hat{x}$

$$\partial f(\hat{x}) = \{\zeta \in \mathbb{R}^1 : f^\circ(\hat{x}; v) \geq v\zeta \text{ for all } v \in \mathbb{R}^1\}$$

contains zero. □

Both examples have differentiable Lipschitz functions f that are not strictly differentiable and produce a single limit point $\hat{x}$. In both cases the gradient exists at $\hat{x}$ and is non-zero. Example 5.1 is in $\mathbb{R}^1$ and therefore the Clarke derivatives are non-negative for all directions (the above proposition guarantees that zero belongs to the generalized gradient). The last example shows that in higher dimensions, the generalized gradient at $\hat{x}$ may not contain zero.

5.1. Example E: Single limit point with non-zero gradient in $\mathbb{R}^1$

The example presented in this section applies GPS to a Lipschitz function in $\mathbb{R}^1$ which is not strictly differentiable (and thus not continuously differentiable) but is continuous and differentiable everywhere. The sequence of iterates generated by the algorithm possesses a single limit point. This limit point is a Clarke stationary point where the gradient exists and is non-zero.

First, let us define for $\alpha = 2^{-\ell}$, $\ell \in \{2, 3, \ldots\}$ the cubic functions $p_\alpha : [\alpha - \frac{\alpha^2}{2}, \alpha[\to \mathbb{R}$ and $q_\alpha : [\alpha, 2\alpha - 2\alpha^2[\to \mathbb{R}$ as

$$\begin{aligned} p_\alpha(x) &= \frac{8}{\alpha^4}x^3 + \frac{6(\alpha - 4)}{\alpha^3}x^2 + \frac{12(2 - \alpha)}{\alpha^2}x + \frac{(\alpha - 2)(4 - \alpha - \alpha^2)}{\alpha} \\ q_\alpha(x) &= \frac{\alpha - 1}{(2\alpha - 1)^3}\left[\frac{-2}{\alpha^2}x^3 + \frac{3(3 - 2\alpha)}{\alpha}x^2 + 12(\alpha - 1)x \right. \\ &\quad \left. - 2\alpha(4\alpha^3 - 6\alpha^2 + 6\alpha - 3)\right] \end{aligned}$$

Consider the piecewise function in R^1 defined as:

$$f(x) = \begin{cases} x & \text{if } x \leq 0, \\ p_\alpha(x) & \text{if } \alpha - \frac{\alpha^2}{2} \leq x < \alpha, \qquad \text{where } \alpha = 2^{-\ell} \\ q_\alpha(x) & \text{if } \alpha \leq x < 2\alpha - 2\alpha^2, \quad \text{for some } \ell \in \{2, 3, \ldots\}. \\ \dfrac{3}{8} & \text{if } x \geq \dfrac{3}{8}, \end{cases}$$

The plot of the function f appears on the top part of Figure 3 and its derivative on the bottom. Using elementary calculus, one can verify that f is continuous, bounded above by

