SIAM J. OPTIM.
Vol. 13, No. 3, pp. 889–903

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 generalized pattern search (GPS) class of methods for unconstrained and linearly constrained optimization. This analysis is motivated by a desire to understand the successful behavior of the algorithm under hypotheses that are satisfied by many practical problems. 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 depends not only on the algorithm, but also on the directions it uses and on the smoothness of the objective at the limit point in question. The 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 to the defining directions for the algorithm, and it gives previous results as corollaries.

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

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

PII. S1052623400378742

1. Introduction.

Generalized pattern search (GPS) algorithms were defined and analyzed by Torczon [29] for derivative-free unconstrained optimization on continuously differentiable functions using positive spanning directions. Lewis and Torczon [24] introduced the idea of using positive spanning directions with GPS. In [23], they showed that if the objective is continuously differentiable and if the set of directions that define the local search is chosen properly with respect to the boundary of the feasible region, then the GPS framework and convergence theory extend to bound-constrained optimization. In [25], they showed the same results for problems with a finite number of linear constraints. Both these extensions use the appealing "barrier" strategy of declaring any infeasible point to be unacceptable as a next iterate. Our purpose here is to provide a new unified analysis for the methods in [29, 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

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

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 $\ell \le u$.

*Received by the editors September 26, 2000; accepted for publication (in revised form) July 6, 2002; published electronically February 12, 2003. This work was 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.

http://www.siam.org/journals/siopt/13-3/37874.html

†Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal, QC, H3C 3A7 Canada (Charles.Audet@gerad.ca, http://www.gerad.ca/Charles.Audet). This author was supported by NSERC (Natural Sciences and Engineering Research Council) fellowship PDF-207432-1998 during a postdoctoral stay at Rice University.

‡Computational and Applied Mathematics Department, Rice University, MS 134, 6100 Main Street, Houston, TX 77005-1892 (dennis@caam.rice.edu, http://www.caam.rice.edu/~dennis).

GPS methods are extremely effective for some engineering design problems with expensive function evaluations when used with less expensive surrogates [5, 6]. For these and many other applied problems, a call to the subroutine that evaluates $f(x)$ may result unexpectedly in no value being returned even for a feasible x, which we model as $f(x) = \infty$. Reasons for this behavior are discussed in [5], where GPS with surrogates is shown to be 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]. In such instances, we cannot assume global smoothness, not even continuity. We are not the first to observe that GPS can work well on nonsmooth problems, but previous convergence theorems do not apply to such problems.

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 $+\infty$:

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

The reason that we treat together all the methods in [29, 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. To make statements about optimality conditions at these limit points, we work not with f_Ω but with f. 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 (e.g., [27, Theorem 8.6, p. 216] or virtually all of [22]), but it has not been the norm for direct search methods.

Specifically, without assuming any smoothness, we observe that there is a convergent subsequence of the sequence $\{x_k\}$ of iterates produced by the algorithm. Since $\{f(x_k)\}$ generated by the algorithm is nonincreasing, it is convergent to a finite limit if it is bounded below. 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 given in [21], where f is not continuous at one of the limit points generated, but a plot suggests that it is lower semicontinuous.

Again without any smoothness assumptions, we show that there is a limit point $\hat{x}$ of a subsequence of $\{x_k\}$ consisting of iterates on progressively finer meshes. (A formal definition of the mesh is given in section 2.) These specific iterates of interest are *mesh local optimizers* in that they minimize the function on a positive spanning set of neighboring mesh points. This will be made precise in section 2.

The directional tests that led GPS to refine the mesh at mesh local optimizers are exactly that difference quotients be nonnegative for the Clarke generalized directional derivative at $\hat{x}$. 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 on a set of positive spanning directions for $\Re^n$ are a necessary condition for optimality, but that is not the usual first order condition. To get the usual condition that the gradient is zero, 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 $\Re^n$. We give examples that supplement those in [1] and show that our results are sharp in that they predict the behavior of the algorithm.

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. In section 3, we present the assumptions together with a discussion of our local smoothness conditions, give the key result and some immediate corollaries for unconstrained problems together with a discussion of these results, and then go on to the results for the linear constraints. Section 4 is devoted to some concluding remarks.

2. GPS algorithms.

GPS algorithms for unconstrained or linearly constrained minimization generate a sequence of iterates $\{x_k\}$ in $\Re^n$ with nonincreasing objective function values. Each iteration is 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 his or her 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. In [13], they apply the Clarke analysis given 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 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 [28] of the Dennis–Torczon MDS/PDS algorithm [14]. This is not to disparage the MDS algorithm, which is very robust on that example.

