ANALYSIS OF GENERALIZED PATTERN SEARCHES *

CHARLES AUDET † AND J.E. DENNIS JR. ‡

Abstract. This paper contains a new convergence analysis for the Lewis and Torczon GPS class of pattern search methods for linearly constrained optimization. The analysis is motivated by the desire to understand the behavior of the algorithm under hypotheses more consistent with properties satisfied in practice for a class of problems, discussed at various points in the paper, for which these methods are successful. Specifically, even if the objective function is discontinuous or extended valued, the methods find a limit point with some minimizing properties. Simple examples show that the strength of the optimality conditions at a limit point does not depend only on the algorithm, but also on the directions it uses, and on the smoothness of the objective at the limit point in question. This contribution of this paper is to provide a simple convergence analysis that supplies detail about the relation of optimality conditions to objective smoothness properties, and the defining directions for the algorithm, and it gives older results as easy corollaries.

Key words. Pattern search algorithm, linearly constrained optimization, surrogate-based optimization, nonsmooth optimization, derivative-free convergence analysis.

AMS subject classifications. 90C30, 90C56, 65K05

1. Introduction. Generalized pattern search (GPS) algorithms were defined and analyzed by Torczon [28] for derivative-free unconstrained optimization on continuously differentiable functions using positive spanning directions [24]. Lewis and Torczon showed that if the objective is continuously differentiable and if the set of directions that define the local search is chosen properly, then the GPS framework and convergence theory extends to bound constrained optimization [23] and more generally for problems with a finite number of linear constraints [25] by the appealing "barrier" strategy of declaring any infeasible point to be unacceptable as a next iterate. Our purpose here is to provide a new simpler unified analysis for the methods in [28, 23, 25], and to help elucidate the relationship between the algorithm, the search directions, and the local smoothness properties of the objective at certain specified limit points of the algorithm.

The optimization problem considered in this paper is:

$$\min_{x\in\Omega} f(x)\ ,\ \text{where } f:\Re^n \to \Re\cup\{\infty\}\ . \tag{1.1}$$

We assume as in [25] that $\Omega = \{x \in \Re^n : \ell \le Ax \le u\}$ where $A \in \mathcal{Q}^{m\times n}$ is a rational matrix, $\ell, u \in \{\Re \cup \{\pm\infty\}\}^m$ and $\ell < u$. The way of handling the linear constraints here, and indeed the entire algorithm, is the same as in [23] and [25]; but a key part of the analysis here is more general and much shorter.

We believe that the primary niche of GPS methods within nonlinear optimization stems from their effectiveness when used with surrogates [5, 6] for what are generally expensive objective function evaluations. Certainly our interest in them is based on

* Work of the first author was supported by NSERC (Natural Sciences and Engineering Research Council) fellowship PDF-207432-1998 during a post-doctoral stay at Rice University, and both authors were supported by DOE DE-FG03-95ER25257, AFOSR F49620-01-1-0013, The Boeing Company, Sandia LG-4253, ExxonMobil, and the LANL Computer Science Institute (LACSI) contract 03891-99-23.

†Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal (Québec), H3C 3A7 Canada (charlesa@gerad.ca)

‡Computational and Applied Mathematics Department, Rice University - MS 134, 6100 Main Street, Houston, Texas, 77005-1892 (dennis@caam.rice.edu)

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**
4. TITLE AND SUBTITLE **Analysis of Generalized Pattern Searches**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES **The original document contains color images.**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**15**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

success using this approach on some interesting engineering design problems. This motivation influences the way we like to view the methods, as well as what we are willing to assume about the objective function f to do the analysis. We will give contextual discussions of several ways in which GPS methods fill this niche - the first just below.

For many applied problems, a call to the subroutine that evaluates $f(x)$ may result unexpectedly in no value being returned, which we model as $f(x) = \infty$. This important issue is discussed in detail in [5], where GPS is effective on a helicopter rotor design example for which no value is returned roughly 66% of the time. The issue is discussed in a different algorithmic and application context in [7, 8]. The point is that because this happens in many applications, we are precluded from making global smoothness assumptions, including even continuity. We are not the first to observe that GPS can work well on nonsmooth problems. Hough, Kolda and Torczon note in an earlier version of [20] that "while the theory for pattern search assumes that f is continuously differentiable, pattern search methods can be effective on nondifferentiable (and even discontinuous) problems precisely because they do not explicitly rely on derivative information to drive the search."

We view the barrier approach as applying the algorithm not to f, but to the barrier function $f_\Omega = f + \psi_\Omega$, where ψ_Ω is the indicator function for Ω. It is zero on Ω and ∞ elsewhere. Clearly then, we do not evaluate $f(x)$ if x is infeasible because we know that its value is immaterial since the algorithm works with f_Ω, and the value of f_Ω is $+\infty$ on all points that are either infeasible or at which f is declared to be $+\infty$:

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

The reason that we treat together all the methods in [28, 23, 25] that use the barrier approach is that by viewing them as the same algorithm applied to f_Ω, we can treat them by corollaries of a single result, Theorem 3.7, that allows for extended values and other nonsmooth behavior. Our approach is first to identify a class of promising limit points produced by GPS applied to extended-valued discontinuous functions like f_Ω. If f is lower semicontinuous at such a limit point, we can make a weak optimality statement. Then we apply the Clarke calculus [9] locally to f at such a point to relate progressively stronger optimality conditions to progressively stronger local smoothness assumptions at the limit point.

Thus, the structure of our results will be that at some limit point whose existence is asserted independent of certain assumptions, we make those additional assumptions to draw stronger conclusions. This is standard for Newton or quasi-Newton methods ([27], *e.g.*, Theorem 8.6 pg 216 or virtually all of [22]), but it has not been the norm for direct search methods.

Specifically, we observe without assuming any smoothness that there is a convergent subsequence of the sequence $\{x_k\}$ of iterates produced by the algorithm. Obviously if $\{f(x_k)\}$ is bounded below, then $\lim_k f(x_k)$ is finite since the sequence is nonincreasing. Thus, if f is lower semicontinuous at any limit point $\bar{x}$ of the sequence of iterates, then $f(\bar{x}) \leq \liminf_k f(x_k) = \lim_k f(x_k)$. Our analysis is of interest for the heat intercept design problem we give in [21] where f is not continuous at one of the limit points generated, but a plot suggests that it is lower semicontinuous. In a case where $f(x_k) = \infty$, we believe that an optimization code should notify the user that it has found an interesting point at which the subroutine that evaluates f should be carefully examined in hopes of obtaining a value, which may correspond to a good design.

Again without any smoothness assumptions, we show that there is a limit point $\hat{x}$ of a subsequence of $\{x_k\}$ consisting of iterates that are local optimizers of $f(x)$ to a progressively finer resolution of the current mesh at those iterates (a formal definition of the mesh is given in Section 2). The directional tests that led GPS to refine the mesh at the terms of the subsequence are exactly that difference quotients for the Clarke generalized directional at $\hat{x}$ are nonnegative. If the Clarke derivatives exist at $\hat{x}$, as they will if f is locally Lipschitz at $\hat{x}$, then these nonnegative difference quotients pass through the limit to be nonnegative Clarke derivatives in the directions used.