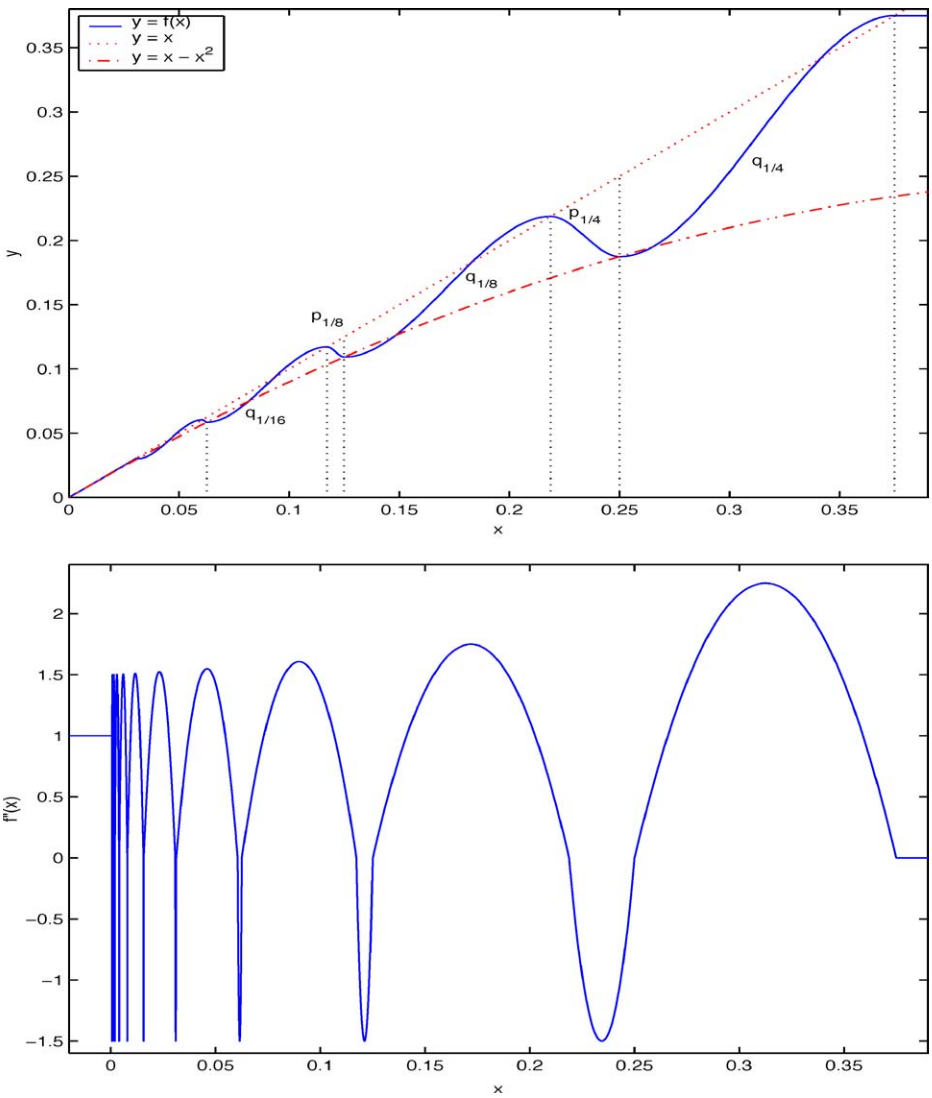

Figure 3. A continuous function that is not strictly differentiable at the origin.

$x + x^2$, bounded below by $x - 2x^2$, Lipschitz and differentiable everywhere, and

$$1 = \lim_{t \downarrow 0} \frac{t - 2t^2}{t} \le f'(0; 1) = \lim_{t \downarrow 0} \frac{f(t)}{t} \le \lim_{t \downarrow 0} \frac{t + t^2}{t} = 1,$$

therefore the gradient at the origin is $\nabla f(0) = 1$. One can also show that for any $\ell \in \{2, 3, \ldots\}$, the point $x = 2^{-\ell}$ is a local minimizer of f with $f(x) = q_x(x) = x - x^2$ (this is in fact the key to this example, the iterates will move from one local minimum to another and will converge to the origin).

The following GPS parameters are used: $x_0 = 2^{-2}$, $\Delta_0 = 2^{-5}$, $G = 1$, $Z = D = [-1,\ +1]$, $\tau = 2^2$, $w^- = -1$ and $w^+ = 0$. When the iteration number k is odd, the SEARCH step considers the mesh trial point $x_k + z_k\Delta_k$ where z_k is the integer part of $\frac{-4}{x_k}$. The SEARCH is empty when k is even.

Polling around the local minimum x_0 forces the algorithm to reduce the mesh size parameter since the value $f(x_0) = \frac{3}{16}$ is less than both values

$$
\begin{aligned}
f(x_0 - \Delta_0) &= p_{x_0}\left(x_0 - \frac{x_0^2}{2}\right) = p_{1/4}\left(\frac{7}{32}\right) = \frac{7}{32} \ > \ f(x_0),\\
f(x_0 + \Delta_0) &= q_{x_0}\left(x_0 + \frac{x_0^2}{2}\right) = q_{1/4}\left(\frac{9}{32}\right) = \frac{111}{512} > f(x_0).
\end{aligned}
$$

Therefore $x_1 = x_0$ and $\Delta_1 = \tau^{-1}\Delta_0 = 2^{-7}$. The other iterates are described by the following result.