Below, we will offer terminology consistent with that of Coope and Price 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 to provide 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 whether a lower function value can be found there.

When the POLL step fails to provide 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 $w^- \leq -1$.

A feature first noted in Torczon [29] and also supported in the analysis given here is that if either the SEARCH or POLL step produces an improved mesh point, the current iteration can stop, and 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 to carry out the next SEARCH step, 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 to show only 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 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 nonsingular 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 [29, 23, 25].

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 as follows.

A BASIC GPS ALGORITHM.

- *Initialization*:
  Let x_0 be such that $f_\Omega(x_0)$ is finite. Let D be a positive spanning set, and let M_0 be the mesh on $\Re^n$ defined by $\Delta_0 > 0$ and D_0 (see (2.4)). Set the iteration counter k to 0.
- SEARCH *and* POLL *steps*:
  Perform the SEARCH and possibly the POLL steps (or only part of them) until an improved mesh point x_{k+1} with the lowest f_Ω value so far is found on the mesh M_k defined by (2.4).
  - Optional SEARCH: Evaluate f_Ω on a finite subset of trial points on the mesh M_k defined by (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 (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 $x_{k+1} = x_k$ and 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.

The SEARCH strategy is the key to the algorithm's effectiveness, as we discussed above. 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. 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 theorem, even if some of them can essentially 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, 16, 17, 18]). A sufficient condition for this to hold is that the level set $L(x_0) = \{x \in \Omega : f(x) \leq f(x_0)\}$ be 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, arbitrarily near a point at which it is lower semicontinuous, f can be infinite, which means that there can be points of the sort mentioned above at which f fails to evaluate arbitrarily near $\bar{x}$, but it also means that 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. The analysis would be simpler if we assumed that the mesh size was never coarsened, since obviously then the meshes would 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 deeper basin than might be found without it.

To summarize, the convergence analysis provided below relies only on the following assumptions.

A1: A function $f_\Omega = f + \psi_\Omega : \Re \rightarrow \Re \cup \{+\infty\}$ and initial point $x_0 \in \Re^n$ (with $f_\Omega(x_0) < \infty$) are available.
A2: The constraint matrix A is rational.
A3: All iterates $\{x_k\}$ produced by the GPS algorithm lie in a compact set.

We now prove the following result with an immediate, but rather strange, implication—stationary points are the least interesting locally smooth limit points that GPS produces, in the sense that all limit points have the same function value but there are descent directions leading from any locally smooth nonstationary points. 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 limit inferior. Thus, the entire sequence is also bounded below by $f(\bar{x})$, and thus 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 $L(x_0) = \{x \in \Omega : f(x) \leq f(x_0)\}$ (see [2, 23, 25, 29, 11, 12]). In the unconstrained case, Torczon [29] 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 point with a better value of f. Our analysis below uses the weaker assumption of strict differentiability (defined in the first paragraph of section 3.4) at such a limit point 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 at which 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, 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 proofs of the mesh refinement results were first given in [29] 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 [11] 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 into 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 defined by (2.4), the minimal distance over all pairs of distinct mesh points is bounded below by the mesh size parameter Δ_k times a scalar. In the Euclidean norm, the proof involves the smallest singular value of G (see [29]).

LEMMA 3.2. *For any integer $k \geq 0$, any norm for which any nonzero integer vector has norm at least* 1*, and M_k defined by* (2.4)*,*

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

Proof. Using (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 separation between mesh points shown by Lemma 3.2 depends on the directions in D being integer linear combinations of the columns of a fixed nonsingular $n \times n$ generating matrix. For example, in $\Re^1$, positive integer combinations of the columns of $D = [-1, +\pi]$ are a dense subset of the real line. This is not a counterexample to Lemma 3.2, because the matrix $[-1, +\pi]$ cannot be written as a scalar multiple of a 1×2 integer matrix.

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. *Under assumptions* A1 *and* A3*, 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 exceeds $\gamma \cdot \|G^{-1}\|$, 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 [29] and that adapted in [2] for categorical variables.

PROPOSITION 3.4. *Under assumptions* A1 *and* A3*, the mesh size parameters satisfy* $\liminf_{k\to+\infty} \Delta_k = 0$.

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

$$x_N = x_0 + \sum_{k=0}^{N-1} \Delta_k D z_k \;=\; x_0 + \Delta_0 D \sum_{k=0}^{N-1} \tau^{r_k} z_k \;=\; x_0 + \frac{p^\rho}{q^{r^+}} \Delta_0 D \sum_{k=0}^{N-1} p^{r_k-\rho} q^{r^+-r_k} z_k,$$

where p and q are relatively prime integers satisfying $\tau = p/q$. 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 $p^\rho/q^{r^+}\Delta_0 D$.