Nonnegative directional derivatives in a set of directions are necessary conditions for optimality, but they are not the usual first order conditions. To get those, we assume in addition that the generalized gradient of f is a singleton. This extra smoothness causes the above directional optimality conditions to hold for all directions in the positive cone of those directions, and this together with the right choice of directions leads to the familiar first order optimality conditions. We give examples that supplement those in [1] and show that our results are sharp in that they predict the behavior of the algorithm.

We believe that it is useful to understand how the algorithm behaves in such cases because there will generally be no way of knowing beforehand whether the "blackbox" function given to the algorithm is at all smooth, and our analysis describe the minimal optimality conditions that can be guaranteed. We obtain as immediate corollaries earlier results that assumed global continuous differentiability.

The remainder of the paper is organized as follows: in the next section, we will give a brief description of the GPS algorithm class. We adhere to a slightly different, but equivalent version of the Lewis and Torczon algorithm, because our major interest in these algorithms is for problems where they are used with inexpensive surrogates for an expensive function. To see how easily and effectively surrogates can be incorporated into this version of GPS, see [5, 6]. In Section 3, we present the assumptions together with a discussion of our local smoothness conditions, then we give the key result, some easy corollaries for unconstrained problems together with a discussion of these results before we go on to the results for the linear constraints. Section 4 is devoted to some concluding remarks.

2. Generalized pattern search algorithms. Generalized pattern search algorithms for unconstrained or linearly constrained minimization generate a sequence of iterates $\{x_k\}$ in $\Re^n$ with non-increasing objective function values. Because of our interest in surrogate-based optimization, we like to view each iteration as being divided into two phases: an optional SEARCH and a local POLL, defined next.

In the SEARCH step, the barrier objective function f_Ω is evaluated at a finite number of points on a mesh (a discrete subset of $\Re^n$ defined below whose fineness is parameterized by the *mesh size parameter* $\Delta_k > 0$) to try to find one that yields a lower objective function value than the incumbent. Any strategy may be used to select the mesh points that are candidates to replace the incumbent, as long as only finitely many points (including none) are selected.

This is a key point. The SEARCH step accommodates whatever heuristics the user was already using to attack their problem using surrogates. One might do some random search on the mesh using the surrogate, or, as in the Boeing Design Explorer software [4], one might apply SQP to the surrogate problem and then move the solution to a nearby mesh point to choose the candidates at which to evaluate the expensive objective function in hopes of obtaining a better next iterate. Coope and Price [11] offer a possibility for a related framework that does not require pushing a surrogate solution to the mesh for it to become an acceptable trial point. It would be interesting to blend the analysis here with their related methods.

On the other hand, the freedom of the SEARCH step is definitely a theoretical liability. In [1] and here, there are examples of nonempty searches that spoil chances for the algorithm to find KKT points and of empty searches that mire the algorithm in at a poor point when a naive random selection from the current mesh in the SEARCH would generally lead to success. Regardless, this freedom must be retained. Indeed, for the Boeing example [5, 6], the algorithm with surrogates is much more efficient than Serafini's implementation of the Dennis-Torczon MDS/PDS algorithm [13]. This is not to disparage the MDS algorithm, which is very robust on that example.

Below, we will offer terminology consistent with Coope and Price [12] to replace the usual "successful/unsuccessful" terminology in the GPS literature. The original terminology was adequate until it was recognized that the "unsuccessful" iterations were the important ones because they produce *mesh local optimizers*, while successful iterations produce only *improved mesh points*, which we define now.

When the incumbent is replaced, *i.e.*, when $f_\Omega(x_{k+1}) < f_\Omega(x_k)$, or equivalently when $f(x_{k+1}) < f(x_k)$, then x_{k+1} is said to be an *improved mesh point*. When the SEARCH step fails in providing an improved mesh point, the POLL step is invoked. This second step consists of evaluating the barrier objective function at the neighboring mesh points to see if a lower function value can be found there. A crucial practical feature supported by the theory here, but originally in Torczon [28], is that as soon as an improved mesh point is found, polling can stop immediately.

When the POLL step fails in providing an improved mesh point, then the current incumbent solution is said to be a *mesh local optimizer* (*i.e.*, its objective function value is less than or equal to that of neighboring mesh points). The algorithm then refines the mesh by setting the mesh size parameter

$$\Delta_{k+1} = \tau^{w_k} \Delta_k \tag{2.1}$$

for $0 < \tau^{w_k} < 1$, where $\tau > 1$ is a rational number that remains constant over all iterations, and $w_k \leq -1$ is an integer bounded below by the constant $w^- \leq -1$.

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 (there is no sufficient decrease condition, another crucial practical feature supported by the theory in Torczon [28]) and here, the mesh size parameter is kept the same or is increased to carry out far reaching and inexpensive (if surrogates are used) SEARCH steps, and the process is reiterated. The coarsening of the mesh follows the rule

$$\Delta_{k+1} = \tau^{w_k} \Delta_k \tag{2.2}$$

where $\tau > 1$ is defined above and $w_k \geq 0$ is an integer bounded above by $w^+ \geq 0$. Our experience with surrogate-based SEARCH steps [5], [6] is that a great deal of progress can be made with few function values, and at least $n + 1$ function evaluations are needed only to show local mesh optimality, which indicates that the mesh needs to be refined (see [24] for defining a minimal number of polling directions).

By modifying the mesh size parameters as above, it follows that for any $k \geq 0$, there exists an integer $r_k \in \mathcal{Z}$ such that

$$\Delta_k = \tau^{r_k} \Delta_0. \tag{2.3}$$

The basic ingredient in the definition of the mesh is a set of positive spanning directions D in $\Re^n$ (more precisely, nonnegative linear combinations of the elements

- Initialization:
  Let x_0 be such that $f_\Omega(x_0)$ is finite, and let M_0 be the mesh on $\Re^n$ defined by $\Delta_0 > 0$, and let D_0 and x_0 be given (see equation (2.4)). Set the iteration counter k to 0.
- Search and poll step:
  Perform the search and possibly the poll steps (or only part of them) until an improved mesh point x_{k+1} with the lowest so far f_Ω value is found on the mesh M_k defined by equation (2.4).
  - Optional search: Evaluate f_Ω on a finite subset of trial points on the mesh M_k defined by equation (2.4) (the strategy that gives the set of points is usually provided by the user; it must be finite and the set can be empty).
  - Local poll: Evaluate f_Ω on the poll set defined in equation (2.5).
- Parameter update:
  If the search or the poll step produced an improved mesh point, *i.e.*, a feasible iterate $x_{k+1} \in M_k \cap \Omega$ for which $f_\Omega(x_{k+1}) < f_\Omega(x_k)$, then update $\Delta_{k+1} \geq \Delta_k$ according to rule (2.2).
  Otherwise, $f_\Omega(x_k) \leq f_\Omega(x_k + \Delta_k d)$ 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 (2.1).
  Increase $k \leftarrow k + 1$ and go back to the search and poll step.