Theorem 5.2. *For any integer $\ell \in \mathbb{N}$, the iterates and mesh size parameters generated by the above instance of GPS are $x_{2\ell} = x_{2\ell+1} = 2^{-(\ell+2)}$, $\Delta_{2\ell} = 2^{-(2\ell+5)}$ and $\Delta_{2\ell+1} = 2^{-(2\ell+7)}$.*

Proof: The proof is by induction on ℓ. We already showed the result for $\ell = 0$.

Suppose that the odd iteration $k = 2(\ell - 1) + 1$ is initiated with $x_k = 2^{-(\ell+1)}$, $\Delta_k = 2^{-(2\ell+5)}$ and $f(x_k) = 2^{-(\ell+1)} - 2^{-(2\ell+2)}$ for some $\ell - 1 \geq 0$. The SEARCH step considers the trial point $t = x_k + z_k\Delta_k$ with $z_k = \frac{-4}{2^{-(\ell+1)}} = -2^{\ell+3}$. The trial point is therefore $t = 2^{-(\ell+1)} - 2^{\ell+3}2^{-(2\ell+5)} = 2^{-(\ell+2)}$ and its function value is $f(t) = 2^{-(\ell+2)} - 2^{-(2\ell+4)} < f(x_k)$. Therefore $x_{k+1} = x_{2\ell} = t$ is an improved mesh point and $\Delta_{k+1} = \Delta_{2\ell} = \Delta_k$.

Now, at the even iteration $k' = k + 1 = 2\ell$ no SEARCH is performed, and the POLL step identifies a mesh local optimizer since

$$
\begin{aligned}
f(x_{k'} - \Delta_{k'}) &= f\left(x_{k'} - \frac{x_{k'}^2}{2}\right) \ = \ p_{x_{k'}}\left(x_{k'} - \frac{x_{k'}^2}{2}\right)\\
&= x_{k'} - \frac{x_{k'}^2}{2} \ > \ x_{k'} - x_{k'}^2 \ = \ q_{x_{k'}}(x_{k'}) \ = \ f(x_{k'})\\
f(x_{k'} + \Delta_{k'}) &= f\left(x_{k'} + \frac{x_{k'}^2}{2}\right) \ = \ q_{x_{k'}}\left(x_{k'} + \frac{x_{k'}^2}{2}\right)\\
&= \frac{\left(x_{k'} - x_{k'}^2\right)\left(39x_{k'}^3 - 51x_{k'}^2 + 24x_{k'} - 4\right)}{4(2x_{k'} - 1)^3}\\
&> x_{k'} - x_{k'}^2 \ = \ q_{x_{k'}}(x_{k'}) \ = \ f(x_{k'}).
\end{aligned}
$$

Therefore $x_{2\ell+1} = x_{2\ell} = 2^{-(\ell+2)}$ and $\Delta_{2\ell+1} = \frac{\Delta_{2\ell}}{4} = 2^{-(2\ell+7)}$. $\square$

The previous result shows that the sequence of iterates converges to the origin where the function is Lipschitz but not continuously differentiable. Proposition 5.1 is tight in $\mathbb{R}^1$ in the sense that even if Clarke derivatives are non-negative in every direction and if zero belongs to the generalized gradient (this can also be seen by looking at the sequence of mesh local optimizers in Figure 3), then it is possible that the gradient exists but is non-zero.

5.2. *Example F: Single limit point with zero excluded from the generalized gradient*

When $n = 1$ and f is Lipschitz at a limit point $\hat{x}$ of a refining sequence, Proposition 5.1 ensures that zero is contained in the generalized gradient. The previous example showed that even if the gradient exists, it might be non-zero. We now present an example with $n = 2$, where zero is not in the generalized gradient but where the gradient exists at the limit of the sequence of iterates. We will need the following lemma.