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 (2.2) is applied only finitely many times, and the one in (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 were not allowed (i.e., w^+ were set at 0 in (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. *Under assumptions* A1 *and* A3*, 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 directional 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 (3.1)).

THEOREM 3.7. *Under assumptions* A1–A3*, if $\hat{x}$ is any limit of a refining subsequence, 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

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

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. But since they are defined, 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 is 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 [19].

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.

Theorem 3.7 is the key to our analysis. Its proof follows so directly from Clarke's

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

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
$4(i+1)$	$\frac{1}{4\alpha}$	$\frac{1}{8\alpha^2}$	$\frac{3}{4\alpha}$			

definition of the generalized directional derivative because unsuccessful polling at mesh local optimizers belonging to convergent refining sequences provides 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 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 $\lim_{y \to \hat{x}, t \downarrow 0} \frac{f(y+tw)-f(y)}{t} = D_s f(\hat{x})^T w$ 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 let $\hat{x}$ 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 $\nabla f(\hat{x})^T w \geq 0$. However, 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 [20], 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. Previous GPS analysis techniques that use global continuous differentiability do not apply to this example.

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 [20] 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 Lipschitz everywhere, and 0 is the only point at which Clarke's generalized derivatives are nonnegative in all directions of a positive spanning set.

We certainly are not claiming that the weaker smoothness conditions that 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. 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 following two corollaries assume continuous differentiability. We have discussed how, for our applications, this assumption is unlikely to be satisfied, except perhaps locally. We include these results only to tie our results here to earlier results that use global continuous differentiability. 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* A1 *and* A3 *hold for* $\Omega = \Re^n$ *and* f *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. 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 [15]. Earlier, under strong restrictions on the algorithm, it was shown in [29] that $0 = \lim_k \nabla f(x_k)$. One of those restrictions is that $\lim \Delta_k = 0$, which we proved above 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 is Torczon's result from [29], strengthened by the same result from [15].

COROLLARY 3.12. *Let* A1 *and* A3 *hold for* $\Omega = \Re^n$, *and let* 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, although there is always at least one limit point that is a stationary point, if GPS produces a nonstationary limit point [1], then it is more promising than any stationary limit point because they have the same function value, but there is a descent direction from the nonstationary limit point. The conclusion is that the directions used 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[1] 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 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 leads to lower function values.

3.5. Linearly constrained convergence results. In this section, we will consider only the case in which 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 $D_k \subseteq D$. The convergence analysis 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 $T_\Omega(x) = \text{cl}\{\mu(w - x) : \mu \geq 0, w \in \Omega\}$. 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

[1]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]).

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 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 when the iterates approach the boundary with small values of the mesh size parameter, the rule for selecting the mesh size parameter conforms to the boundary of Ω. This is mitigated somewhat by allowing variable coarsening of the mesh as in (2.2).

THEOREM 3.14. *Under assumptions* A1–A3*, if* f *is strictly differentiable at a limit 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 thus we can proceed directly to the case in which $\hat{x}$ 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 $D_k^{\hat{x}}$ generates $T_\Omega(\hat{x})$ for large $k \in K$. It follows that there can be only finitely many different such sets $D_k^{\hat{x}}$ for $k \in K$. Let $D^{\hat{x}} \subseteq D$ be one of them that occurs 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* A1–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 exposing 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 an observation 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. This is because, to prove the efficacy of the barrier strategy, the positive spanning set D, from which all the GPS directions are chosen, is finite, and thus it cannot be certain to generate the tangent cone at every boundary point of a nonpolygonal feasible region that the iteration approaches.

In [3], we suggest and analyze a GPS algorithm for general constraints, based not on a single objective but on the new filter approach of Fletcher and collaborators [16, 17, 18]. In [26], Lewis and Torczon give a successive augmented Lagrangian pattern search approach together with its convergence analysis. Ongoing work by Coope and Price along the lines of [12] and [13] promises alternatives for general constraints yet to be realized.

Acknowledgments. We wish to acknowledge a helpful referee and Major Mark Abramson, USAF, for many insightful comments that improved the presentation of this work.

REFERENCES

[1] C. Audet, *Convergence Results for Pattern Search Algorithms Are Tight*, Technical report TR98-24, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1998.

