

## SECOND ORDER BEHAVIOR OF PATTERN SEARCH

MARK A. ABRAMSON \*

**Abstract.** Previous analyses of pattern search algorithms for unconstrained and linearly constrained minimization have focused on proving convergence of a subsequence of iterates to a limit point satisfying either directional or first-order necessary conditions for optimality, depending on the smoothness of the objective function in a neighborhood of the limit point. Even though pattern search methods require no derivative information, we are able to prove some limited directional second-order results. Although not as strong as classical second-order necessary conditions, these results are stronger than the first order conditions that many gradient-based methods satisfy. Under fairly mild conditions, we can eliminate from consideration all strict local maximizers and an entire class of saddle points.

**Key words.** nonlinear programming, pattern search algorithm, derivative-free optimization, convergence analysis, second order optimality conditions

**1. Introduction.** In this paper, we consider the class of generalized pattern search (GPS) algorithms applied to the linearly constrained optimization problem,

$$\min_{x \in X} f(x), \quad (1.1)$$

where the function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ , and  $X \subseteq \mathbb{R}^n$  is defined by a finite set of linear inequalities *i.e.*,  $X = \{x \in \mathbb{R}^n : Ax \geq b\}$ , where  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$ . We treat the unconstrained, bound constrained, and linearly constrained problems together because in these cases, we apply the algorithm, not to  $f$ , but to the “barrier” objective function  $f_X = f + \psi_X$ , where  $\psi_X$  is the indicator function for  $X$ ; *i.e.*, it is zero on  $X$ , and infinity elsewhere. If a point  $x$  is not in  $X$ , then we set  $f_X(x) = \infty$ , and  $f$  is not evaluated. This is important in many practical engineering problems in which  $f$  is expensive to evaluate.

The class of derivative-free pattern search algorithms was originally defined and analyzed by Torczon [27] for unconstrained optimization problems with a continuously differentiable objective function  $f$ . Torczon’s key result is the proof that there exists a subsequence of iterates that converges to a point  $x^*$  which satisfies the first-order necessary condition,  $\nabla f(x^*) = 0$ . Lewis and Torczon [20] add the valuable connection between pattern search methods and positive basis theory [16] (the details of which are ingrained into the description of the algorithm in Section 2). They extend the class to solve problems with bound constraints [21] and problems with a finite number of linear constraints [22], showing that if  $f$  is continuously differentiable, then a subsequence of iterates converges to a point satisfying the Karush-Kuhn-Tucker (KKT) first-order necessary conditions for optimality.

Audet and Dennis [7] add a hierarchy of convergence results for unconstrained and linearly constrained problems whose strength depends on the local smoothness of the objective function. They apply principles of the Clarke [12] nonsmooth calculus to show convergence to a point having nonnegative generalized directional derivatives in a set of directions that positively span the tangent cone there. They show convergence to a first-order stationary (or KKT) point under the weaker hypothesis of strict differentiability at the limit point, and illustrate how results of [21], [22], and [27] are corollaries of their own work.

---

\*Department of Mathematics and Statistics, Air Force Institute of Technology, AFIT/ENC, 2950 Hobson Way, Building 641, Wright Patterson AFB, Ohio, 45433-7765 (Mark.Abramson@afit.edu, <http://en.afit.edu/enc/Faculty/MAbramson/abramson.html>)

Audet and Dennis also extend GPS to categorical variables [6], which are discrete variables that cannot be treated by branch and bound techniques. This approach is successfully applied to engineering design problems in [2] and [19]. The theoretical results here can certainly be applied to these *mixed variable problems*, with the caveat that results would be with respect to the continuous variables (*i.e.*, while holding the categorical variables fixed). An adaptation of the results in [6] to more general objective functions using the Clarke [12] calculus can be found in [1].

The purpose of this paper is to provide insight into the second order behavior of the class of GPS algorithms for unconstrained and linearly constrained optimization. This may seem somewhat counterintuitive, in that, except for the approach described in [3], GPS methods do not even use first derivative information. However, the nature of GPS in evaluating the objective in multiple directions does, in fact, lend itself to some limited discovery of second order theorems, which are generally stronger than what can be proved for many gradient-based methods. Specifically, while we cannot ensure positive semi-definiteness of the Hessian matrix in all directions (and in fact, we show a few counter-examples), we can establish this result with respect to a certain subset of the directions, so that the likelihood of convergence to a point that is not a local minimizer is reasonably small.