Lemma 5.3. *If there exists an $\epsilon > 0$ such that the functions ℓ, f and u satisfy*

$$\ell(\hat{x}) = f(\hat{x}) = u(\hat{x}) \quad \textit{and} \quad \ell(x) \le f(x) \le u(x) \quad \forall\, x \in B_\epsilon(\hat{x})$$

where $B_\epsilon(\hat{x})$ is a ball of radius ϵ centered at $\hat{x} \in \mathbb{R}^n$, and if both ℓ and u are differentiable at $\hat{x}$, and if $\nabla \ell(\hat{x}) = \nabla u(\hat{x})$ then f is differentiable at $\hat{x}$ and $\nabla f(\hat{x}) = \nabla \ell(\hat{x})$.

Proof: By definition of differentiability,

$$\begin{aligned} 0 = 7 \liminf_{h\to 0} \tfrac{\ell(\hat{x}+h)-\ell(\hat{x})-h^T\nabla\ell(\hat{x})}{\|h\|} &\le \liminf_{h\to 0} \tfrac{f(\hat{x}+h)-f(\hat{x})-h^T\nabla\ell(\hat{x})}{\|h\|} \\ \limsup_{h\to 0} \tfrac{f(\hat{x}+h)-f(\hat{x})-h^T\nabla u(\hat{x})}{\|h\|} &\le \limsup_{h\to 0} \tfrac{u(\hat{x}+h)-u(\hat{x})-h^T\nabla u(\hat{x})}{\|h\|} \;=\; 0, \end{aligned}$$

and therefore, $\lim_{h\to 0} \frac{f(\hat{x}+h)-f(\hat{x})-h^T\nabla\ell(\hat{x})}{\|h\|}$ exists and equals zero. The result follows by the definition and uniqueness of the gradient (Rockafellar, 1970). $\square$

Consider the function f defined as follows,

$$f(a,b) = \begin{cases} 2b(1-b^2) & \text{if } 0 < b(1-2b^2) \le a \le b \\ 2a(1-a^2) & \text{if } 0 < a(1-2a^2) \le b < a \\ a+b & \text{otherwise.} \end{cases}$$

f is continuous and Lipschitz everywhere on $\{(a,b)^T : a < \frac{1}{\sqrt{2}}, b < \frac{1}{\sqrt{2}}\}$. The plot of the function is displayed on the left of Figure 4 and the piecewise linear level sets on the right.

We will show that every iterate lies on the diagonal $b = a$ and that they converge to the origin, which is not a Clarke stationary point.

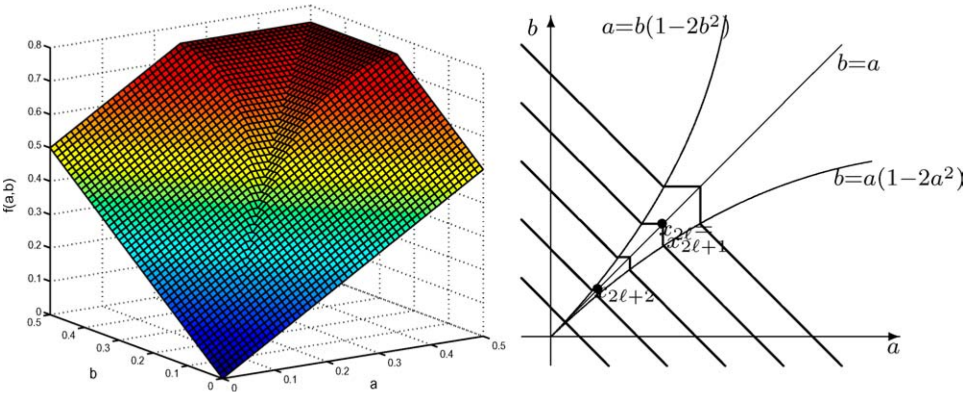

Figure 4. Objective function landscape, and level sets.