[2] C. Audet and J.E. Dennis, Jr., *Pattern search algorithms for mixed variable programming*, SIAM J. Optim., 11 (2000), pp. 573–594.

[3] C. Audet and J.E. Dennis, Jr., *A Pattern Search Filter Method for Nonlinear Programming without Derivatives*, Technical report TR00-09, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 2000.

[4] C. Audet, A.J. Booker, J.E. Dennis, Jr., P.D. Frank, and D. Moore, *A Surrogate-Model-Based Method For Constrained Optimization*, AIAA 2000-4891, in Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA, 2000.

[5] 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 Optim., 17 (1999), pp. 1–13.

[6] A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.W. Moore, and D.B. Serafini (1999), *Managing Surrogate Objectives to Optimize a Helicopter Rotor Design—Further Experiments*, AIAA Paper 98-4717, in Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St. Louis, MO, 1998.

[7] T.D. Choi, O.J. Eslinger, C.T. Kelley, J.W. David, and M. Etheridge, *Optimization of automotive valve train components with implicit filtering*, Optim. Engrg., 1 (2000), pp. 9–27.

[8] T.D. Choi and C.T. Kelley, *Superlinear convergence and implicit filtering*, SIAM J. Optim., 10 (1999), pp. 1149–1162.

[9] F.H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM Classics in Appl. Math. 5, SIAM, Philadelphia, 1990.

[10] A.R. Conn, N.I.M. Gould, and Ph.L. Toint, *A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM J. Numer. Anal., 28 (1991), pp. 545–572.

[11] I.D. Coope and C.J. Price, *On the convergence of grid-based methods for unconstrained optimization*, SIAM J. Optim., 11 (2001), pp. 859–869.

[12] I.D. Coope and C.J. Price, *Positive Bases in Optimization*, Report UCDMS2000/12, Department of Mathematics and Statistics, University of Canterbury, Canterbury, UK, 2000.

[13] I.D. Coope and C.J. Price, *Frames and Grids in Unconstrained and Linearly Constrained Optimization: A Non-Smooth Approach*, Report UCDMS2002/1, Department of Mathematics and Statistics, University of Canterbury, Canterbury, UK, 2002.

[14] J.E. Dennis and V. Torczon, *Direct search methods on parallel machines*, SIAM J. Optim., 1 (1991), pp. 448–474.

[15] E. Dolan, M. Lewis, and V. Torczon, *On the local convergence of pattern search*, ICASE Technical report 2000-36, NASA Langley Research Center, Hampton, VA, 2000.
[16] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Math. Program., 91 (2002), pp. 239–269.
[17] R. Fletcher, S. Leyffer, and Ph.L. Toint, *On the global convergence of a filter–SQP algorithm*, SIAM J. Optim., 13 (2002), pp. 44–59.
[18] R. Fletcher, N.I.M. Gould, S. Leyffer, Ph.L. Toint, and A. Wachter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM J. Optim., 13 (2002), pp. 635–659.
[19] P. Gilmore and C.T. Kelley, *An implicit filtering algorithm for optimization of functions with many local minima*, SIAM J. Optim., 5 (1995), pp. 269–285.
[20] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, New York, 1993.
[21] M. Kokkolaras, C. Audet, and J.E. Dennis, Jr., *Mixed variable optimization of the number and composition of heat intercepts in a thermal insulation system*, Optim. Engrg., 2 (2001), pp. 5–29.
[22] C-J. Lin and J.J. Moré, *Newton's method for large bound-constrained optimization problems*, SIAM J. Optim., 9 (1999), pp. 1100–1127.
[23] R.M. Lewis and V. Torczon, *Pattern search algorithms for bound constrained minimization*, SIAM J. Optim., 9 (1999), pp. 1082–1099.
[24] R.M. Lewis and V. Torczon, *Rank Ordering and Positive Basis in Pattern Search Algorithms*, ICASE Technical report TR 96-71, NASA Langley Research Center, Hampton, VA, 1996.
[25] R.M. Lewis and V. Torczon, *Pattern search methods for linearly constrained minimization*, SIAM J. Optim., 10 (2000), pp. 917–941.
[26] R.M. Lewis and V. Torczon, *A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, SIAM J. Optim., 12 (2002), pp. 1075–1089.
[27] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Ser. Oper. Res., Springer, NY, 1999.
[28] D. Serafini, *A Framework for Managing Models in Nonlinear Optimization of Computationally Expensive Functions*, Ph.D. Thesis, Rice University, Houston, TX, 1999.
[29] V. Torczon, *On the convergence of pattern search algorithms*, SIAM J. Optim., 7 (1997), pp. 1–25.