This paper does not address the question of second order behavior of GPS algorithms for general nonlinear constraints. Extending convergence results of basic GPS to problems with nonlinear constraints requires augmentation to handle these constraints. Lewis and Torczon [23] do this by approximately solving a series of bound constrained augmented Lagrangian subproblems [14], while Audet and Dennis [9] use a filter-based approach [17]. The results presented here may be extendable to the former but not the latter, since the filter approach given in [9] cannot be guaranteed to converge to a first-order KKT point. The direct search algorithm of Lucidi *et al.* [24] applies positive basis theory to handle nonlinear constraints in a way similar to GPS, but it requires constraint derivatives and satisfaction of a sufficient decrease condition to ensure convergence, which [23] and [9] do not. Because of dissatisfaction with these limitations, Audet and Dennis [8] recently introduced the class of mesh-adaptive direct search (MADS) algorithms, a generalization of GPS that achieves first-order convergence for nonlinear constrained problems by generating a set of feasible directions that, in the limit, becomes asymptotically dense in the tangent cone. We plan to study second-order convergence properties of MADS in future work.

The remainder of this paper is organized as follows. In the next section, we briefly describe the basic GPS algorithm, followed by a review of known convergence results for basic GPS algorithms in Section 3. In Section 4, we show that, while convergence to a local maximizer is possible, it can only happen under some very strong assumptions on the both the objective function and the set of directions used by the algorithm. In Section 5, we introduce additional theorems to describe second order behavior of GPS more generally, along with a few examples to illustrate the theory and show that certain hypotheses cannot be relaxed. Section 6 offers some concluding remarks.

**Notation.**  $\mathbb{R}$ ,  $\mathbb{Q}$ ,  $\mathbb{Z}$ , and  $\mathbb{N}$  denote the set of real numbers, rational numbers, integers, and nonnegative integers, respectively. For any set  $S$ ,  $|S|$  denotes the cardinality of  $S$ , and  $-S$  is the set defined by  $-S = \{-s : s \in S\}$ . For any finite set  $A$ , we may also refer to the matrix  $A$  as the one whose columns are the elements of  $A$ . Similarly, for any matrix  $A$ , the notation  $a \in A$  means that  $a$  is a column of  $A$ . For  $x \in X$ , the *tangent cone* to  $X$  at  $x$  is  $T_X(x) = \text{cl}\{\mu(w - x) : \mu \geq 0, w \in X\}$ ,

and the *normal cone*  $N_X(x)$  to  $X$  at  $x$  is the polar of the tangent cone; namely,  $N_X(x) = \{v \in \mathbb{R}^n : v^T w \leq 0 \ \forall w \in T_X(x)\}$ . It is the nonnegative span of all outwardly pointing constraint normals at  $x$ .

**2. Generalized Pattern Search Algorithms.** For unconstrained and linearly constrained optimization problems, the basic GPS algorithm generates a sequence of iterates having nonincreasing function values. Each iteration consists of two main steps, an optional SEARCH phase and a local POLL phase, in which the barrier objective function  $f_X$  is evaluated at a finite number of points that lie on a mesh, with the goal of finding a point with lower objective function value, which is called an *improved mesh point*.

The mesh is not explicitly constructed; rather, it is conceptual. It is defined primarily through a set of positive spanning directions  $D$  in  $\mathbb{R}^n$ ; *i.e.*, where every vector in  $\mathbb{R}^n$  may be represented as a nonnegative linear combination of the elements of  $D$ . For convenience, we also view  $D$  as a real  $n \times n_D$  matrix whose  $n_D$  columns are its elements. The only other restriction on  $D$  is that it must be formed as the product

$$D = GZ, \quad (2.1)$$

where  $G \in \mathbb{R}^{n \times n}$  is a nonsingular real generating matrix, and  $Z \in \mathbb{Z}^{n \times n_D}$  is an integer matrix of full rank. In this way, each direction  $d_j \in D$  may be represented as  $d_j = Gz_j$ , where  $z_j \in \mathbb{Z}^n$  is an integer vector. At iteration  $k$ , the mesh is defined by the set

$$M_k = \bigcup_{x \in S_k} \{x + \Delta_k D z : z \in \mathbb{N}^{n_D}\}, \quad (2.2)$$

where  $S_k \in \mathbb{R}^n$  is the set of points where the objective function  $f$  had been evaluated by the start of iteration  $k$ , and  $\Delta_k > 0$  is the mesh size parameter that controls the fineness of the mesh. This construction is the same as that of [8] and [9], which generalizes the one given in [7]. It ensures that all previously computed iterates will lie on the current mesh.