The GPS parameters are $x_0^T = (\frac{1}{2}, \frac{1}{2})$, $\Delta_0 = \frac{1}{8}$, $G = I$ (the identity matrix),

$$Z = D = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}, \tau = 8, w^- = -1 \quad \text{and} \quad w^+ = 0.$$

When the iteration number k is odd, the SEARCH step considers the mesh trial point $x_k + 2^{k+3}\Delta_k(-1, -1)^T$. There is no SEARCH when k is even.

The next result shows that the entire sequence of iterates converges to the origin.

Theorem 5.4. *For any integer $\ell \in \mathbb{N}$, the iterates and mesh size parameters generated by the above instance of GPS are $x_{2\ell} = x_{2\ell+1} = (2^{-(\ell+1)}, 2^{-(\ell+1)})^T$, $\Delta_{2\ell} = 8^{-(\ell+1)}$ and $\Delta_{2\ell+1} = 8^{-(\ell+2)}$.*

Proof: The proof is by induction on ℓ. When $\ell = 0$, polling around x_0 yields a mesh local optimizer: $f(\frac{3}{8}, \frac{1}{2}) = f(\frac{1}{2}, \frac{3}{8}) = \frac{3}{4} = f(x_0)$ and $f(\frac{5}{8}, \frac{5}{8}) = \frac{195}{256} > f(x_0)$. So $x_1 = x_0$ and $\Delta_1 = \frac{1}{64}$.

Suppose that the result is true up to $k = 2\ell$, and therefore $x_{2\ell} = (2^{-(\ell+1)}, 2^{-(\ell+1)})$ and $\Delta_{2\ell} = 8^{-(\ell+1)}$. Then polling around x_k yields a mesh local optimizer:

$$\begin{aligned} f\big(2^{-(\ell+1)} - 8^{-(\ell+1)}, 2^{-(\ell+1)}\big) &= f\big(2^{-(\ell+1)}, 2^{-(\ell+1)} - 8^{-(\ell+1)}\big) = 2^{-\ell} - 2^{-(3\ell+2)} \\ &= f(x_k) < f\big(2^{-(\ell+1)} + 8^{-(\ell+1)}, 2^{-(\ell+1)} + 8^{-(\ell+1)}\big). \end{aligned}$$

So $x_{2\ell+1} = x_{2\ell}$ and $\Delta_{2\ell+1} = 8^{-(\ell+2)}$.

Iteration $k = 2\ell + 1$ is odd, and so the SEARCH trial point is $t = (2^{-(\ell+2)}, 2^{-(\ell+2)})^T$ with $f(t) < f(x_k)$. So $x_{2\ell+2} = t$ and $\Delta_{2\ell+2} = \Delta_{2\ell+1}$. □

The previous result shows that the entire sequence of iterates converges to the origin.

Proposition 5.5. *The generalized gradient $\partial f(0)$ does not contain zero, and $\nabla f(0)$ exists.*

Proof: The generalized gradient $\partial f(0)$ is the convex hull of the set $\{(2,0)^T, (0,2)^T, (1,1)^T\}$. It clearly does not contain zero.

Define the functions $\ell(a,b) = (a+b)(1-ab)$ and $u(a,b) = a+b$. If $0 < b(1-2b^2) \leq a \leq b$ then $f(a,b) = 2b(1-b^2) =\leq a+b = u(a,b)$. Let $\epsilon > 0$ be such that $2b - ab < 1$ for all $(a,b)^T \in B_\epsilon(0)$, then

$$\begin{aligned} f(a,b) - \ell(a,b) &= 2b(1-b)(1+b) - (1-ab)(a+b) \\ &\geq 2b(1-b)(1+a) - (1-ab)(a+b) \\ &= (b-a)(1-2b-ab) \ \geq \ 0. \end{aligned}$$

Similarly, $\ell(a,b) \leq f(a,b) \leq u(a,b)$ when $0 < a(1-2a^2) \leq b < a$. Therefore, Lemma 5.3 ensures that $\nabla f(0)$ exists and equals $\nabla u(0) = (1,1)^T \neq 0$. □

This example shows that the sequence of iterates produced by a GPS algorithm on a Lipschitz function can converge to a point where the gradient exists, is non-zero, but where zero does not belong to the generalized gradient. Proposition 5.1 is tight since it is only valid in $\mathbb{R}^1$. Therefore, when GPS is applied to a function in $\mathbb{R}^2$ that is not strictly differentiable, then it is possible that the limit of a refining subsequence is not a Clarke stationary point.