Fig. 2.1. *A basic GPS algorithm*

of the set D span $\Re^n$). There is great freedom in choosing these directions, only the following additional rule needs to be respected: each direction $d_j \in D$ (for $j = 1, 2, \ldots, |D|$) is the product $G\bar{z}_j$ of the non-singular generating matrix $G \in \Re^{n\times n}$ by an integer vector $\bar{z}_j \in \mathcal{Z}^n$. Note that the same generating matrix is used for all directions. For convenience, the set D is also viewed as a real $n \times |D|$ matrix. Similarly, we denote the matrix whose columns are $\bar{z}_j$, for $j = 1, 2, \ldots, |D|$ by $\bar{Z}$; we can therefore write $D = G\bar{Z}$. At iteration k, the mesh is centered around the current iterate $x_k \in \Re^n$ and its fineness is parameterized through the mesh size parameter Δ_k as follows

$$M_k = \{x_k + \Delta_k Dz : z \in \mathcal{Z}_+^{|D|}\}, \tag{2.4}$$

where $\mathcal{Z}_+$ is the set of nonnegative integers. This way of describing the mesh differs from [28, 23, 25] because we think it easier to understand and work with.

At each iteration, some positive spanning matrix D_k composed of columns of D is used to construct the poll set. We write $D_k \subseteq D$ to signify that the matrix D_k is composed of columns of D. The poll set is composed of mesh points neighboring the current iterate x_k in the directions of the columns of D_k:

$$\text{Poll set:} \quad \{x_k + \Delta_k d : d \in D_k\}. \tag{2.5}$$

Rules for selecting D_k may depend on the user's dynamic intervention during the current run, or, for example, on the iteration number or the current iterate, *i.e.*, $D_k = D(k, x_k) \subseteq D$.

The algorithm is stated formally in Figure 2.1.

The search strategy is the key to effectiveness. In practice it allows the use of heuristic and surrogate methods to explore the domain of the variables. For example, one might apply a few generations of a genetic algorithm on the mesh to f_Ω or to a surrogate. The convergence analysis is independent of the SEARCH step, provided that it is finite and returns a point (or points) on the mesh. The POLL step applied to f_Ω, as we will see, guarantees that the limit point provided by the algorithm satisfies optimality conditions whose strength depends on the local smoothness of f at the limit point.

3. Convergence analysis.

Theorem 3.7 is our main result. It and Theorem 3.1 make no special assumptions about the crucial relationship between the directions D and the feasible region Ω. This means that they apply to quite general uses of GPS (see also the remark following Theorem 3.14), but without a connection between Ω and D, the resulting constrained optimality conditions are weak even when f is smooth. Its immediate corollary (Theorem 3.9) is the strongest result we expect for stationarity in the unconstrained case (see [1] for supporting examples).

Since one of the objectives of the paper is to simplify the convergence analysis of GPS, we include the proofs of all the results leading to our main one, even if some of them essentially can be found in previous work modulo the slightly different way of defining the mesh (we indicate the appropriate references).

3.1. Assumptions and smoothness requirements.

We make the standard assumption that all iterates produced by GPS lie in a compact set (see [2, 3, 10, 11, 12, 15, 16, 17]). A sufficient condition for this to hold is that the level set $L(x_0) = \{x \in \Omega : f(x) \leq f(x_0)\}$ is compact. We cannot assume that $L(x_0)$ is compact because we allow discontinuities and even $f(x) = \infty$, and so we do not know that $L(x_0)$ is closed. However we can assume that $L(x_0)$ is bounded so that its closure is compact.

Whatever we assume to ensure that the iterates are in a compact set, this already implies that there are convergent subsequences of the iteration sequence. This is enough to say that if f is lower semicontinuous at such a limit point $\bar{x}$, then $f(\bar{x}) \leq \lim_k f(x_k)$ for the entire iteration sequence. Of course, f can be infinite arbitrarily near a point where it is lower semicontinuous, and so we can say nothing about any derivatives at such an $\bar{x}$. For that, we will consider an interesting set of subsequences identified by the algorithm. Specifically, we will be concerned here, as in [2, 11, 12] with the iterates x_k that are mesh local optimizers for meshes that get infinitely fine. We will use $\bar{x}$ to denote generic limit points of the sequence of iterates, and $\hat{x}$ for limit points of mesh local optimizers for meshes that get infinitely fine. It is only at mesh local optimizers that Δ_k is reduced. This is not to say that other subsequences may not exhibit interesting first order behavior, but we can prove that these do, and that is more specific. The analysis is simpler if we assume that the mesh size is never coarsened, since obviously then the meshes become infinitely fine for every sequence of mesh local optimizers. However, we will not use this assumption since mesh coarsening can lead more rapidly to a more global solution.

To summarize, the convergence analysis provided below relies only on the following assumptions, and some results are stated in terms of the set of directions D.

A1: A function $f_\Omega = f + \psi_\Omega : \Re \to \Re \cup \{+\infty\}$ is available.
A2: The constraint matrix A is rational.
A3: All iterates $\{x_k\}$ produced by the algorithm lie in a compact set.

This allows us to prove the following result with an immediate, but rather strange implication - stationary points are the least interesting limit points GPS produces. Of course, if all the limit points are stationary points, then all are equally interesting.

THEOREM 3.1. *Under assumptions A1 and A3, there exists at least one limit point of the iteration sequence $\{x_k\}$. If f is lower semicontinuous at such a limit point $\bar{x}$, then $\lim_k f(x_k)$ exists and is greater than or equal to $f(\bar{x})$. If f is continuous at every limit point of $\{x_k\}$, then every limit point has the same function value.*

Proof. Since f is lower semicontinuous at $\bar{x}$, we know that for any subsequence $\{x_k\}_{k\in K}$ of the iteration sequence that converges to $\bar{x}$, $\liminf_{k\in K} f(x_k) \geq f(\bar{x})$, which is finite. But since the subsequence of function values is a subsequence of a nonincreasing sequence, they have the same lim inf. Thus, the entire sequence is also bounded below by $f(\bar{x})$, and so it converges. ∎

To prove more, we will need to assume more. In addition to A1-A3, previous work on pattern search algorithms assumes continuous differentiability of the function f on a neighborhood of the level set $L(x_0) = \{x \in \Omega : f(x) \leq f(x_0)\}$ ([2, 23, 25, 28, 11, 12]). In the unconstrained case, Torczon [28] shows that for GPS there exists a limit point $\bar{x}$ satisfying $\nabla f(\bar{x}) = 0$, and our [2] shows the same result for every limit point $\hat{x}$ of any sequence of mesh local optimizers for which $\lim_k \Delta_k = 0$. Note that since every limit point of the GPS sequence is a point of continuity in this case, nonstationary limit points, whose possible existence is shown in [1], are very interesting because with the right search step, or the right choice of directions, one can proceed to a feasible point with a better value of f. Our analysis below uses a weaker assumption at such a limit point (strict differentiability[1] of f at $\hat{x}$ instead of continuous differentiability on $L(x_0)$).