The SEARCH step is simply an evaluation of a finite number of mesh points. It retains complete flexibility in choosing the mesh points, with the only caveat being that the points must be finite in number (including none). This could include a few iterations using a heuristic, such as a genetic algorithm, random sampling, etc., or, as is popular among many in industry (see [5, 10, 11, 25]), the approximate optimization on the mesh of a less expensive surrogate function. A related algorithm that does not require the surrogate solution to lie on the mesh (but requires additional assumptions for convergence) is found in [15].

If the SEARCH step fails to generate an improved mesh point, the POLL step is performed. This step is much more rigid in its construction, but this is necessary in order to prove convergence. The POLL step consists of evaluating  $f_X$  at points neighboring the current iterate  $x_k$  on the mesh. This set of points  $P_k$  is called the *poll set* and is defined by

$$P_k = \{x_k + \Delta_k d : d \in D_k \subseteq D\} \subset M_k, \quad (2.3)$$

where  $D_k$  is a positive spanning set of directions taken from  $D$ . We write  $D_k \subseteq D$  to mean that the columns of  $D_k$  are taken from the columns of  $D$ . Choosing a subset

$D_k \subset D$  of positive spanning directions at each iteration also adds the flexibility that will allow us to handle linear constraints in an efficient fashion.

If either the SEARCH or POLL step is successful in finding an improved mesh point, then the iteration ends immediately, with that point becoming the new iterate  $x_{k+1}$ . In this case, the mesh size parameter is either retained or increased (*i.e.*, the mesh is coarsened). If neither step finds an improved mesh point, then the point  $x_k$  is said to be a *mesh local optimizer* and is retained as the new iterate  $x_{k+1} = x_k$ , and the mesh size parameter is reduced (*i.e.*, the mesh is refined).

The rules that govern mesh coarsening and refining are as follows. For a fixed rational number  $\tau > 1$  and two fixed integers  $w^- \leq -1$  and  $w^+ \geq 0$ , the mesh size is updated according to the rule,

$$\Delta_{k+1} = \tau^{w_k} \Delta_k, \quad (2.4)$$

where  $w_k \in \{0, 1, \dots, w^+\}$  if the mesh is coarsened, or  $w_k \in \{w^-, w^- + 1, \dots, -1\}$  if the mesh is refined.

From (2.4), it follows that, for any  $k \geq 0$ , there exists an integer  $r_k$  such that

$$\Delta_{k+1} = \tau^{r_k} \Delta_0. \quad (2.5)$$

The basic GPS algorithm is given in Figure 2.1.

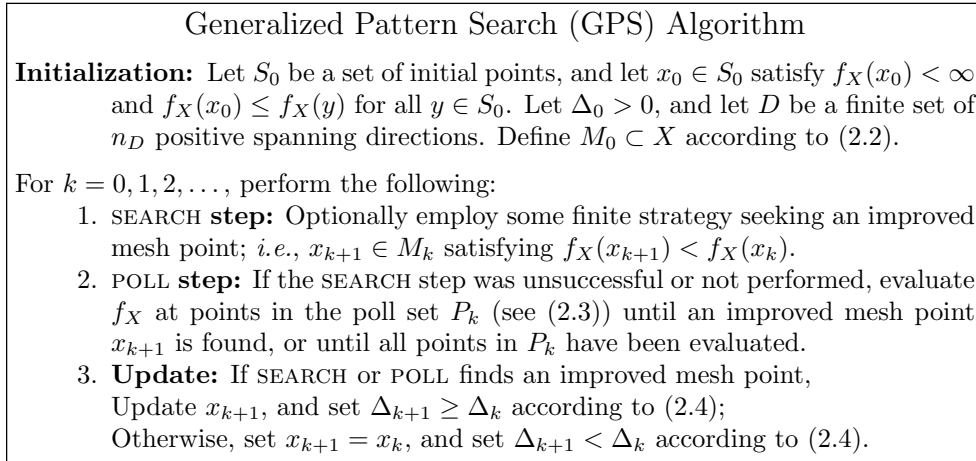


FIG. 2.1. *Basic GPS Algorithm*

With the addition of linear constraints, in order to retain first-order convergence properties, the set of directions  $D_k$  must be chosen to conform to the geometry of the constraints. The following definition, found in [7] (as an abstraction of the ideas and approach of [22]), gives a precise description for what is needed for convergence.