6. Discussion

GPS algorithms are well-suited for non-smooth optimization problems. The convergence behavior of these algorithms has been analyzed under a hierarchy of differentiability hypotheses.

The first analysis is due to Torczon (1997). She shows that the limit inferior of the mesh size parameter goes to zero. She also shows that when f is globally continuously differentiable then the limit inferior of the norm of the gradient of the sequence of iterates produced by the algorithm goes to zero. Audet and Dennis (2003) analyze convergence under less restrictive differentiability assumptions. They first identify limits of refining subsequences, regardless of the differentiability of f. Then they show that if f is strictly differentiable near the limit of any refining subsequence, the gradient at that point is zero. They also analyze the situation where f is simply Lipschitz continuous at the limit point. They show that Clarke derivatives are non-negative in a set of positive spanning directions. This does not imply Clarke stationarity.

The GPS examples presented in this paper confirm that these theoretical convergence results cannot be improved without either assuming more about the optimization problem, or without adding more requirements on the algorithm.

In particular, we showed that the requirements on the set of directions D and on the mesh update parameter τ cannot be relaxed, otherwise, the existence of a refining subsequence would be compromised.

We also showed that, even for continuously differentiable functions, the sequence of iterates produced by GPS algorithms can possess infinitely many limit points. Furthermore, the gradient at some (but not all) of these limit points can be non-zero.

Finally, we also showed that even when the sequence of iterates produced by GPS algorithms possesses a single limit point where the gradient exists, it can be non-zero. In $\mathbb{R}^1$, the limit point is necessarily a Clarke stationary point. But, in larger dimensions, zero may be excluded from the generalized gradient. This shows that the standard GPS class of algorithm does not guarantee to generate a Clarke stationary point for Lipschitz functions (the theory (Audet and Dennis, 2003) guarantees it only for strictly differentiable functions).

Recent work (Audet and Dennis, 2004) propose a new class of direct search algorithms that generalizes GPS by allowing polling directions to be chosen in a larger set. As the mesh size parameter goes to zero, the normalized polling directions become dense in the unit sphere. They show that the new method can produce a Clarke stationary point for locally Lipschitz functions.

Acknowledgments

The author is very grateful to John Dennis, whose questions prompted these examples. The author would also like to thank Mark Abramson and two anonymous referees for suggestions and corrections that improved the paper.

References

C. Audet and J. E. Dennis, Jr., "Analysis of generalized pattern searches," *SIAM Journal on Optimization* vol. 13, pp. 889–903, 2003.

C. Audet and J. E. Dennis, Jr., "Mesh adaptive direct search algorithms for constrained optimization," *Les cahiers du GERAD* G-2004-04, Montréal, 2004.

A. J. Booker, J. E. Dennis, Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, "A rigorous framework for optimization of expensive functions by surrogates," *Structural Optimization* vol. 17, no. 1, pp. 1–13, 1999.

F. H. Clarke, "Optimization and nonsmooth analysis," *SIAM Classics in Applied Mathematics* vol. 5, 1990, Philadelphia.

I. D. Coope and C. J. Price, "On the convergence of grid-based methods for unconstrained optimization," *SIAM Journal on Optimization* vol. 11, pp. 859–869, 2001.

R. M. Lewis and V. Torczon, "Rank ordering and positive basis in pattern search algorithms," *ICASE* TR 96-71, 1996.

R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

V. Torczon, "On the convergence of pattern search algorithms," *SIAM Journal on Optimization* vol. 7, no. 1, pp. 1–25, 1997.