First we easily show (under no smoothness assumptions) the existence of at least one limit point of a subsequence of mesh local optimizers on meshes that get infinitely fine. Then, for those limit points where f is strictly differentiable, we show that the gradient is zero. To avoid confusion about the relative strength of assuming in the context of GPS that f is locally Lipschitz, or strictly differentiable at a point, or continuously differentiable, we will provide examples following Theorems 3.7 and 3.9 for which those results apply and earlier results do not. The original proof of the mesh refinement results were first given in [28] with a different description of the meshes.

We now proceed with some results on the behavior of the mesh and mesh size parameter. These results do not depend at all on the smoothness of f_Ω; they use just the definition of the algorithm and integrality of the matrix $\overline{Z}$ used to construct the set of directions D. For a different framework, Coope and Price relax the conditions on the mesh but they assume that the meshes become infinitely fine. This is an interesting tradeoff that puts the burden for ensuring that the meshes become infinitely fine onto the implementation, but allows for search points off the mesh and more freedom in the definition of the meshes.

3.2. Mesh refinement. The main result of this section is that there is a subsequence of mesh local optimizers for which the mesh size parameter goes to zero. The first lemma shows that for each mesh M_k, the minimal distance over all pairs of distinct mesh points is bounded below by the mesh size parameter Δ_k times a scalar.

[1]The function f is said to be *strictly differentiable* at x if for all v, $\displaystyle\lim_{y\to x, t\downarrow 0} \frac{f(y+tv)-f(y)}{t} = \nabla f(x)^T v$ (see Clarke [9]).

In the Euclidean norm, the proof involves the smallest singular value of G [28].

LEMMA 3.2. *For any integer $k \geq 0$, and any norm for which any nonzero integer vector has norm at least 1,*

$$\min_{u \neq v \in M_k} \|u - v\| \;\geq\; \frac{\Delta_k}{\|G^{-1}\|}.$$

Proof. Using equation (2.4), we let $u = x_k + \Delta_k D z_u$ and $v = x_k + \Delta_k D z_v$ be two distinct points on M_k with both z_u and z_v in $\mathcal{Z}_+^{|D|}$. Then

$$\|u - v\| \;=\; \Delta_k \|D(z_u - z_v)\| \;=\; \Delta_k \|G\bar{Z}(z_u - z_v)\| \;\geq\; \Delta_k \frac{\|\bar{Z}(z_u - z_v)\|}{\|G^{-1}\|} \;\geq\; \frac{\Delta_k}{\|G^{-1}\|}.$$

The last part of the inequality is due to the fact that $\bar{Z}(z_u - z_v)$ is a nonzero integer vector, thus its norm is greater than or equal to one. ∎

The previous result would not be true if the directions of D were not constructed through an integral matrix $\overline{Z}$. For example, in $\Re^1$ positive integer combinations of the columns of $D = [-1, +\pi]$ are a dense subset of the real line. Indeed, there are no $\overline{Z} = [\overline{z}_1, \overline{z}_2]$ with $\overline{z}_1, \overline{z}_2 \in \mathcal{Z}^1$ and $G \in \Re^1$ such that $D = G\overline{Z}$.

The next lemma shows that the mesh size parameters generated by the algorithm are bounded above (it is similar to a result in [2] for categorical variables).

LEMMA 3.3. *There exists a positive integer r^+ such that $\Delta_k \leq \Delta_0 \tau^{r^+}$ for any integer $k \geq 0$.*

Proof. Using assumption A3, we let $\mathcal{X}$ be a compact set in $\Re^n$ that contains all iterates, and denote its diameter by γ (i.e., the maximal distance between two of its points). If $\Delta_k > \gamma \cdot \|G^{-1}\|$, then Lemma 3.2 with ($v = x_k$) ensures that any trial point $u \in M_k$ different from x_k would have been outside of $\mathcal{X}$. But since no iterate is outside $\mathcal{X}$, it follows that at any iteration whose mesh size parameter exceeded $\gamma \cdot \|G^{-1}\|$, the iterate x_k is a mesh local optimizer. Thus Δ_k is bounded above by $\gamma \cdot \|G^{-1}\| \tau^{w^+}$ and the result follows by setting r^+ large enough so that $\Delta_0 \tau^{r^+} \geq \gamma \cdot \|G^{-1}\| \tau^{w^+}$. ∎

The proof of the next result is identical in spirit to that of the same result in Torczon [28] and adapted in [2] for categorical variables.

PROPOSITION 3.4. *The mesh size parameters satisfy* $\liminf\limits_{k \to +\infty} \Delta_k = 0$.

Proof. Suppose by way of contradiction that there exists a negative integer ρ such that $0 < \Delta_0 \tau^\rho \leq \Delta_k$ for all $k \geq 0$. Combining equation (2.3) with Lemma 3.3 implies that for any $k \geq 0$, r_k takes its value among the integers of the finite set $\{\rho, \rho + 1, \ldots, r^+\}$.

Since $x_{k+1} \in M_k$, equation (2.4) assures that $x_{k+1} = x_k + \Delta_k D z_k$ for some $z_k \in \mathcal{Z}_+^{|D|}$. Using equation (2.3) by substituting $\Delta_k = \Delta_0 \tau^{r_k}$ it follows that for any integer $N \geq 1$:

$$x_N = x_0 + \sum_{k=1}^{N-1} \Delta_k D z_k \;=\; x_0 + \Delta_0 D \sum_{k=1}^{N-1} \tau^{r_k} z_k \;=\; x_0 + \frac{p^\rho}{q^{r^+}} \Delta_0 D \sum_{k=1}^{N-1} p^{r_k - \rho} q^{r^+ - r_k} z_k$$

where p and q are relatively prime integers satisfying $\tau = \frac{p}{q}$. Since for any k the term $p^{r_k - \rho} q^{r^+ - r_k} z_k$ appearing in this last sum is an integer, it follows that all iterates lie on the translated integer lattice generated by x_0 and the columns of $\frac{p^\rho}{q^{r^+}} \Delta_0 D$.

Therefore, since all iterates belong to a compact set, it follows that there are only finitely many different iterates, and thus one of them must be visited infinitely many times. Therefore the rule presented in equation (2.2) is only applied finitely many times, and the one in equation (2.1) is applied infinitely many times. This contradicts the hypothesis that $\Delta_0\tau^\rho$ is a lower bound for the mesh size parameter. ∎

3.3. Main convergence result. Since the mesh size parameter shrinks only when a mesh local optimizer is detected, Proposition 3.4 guarantees that there are infinitely many mesh local optimizers. The following definition specifies the subsequences we use.

DEFINITION 3.5. *A subsequence of the GPS iterates consisting of mesh local optimizers, $\{x_k\}_{k\in K}$ (for some subset of indices K), is said to be a* refining subsequence *if $\{\Delta_k\}_{k\in K}$ converges to zero.*

The following shows the existence of convergent refining subsequences. Notice that if coarsening of the mesh was not allowed (*i.e.*, w^+ is set at 0 in equation (2.2)), then every subsequence of mesh local optimizers would be a refining subsequence, and so the next result would be trivial.

THEOREM 3.6. *There exists at least one convergent refining subsequence.*

Proof. Let K'' be the set of indices of iterates that are mesh local optimizers. Since the mesh is refined only at iterations when a local mesh optimizer is detected, Proposition 3.4 guarantees that there exists a subset of indices $K' \subset K''$ for which $\{\Delta_k\}_{k\in K} \downarrow 0$. Assumption $A3$ ensures that there exists a subset of indices $K \subset K'$ for which the subsequence of iterates $\{x_k\}_{k\in K}$ converges. ∎

We show below that the limit of any refining subsequence satisfies first order optimality conditions appropriate to the local smoothness of f. It is shown in [1] that even for a continuously differentiable f, the entire iteration sequence might not converge. There may even be infinitely many limit points, and not all of these limit points are stationary points.

Next is our basic, but key, result in which we apply Clarke's [9] generalized directional derivatives in a very straightforward way to the pattern search analysis. The results that follow specialize this result. Clarke's derivative at $\hat{x}$ in the direction d is defined for locally Lipschitz functions. Loosely speaking, it is defined to be the limit superior of the directional derivatives (in the direction d) of sequences converging to $\hat{x}$. The precise definition is given in the proof (see equation (3.1)).

THEOREM 3.7. *Under assumptions A1-A3, if $\hat{x}$ is any limit of a refining subsequence, and if d is any direction in D for which f at a poll step was evaluated for infinitely many iterates in the subsequence, and if f is Lipschitz near $\hat{x}$, then the generalized directional derivative of f at $\hat{x}$ in the direction d is nonnegative, i.e., $f^\circ(\hat{x}; d) \geq 0$.*

Proof. Let $\{x_k\}_{k\in K}$ be a refining subsequence and $\hat{x}$ its limit point obtained as in the statement of the Theorem. Since f is locally Lipschitz near $\hat{x}$, we have from Clarke [9] by definition that:

$$f^\circ(\hat{x}; d) \equiv \limsup_{y\to\hat{x},\ t\downarrow 0} \frac{f(y+td)-f(y)}{t} \ \geq \ \limsup_{k\in K} \frac{f(x_k+\Delta_k d)-f(x_k)}{\Delta_k}\ . \tag{3.1}$$

We need to know that the difference quotients are defined. First note that since f is Lipschitz near $\hat{x}$, it must be finite near $\hat{x}$. Note also that since a main point of the paper is to allow for extended valued functions and to justify the expedient of dealing with constraints by declining to evaluate the function f at infeasible points, we made the hypothesis that f was actually evaluated infinitely many times in the direction d. Therefore, for k sufficiently large all the poll steps in the direction d, $x_k + \Delta_k d$, are feasible. If they had not been, then f_Ω would have been infinite there and so f would not have been evaluated (recall that if $x \notin \Omega$, then $f_\Omega(x)$ is set at $+\infty$ and $f(x)$ is not evaluated).

Thus, we have that infinitely many of the right hand quotients of (3.1) are defined, and in fact they are the same as for f_Ω. This allows us to conclude that all of them must be nonnegative or else the corresponding poll step would have been successful in identifying an improved mesh point (recall that refining subsequences are constructed from mesh local optimizers). ∎

In the unconstrained case, there will always be a positive spanning set of directions that satisfy the hypotheses of the previous theorem. In the constrained case, there may be no such d if D were defined in a way incompatible with the geometry of the constraints (see the example in [23]). Thus in the next section, we will appeal to the construction in [25] to ensure that a sufficiently rich set of directions is used for bound or linear constraints. Again, we emphasize that GPS is a directional method, and the choice of directions is crucial.

The following example illustrates Theorem 3.7 on a Lipschitz function. This function looks like a convex function (quadratic in fact) that has been contaminated by local noise that decreases in amplitude near the minimizer. This behavior is common enough in practice to be the target class for implicit filtering algorithms [18].

EXAMPLE 3.8. *Consider the function* $f : \Re \to \Re$ *defined as* $f(x) = x^2(2+\sin(\frac{\pi}{x}))$. *This function possesses infinitely many local optima near* 0. *One can show that* f *is Lipschitz near* 0, *but it is not strictly differentiable there, and so certainly it is not continuously differentiable. In fact, the generalized gradient satisfies* $\partial f(0) = [-\pi, \pi]$.

If the GPS algorithm with empty SEARCH *steps,* $x_0 = \frac{1}{3}$, $\Delta_0 = 1$, $D = \{-1, 1\}$, $\Delta_{k+1} = \Delta_k$ *when an improved mesh point is found, and* $\Delta_{k+1} = \frac{1}{2}\Delta_k$ *when a mesh local optimizer is detected, is applied to this problem, then the sequence of iterates* $\{x_k\}$ *converges to* 0, *where* $f^\circ(0; \pm 1) = \pi \geq 0$ *as Theorem 3.7 guarantees. The proof of this claim can be seen from Table 3.1.*

TABLE 3.1
In four consecutive iterations, the iterates go from $x_k = \frac{1}{\alpha}, \Delta_k = \frac{3}{\alpha}$ *where* α *is a positive integer to* $x_{k+4} = \frac{x_k}{4}, \Delta_{k+4} = \frac{\Delta_k}{4}$.

k	x_k	$f(x_k)$	Δ_k	$f(x_k - \Delta_k)$	$f(x_k + \Delta_k)$	Iteration status
$4i$	$\frac{1}{\alpha}$	$\frac{2}{\alpha^2}$	$\frac{3}{\alpha}$	$f(\frac{1-3}{\alpha}) \geq \frac{4}{\alpha^2}$	$f(\frac{1+3}{\alpha}) \geq \frac{16}{\alpha^2}$	mesh local optimizer
$4i+1$	$\frac{1}{\alpha}$	$\frac{2}{\alpha^2}$	$\frac{3}{2\alpha}$	$f(\frac{2-3}{2\alpha}) = \frac{1}{2\alpha^2}$	$f(\frac{2+3}{2\alpha}) \geq \frac{25}{4\alpha^2}$	improved mesh point
$4i+2$	$\frac{-1}{2\alpha}$	$\frac{1}{2\alpha^2}$	$\frac{3}{2\alpha}$	$f(\frac{-1-3}{2\alpha}) \geq \frac{4}{\alpha^2}$	$f(\frac{-1+3}{2\alpha}) = \frac{2}{\alpha^2}$	mesh local optimizer
$4i+3$	$\frac{-1}{2\alpha}$	$\frac{1}{2\alpha^2}$	$\frac{3}{4\alpha}$	$f(\frac{-2-3}{4\alpha}) \geq \frac{25}{16\alpha^2}$	$f(\frac{-2+3}{4\alpha}) = \frac{1}{8\alpha^2}$	improved mesh point
$4(i+1)$	$\frac{1}{4\alpha}$	$\frac{1}{8\alpha^2}$	$\frac{3}{4\alpha}$			

Theorem 3.7 is the key to our analysis. The fact that its proof follows so directly from Clarke's definition of the generalized directional derivative is because unsuccessful polling at mesh local optimizers belonging to convergent refining sequences provide exactly the nonnegative difference quotients that Clarke's derivatives need since $x_k \to \hat{x}$ and $\Delta_k \downarrow 0$. We believe that this illustrates an intimate relationship between Clarke's generalized directional derivatives and the directional algorithm GPS.

3.4. Corollaries for unconstrained optimization. Before we add the complication of choosing directions for linear constraints, we give some easy corollaries of Theorem 3.7 for the unconstrained case.

In addition to the assumption that f is Lipschitz near $\hat{x}$, we assume that the generalized gradient of f at $\hat{x}$ is a singleton. This is equivalent to assuming that f is strictly differentiable at $\hat{x}$, *i.e.*, that there exists a $D_s f(\hat{x}) \in \Re^n$ such that $\lim\limits_{y \to \hat{x}, t \downarrow 0} \frac{f(y+tw)-f(y)}{t} = D_s f(\hat{x})^T w$ for all $w \in \Re^n$ (see [9], Proposition 2.2.1 or Proposition 2.2.4). Since the generalized gradient is a singleton $\partial f(\hat{x}) = \{D_s f(\hat{x})\}$, we use the standard notation for the gradient $\nabla f(\hat{x}) = D_s f(\hat{x})$.

THEOREM 3.9. *Under assumptions A1 and A3, let $\Omega = \Re^n$ and $\hat{x}$ be any limit of a refining subsequence. If f is strictly differentiable at $\hat{x}$, then $\nabla f(\hat{x}) = 0$.*

Proof. Again from [9], if f is strictly differentiable at $\hat{x}$, then for any direction $w \neq 0$, $f^\circ(\hat{x}; w) = \nabla f(\hat{x})^T w$. Now let $\hat{D}$ be any positive spanning set that is used infinitely many times in the refining subsequence, there must be at least one since D is finite. Then by Theorem 3.7, for each $d \in \hat{D}$, $0 \leq \nabla f(\hat{x})^T d$. Thus, if we write w as a nonnegative linear combination of the elements of $\hat{D}$, then we see immediately that $\nabla f(\hat{x})^T w \geq 0$. But the same construction for $-w$ shows that $-\nabla f(\hat{x})^T w \geq 0$ and so $\nabla f(\hat{x}) = 0$. ∎

The following example, based on a function taken from [19], illustrates the applicability of Theorem 3.9 by showing that any realization of GPS converges to the global minimizer for this convex function, which is strictly differentiable at its minimizer, but not continuously differentiable. We are not aware of any other results that apply to this example (the previous GPS analysis cannot be applied since they assumed global continuous differentiability).

EXAMPLE 3.10. *Consider the convex function $f : \Re \to \Re$ defined as $f(x) = \int_0^x \varphi(u)du$, where*

$$\varphi(u) = \begin{cases} u & \text{if } u \leq 0 \\ \frac{1}{1+\kappa} & \text{if } \kappa + 1 > \frac{1}{u} \geq \kappa \in \mathcal{Z}_+. \end{cases}$$

The function f is Lipschitz near $\hat{x} = 0$. It is shown in [19] that f has kinks at $\frac{1}{\kappa}$ with $\partial f(\frac{1}{\kappa}) = [\frac{1}{\kappa+1}, \frac{1}{\kappa}]$ for $\kappa = 1, 2, \ldots$ The corollary of Proposition 2.2.4 in [9] guarantees that f is not continuously differentiable near $\hat{x}$. Furthermore, $\partial f(0)$ reduces to the singleton $\{0\}$, and the same Proposition ensures that f is strictly differentiable at $\hat{x}$.

Applying Theorem 3.9 guarantees that any instance of any pattern search algorithm with any set of initial parameters generates a subsequence of iterates that converges to the global minimizer $\hat{x} = 0$ where $\nabla f(\hat{x}) = 0$, since the function is locally Lipschitz everywhere, and 0 is the only point where Clarke's generalized derivatives are nonnegative in all directions of a positive spanning set.

We certainly are not claiming that the weaker smoothness conditions we use imply that GPS methods *always* find a minimizer. This has been known to be false since the inception of GPS methods. Simple convex counterexamples come from starting at just the wrong point and choosing just the right ill-suited directions. This can be seen by considering $f(x) = |x_1| + |x_2|$ on $\Re^2$ and starting with $x_0 = (1, 0)^T$ with $D = \{(1,0)^T, (-1,1)^T, (-1,-1)^T\}$. The initial point x_0 is a mesh local optimizer for every $\Delta > 0$, and so the iteration never moves from x_0 with an empty SEARCH step.

Unlike the corollaries below that require more smoothness, our theorem applies to this simple example and describes exactly what happens; f is regular at $\hat{x}$ and the directional derivatives along the members of D are nonnegative.

The advantage of our analysis over the previous ones is that it can be applied to a wider class of problems, and that it says what actually happens when the algorithm is applied to them.

The following two corollaries assume continuous differentiability. We have discussed how for our applications, this assumption unlikely to be satisfied, except perhaps locally. We include these results only to tie our results here to earlier results that use global continuous differtiabililty. The first corollary strengthens our result in [2]. It shows that the limit of the gradient for any refining subsequence converges to zero, even if the subsequence itself does not converge.

COROLLARY 3.11. *Let $\Omega = \Re^n$ and f be continuously differentiable on a neighborhood of a compact set containing all the iterates $\{x_k\}$. Then for any refining subsequence $\{x_k\}_{k \in K}$, $0 = \lim_{k \in K} \nabla f(x_k)$.*

Proof. We have assumed A3, A2 is vacuous, and continuous differentiability implies assumption A1. If $\hat{x}$ is any limit point of a refining subsequence, then continuous differentiability implies strict differentiability at $\hat{x}$ and so $\nabla f(\hat{x}) = 0$ from Theorem 3.9. Since the continuous image of a compact set is compact, the entire sequence of gradients of any refining subsequence is in a compact set. Thus, there must be a subsequence $\{x_k\}_{k \in K'}$ of the refining subsequence for which $\lim_{k \in K'} \nabla f(x_k) = \limsup_k \nabla f(x_k)$. But then $\{x_k\}_{k \in K'}$ has a convergent subsequence, and its limit point has a zero gradient because it is a limit point of a refining subsequence, and so $0 = \limsup_k \nabla f(x_k)$. ∎

A consequence of the previous result is that under the assumption that f is continuously differentiable, any limit point of a refining sequence has a zero gradient.

The fact that under the assumption of continuous differentiability the limit of the gradients of any refining subsequence is zero was pointed out in [14]. Earlier, under strong restrictions on the algorithm, it was shown in [28] that $0 = \lim_k \nabla f(x_k)$. One of those restrictions is that $\lim \Delta_k = 0$, which we proved above is already is enough to say that the limit of the gradients at the mesh local optimizers is zero since then they are a refining subsequence. Thus, we will not discuss the restrictions needed for the stronger result, since they are too constraining for our class of problems.

The next corollary (really a corollary of Corollary 3.11) is Torczon's result from [28], strengthened by the same result from [14].

COROLLARY 3.12. *Let $\Omega = \Re^n$ and f be continuously differentiable on a neighborhood of a compact set containing all the iterates $\{x_k\}$, then some limit point $\hat{x}$ of $\{x_k\}$ satisfies $\nabla f(\hat{x}) = 0$. The limit of the gradients for any refining subsequence is zero.*

Proof. Every refining subsequence is a subsequence of $\{x_k\}$. ∎

In summary, if assumptions A1 and A3 are satisfied, then the algorithm guarantees the following hierarchy of convergence behavior.

(i) If f is lower semicontinuous at any limit point $\bar{x}$ of the GPS iteration sequence, then Theorem 3.1 says that $f(\bar{x}) \leq \lim_k f(x_k)$.

(ii) Every limit point of the iteration sequence at which f is continuous has the same function value $\lim_k f(x_k)$ whether or not it is a stationary point. Thus, if GPS produces a nonstationary limit point [1], which must necessarily be a limit point of improved mesh points (formerly called successful iterations), then there is a descent direction from that limit point, and so, despite finding a stationary point, the directions were poorly suited to the problem.

(iii) There is at least one $\hat{x}$ that is a limit point of a refining subsequence *i.e.*, $\hat{x}$ is a limit point of a sequence of local optimizers on meshes that get infinitely fine. If the function f is lower semicontinuous but not even Lipschitz near $\hat{x}$, then nothing additional to the above is claimed about optimality conditions satisfied by $\hat{x}$.

(iv) If f is Lipschitz near $\hat{x}$, then Theorem 3.7 holds and Clarke's generalized derivatives satisfy $f^\circ(\hat{x}; d) \geq 0$ for some directions $d \in D$ that form a positive spanning set. In addition, $f(\hat{x}) = \lim_k f(x_k)$ since f is continuous at $\hat{x}$.

(v) If f is regular[2] at $\hat{x}$, then the directional derivatives satisfy $f'(\hat{x}; d) \geq 0$ for some directions $d \in D$, a positive spanning set, and $f(\hat{x}) = \lim_k f(x_k)$.

(vi) If f is strictly differentiable at $\hat{x}$, then Theorem 3.9 holds and $\nabla f(\hat{x}) = 0$, but its function value $\lim_k f(x_k)$ is the same as at any other limit point of the entire GPS iteration sequence at which f is continuous (by (ii)).

(vii) If f is globally continuously differentiable (as assumed in earlier analyses), this means that every limit point of a refining subsequence is a stationary point as in item(vi) and that the gradients of a refining subsequence converge to zero, whether or not the subsequence converges. However, as was shown in [1], there still there can be limit points of the entire GPS iteration sequence that are not stationary points. Though such points have the same function value as the stationary points, there is a descent direction from such points that lead to lower function values.

3.5. Linearly constrained convergence results. In this section, we will consider only the case where Ω is defined through a finite set of linear constraints. In order to prove the relevant optimality results, we will have to assume that D, even though finite, is rich enough to generate POLL sets that conform to the geometry of the boundary of Ω. Furthermore, to apply our proof technique, we must ensure that the spanning sets that reflect this geometry get used infinitely many times as we converge to a point on the boundary. Lewis and Torczon [25] show how to use standard linear algebra tools to generate the requisite positive spanning matrices $D_k \subseteq D$. This relies on assumption A2, the rationality of the constraint matrix A.

We pause to remind the reader that for $x \in \Omega$, the tangent cone to Ω at x is $T_\Omega(x) = \mathrm{cl}\{\mu(w - x) : \mu \geq 0, w \in \Omega\}$. The normal cone to Ω at x is $N_\Omega(x)$ and can be written as the polar of the tangent cone: $N_\Omega(x) = \{v \in \Re^n : \forall w \in T_\Omega(x),\ v^T w \leq 0\}$. It is the nonnegative span of all the outwardly pointing constraint normals at x.

It would add unnecessary length to this paper to rewrite the construction given by Lewis and Torczon [25] for D and the choice rule for D_k from D at each iteration (their notation for D_k is Γ_k). The construction is presented there quite succinctly in Section 8 of [25] where they consider implementation issues, including difficulties inherent to degenerate constraints. We will use the following abstracted version of their direction choice.

DEFINITION 3.13. *A rule for selecting the positive spanning sets $D_k = D(k, x_k) \subseteq D$* conforms to Ω *for some $\epsilon > 0$, if at each iteration k and for each y in the boundary of Ω for which $\|y - x_k\| < \epsilon$, $T_\Omega(y)$ is generated by a nonnegative linear combinations of the columns of a subset D_k^y of D_k.*

With this definition, we are ready for our next convergence result. Note that if $x_k \in \Omega$ is not near the boundary, then D_k need only provide a positive spanning set for $\Re^n$, which is completely sensible. However, in our experience, it is best not to take ϵ too small so that the iterates crowd up against the boundary of Ω and the mesh size becomes small. This is mitigated somewhat by allowing variable coarsening of the mesh as in equation (2.2).

THEOREM 3.14. *Under assumptions A1-A3, if f is strictly differentiable at a limit*

[2] The function f is said to be *regular* at x if for all v, the one-sided directional derivative exists and coincides with $f^\circ(x; v)$ (see Clarke [9]).

point $\hat{x}$ of a refining subsequence, and if the rule for selecting the positive spanning sets $D_k = D(k, x_k) \subseteq D$ conforms to Ω for an $\epsilon > 0$, then $\nabla f(\hat{x})^T w \geq 0$ for all $w \in T_\Omega(\hat{x})$, and $-\nabla f(\hat{x}) \in N_\Omega(\hat{x})$. Thus, $\hat{x}$ is a KKT point.

Proof. If $\hat{x}$ is interior to Ω, then the result is just Theorem 3.9, and so we can proceed directly to the case where $\hat{x}$ is on the boundary of Ω.

Suppose that the rule for selecting $D_k \subseteq D$ conforms to Ω for some fixed $\epsilon > 0$, and that there are finitely many linear constraints, then $D_k^{\hat{x}}$ spans $T_\Omega(\hat{x})$ for large $k \in K$. It follows that there can only be finitely many different such sets $D_k^{\hat{x}}$ for $k \in K$. Let $D^{\hat{x}} \subseteq D$ be one of them that occur infinitely many times.

Theorem 3.7 implies that $\nabla f(\hat{x})^T d \geq 0$ for every column d of $D^{\hat{x}}$. But since every $w \in T_\Omega(\hat{x})$ is a nonnegative linear combination of the columns of $D^{\hat{x}}$, then $\nabla f(\hat{x})^T w \geq 0$. To complete the proof, we multiply both sides by -1 and conclude that $-\nabla f(\hat{x})$ is in $N_\Omega(\hat{x})$. ∎

REMARK 3.15. *If f were only assumed to be Lipschitz near $\hat{x}$, then we could still conclude as in Theorem 3.7, that $f^\circ(\hat{x}; d) \geq 0$ for every column d of $D^{\hat{x}}$.*

The following corollary is Lewis and Torczon's result from [25] which relies on a stronger differentiability assumption.

COROLLARY 3.16. *If A2 and A3 hold and f is continuously differentiable on a neighborhood of a compact set containing all the iterates $\{x_k\}$, and if the rule for selecting the positive spanning sets $D_k = D(k, x_k) \subseteq D$ conforms to Ω for an $\epsilon > 0$, then there exists a limit point $\hat{x}$ of $\{x_k\}$ such that $\nabla f(\hat{x})^T w \geq 0$ for all $w \in T_\Omega(\hat{x})$, and $-\nabla f(\hat{x}) \in N_\Omega(\hat{x})$. Thus, $\hat{x}$ is a KKT point.*

Proof. The proof follows from Theorem 3.14 since every refining subsequence is a subsequence of $\{x_k\}$ and continuous differentiability implies strict differentiability. ∎

4. Concluding remarks. This paper puts together ways to choose the directions and results on properties of the mesh by Lewis and Torczon, some observations of ours about what is needed to obtain convergence of those algorithms (such as refining subsequences), and elements of nonsmooth analysis set forth by Clarke. Clarke's analysis is perfectly suited to expose the first order optimality conditions at limit points of certain subsequences of the GPS iterates under weakened assumptions that correspond to some real problems for which GPS is quite effective.

We believe that our analysis helps confirm a remark of [25] that GPS methods for general constraints will not be based on the appealingly simple barrier strategy of placing a high function value on infeasible trial points. In [3], we suggest and analyze a GPS algorithm for general constraints based not on a single objective, but on the interesting new filter approach of Fletcher *et al.* [15], [16] and [17]. In [26], Lewis and Torczon give a successive augmented Lagrangian pattern search approach together with its convergence analysis.

Finally, we wish to acknowledge a helpful referee and Major Mark Abramson USAF for many insightful comments that improved the presentation.

REFERENCES

[1] AUDET C. (1998), "Convergence results for patterns search algorithms are tight," *TR98-24* Department of Computational & Applied Mathematics, Rice University, Houston Texas.

[2] AUDET C., DENNIS J.E.JR (2000), "On the convergence of pattern search algorithms with mixed variables," *SIAM Journal on Optimization* Vol.11 No.3, 573-594.

[3] AUDET C., DENNIS J.E.JR (2000), "A Pattern Search Filter Method for Nonlinear Programming without Derivatives," *TR00-09*, Department of Computational and Applied Mathematics, Rice University, Houston Texas.

[4] Audet, C., Booker, A.J., Dennis, J.E.Jr., Frank P.D., and Moore, D.(2000), "A Surrogate-Model-Based Method For Constrained Optimization", AIAA no.4891, *AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*.

[5] Booker A.J., Dennis J.E.Jr, Frank P.D., Serafini D.B., Torczon V. and Trosset M.W.(1999), "A rigorous framework for optimization of expensive functions by surrogates," *Structural Optimization* Vol.17 No.1, 1-13.

[6] Booker A.J., Dennis J.E.Jr, Frank P.D., Moore D.W. and Serafini D.B.(1999), "Managing Surrogate Objectives to Optimize a Helicopter Rotor Design - Further Experiments," AIAA Paper 98-4717, St. Louis, September 1998.

[7] Choi T.D., Eslinger O.J., Kelley C.T., David J.W., Etheridge M.(1998), "Optimization of automotive valve train components with implicit filtering," to appear in *Optimization and Engineering*.

[8] Choi T.D., Kelley C.T.(1999), "Superlinear convergence and implicit filtering," *SIAM Journal on Optimization* Vol.10 No.4, 1149–1162.

[9] Clarke, F.H.(1990) "Optimization and Nonsmooth Analysis," SIAM Classics in Applied Mathematics Vol.5, Philadelphia.

[10] Conn A.R., Gould N.I.M., Toint Ph.L.(1991) "A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds," *SIAM Journal on Numerical Analysis* Vol 28, 545–572.

[11] Coope I.D., Price C.J. (2001), "On the convergence of grid-based methods for unconstrained optimization," *SIAM Journal on Optimization* Vol 11, 859-869.

[12] Coope I.D., Price C.J. (2000), "Positive bases in optimization," University of Canterbury, Dept of Mathematics and Statistics, Report UCCDMS2000/12.

[13] Dennis J.E., Torczon V.(1991), "Direct Search Methods on Parallel Machines," *SIAM Journal on Optimization* Vol 1, pp. 448–474.

[14] Dolan E., Lewis M., Torczon V.(2000), "On the local convergence of pattern search", ICASE Technical Report 2000-36.

[15] Fletcher R., Leyffer S.(1997), "Nonlinear Programming without a penalty function," Dundee University, Dept. of Mathematics, Report NA/171.

[16] Fletcher R, Leyffer S., Toint Ph.L.(1998), "On the global convergence of an SLP-Filter algorithm," Dundee University, Dept. of Mathematics, Report NA/183.

[17] Fletcher R, Gould N.I.M., Leyffer S., Toint Ph.L.(1999), "On the global convergence of trust-region SQP-Filter algorithms for general nonlinear programming," Department of Mathematics, FUNDP, Namur (B), Report 99/03.

[18] Gilmore P, Kelley, C.T. (1995), "An implicit filtering algorithm for optimization of functions with many local minima," *SIAM Journal on Optimimization*, Vol. 5, 269285.

[19] Hiriart-Urruty J.-B., Lemaréchal C.(1993) "Convex analysis and minimization algorithms," Springer-Verlag Berlin, New-York.

[20] Hough P.D., Kolda T.G., Torczon V.(2000), "Asynchronous parallel pattern search for nonlinear optimization," Sandia National Laboratories, Report SAND2000-8213.

[21] Kokkolaras M., Audet C., Dennis J.E.Jr (2001), "Mixed variable optimization of the number and composition of heat intercepts in a thermal insulation system," *Optimization and Engineering*, Vol.2 No.1, 5–29.

[22] Lin C-J., Moré J.J.(1999), "Newton"s method for large bound-constrained optimization problems" *SIAM Journal on Optimimization*, Vol 9, 1100-1127.

[23] Lewis R.M., Torczon V.(1999), "Pattern search algorithms for bound constrained minimization," *SIAM Journal on Optimization*, Vol.9, 1082-1099.

[24] Lewis R.M., Torczon V.(1996), "Rank ordering and positive basis in pattern search algorithms," *ICASE NASA Langley Research Center* TR 96-71.

[25] Lewis R.M., Torczon V.(2000), "Pattern search methods for linearly constrained minimization," *SIAM Journal on Optimization*, Vol.10 No.3, 917–941.

[26] Lewis R.M., Torczon V.(1998), "A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds," *ICASE NASA Langley Research Center* TR 98-31.

[27] Nocedal J., Wright S.J.(1999) "Numerical Optimization," Springer Series in Operations Research, NY

[28] Torczon V.(1997), "On the Convergence of Pattern Search Algorithms," *SIAM Journal on Optimization* Vol.7 No.1, 1–25.