**DEFINITION 2.1.** *A rule for selecting the positive spanning sets  $D_k = D(k, x_k) \subseteq D$  conforms to  $X$  for some  $\epsilon > 0$ , if at each iteration  $k$  and for every boundary point  $y \in X$  satisfying  $\|y - x_k\| < \epsilon$ , the tangent cone  $T_X(y)$  is generated by nonnegative linear combinations of the columns of  $D_k$ .*

Using standard linear algebra tools, Lewis and Torczon [22] provide a clever algorithm to actually construct the sets  $D_k$ . If these sets are chosen so that they conform to  $X$ , all iterates lie in a compact set, and  $f$  is sufficiently smooth, then a subsequence of GPS iterates converges to a first-order stationary point [7, 22].

**3. Existing Convergence Results.** Before presenting new results, it is important to state what is currently known about the convergence properties of GPS for linearly constrained problems.

We first make the following assumptions:

- A1:** All iterates  $\{x_k\}$  produced by the GPS algorithm lie in a compact set.  
**A2:** The set of directions  $D = GZ$ , as defined in (2.1), includes tangent cone generators for every point in  $X$ .  
**A3:** The rule for selecting positive spanning sets  $D_k$  conforms to  $X$  for some  $\epsilon > 0$ .

Assumption A1, which is already sufficient to guarantee the existence of convergent subsequences of the iteration sequence, is a standard assumption [6, 7, 9, 14, 15, 17, 21, 22, 27]. A sufficient condition for this to hold is that the level set  $L(x_0) = \{x \in X : f(x) \leq f(x_0)\}$  is compact. We can assume that  $L(x_0)$  is bounded, but not closed, since we allow  $f$  to be discontinuous and extended valued. Thus we can assume that the closure of  $L(x_0)$  is compact. We should also note that most real engineering optimization problems have simple bounds on the design variables, which is enough to ensure that Assumption A1 is satisfied, since iterates lying outside of  $X$  are not evaluated by GPS. In the unconstrained case, note that Assumptions A2 and A3 are automatically satisfied by any positive spanning set constructed from the product in (2.1).

Assumption A2 is automatically satisfied if  $G = I$  and the constraint matrix  $A$  is rational, as is the case in [22]. Note that the finite number of linear constraints ensures that the set of tangent cone generators for all points in  $X$  is finite, which ensures that the finiteness of  $D$  is not violated.

If  $f$  is lower semi-continuous at any GPS limit point  $\bar{x}$ , then  $f(\bar{x}) \leq \lim_k f(x_k)$ , with equality if  $f$  is continuous [7]. Of particular interest are limit points of certain subsequences (indexed by some index set  $K$ ) for which  $\lim_{k \in K} \Delta_k = 0$ . We know that at least one such subsequence exists because of Torczon's [27] key result, restated here for convenience.

**THEOREM 3.1.** *The mesh size parameters satisfy  $\liminf_{k \rightarrow +\infty} \Delta_k = 0$ .*

From this result, we are interested in subsequences of iterates that converge to a limit point associated with  $\Delta_k$  converging to zero. The following definitions are due to Audet and Dennis [7].

**DEFINITION 3.2.** *A subsequence of GPS 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.*

**DEFINITION 3.3.** *Let  $\hat{x}$  be a limit point of a refining subsequence  $\{x_k\}_{k \in K}$ . A direction  $d \in D$  is said to be a refining direction of  $\hat{x}$  if  $x_k + \Delta_k d \in X$  and  $f(x_k) \leq f(x_k + \Delta_k d)$  for infinitely many  $k \in K$ .*

Audet and Dennis [6] prove the existence of at least one convergent refining subsequence. An important point is that, since a refining direction  $d$  is one in which  $x_k + \Delta_k d \in X$  infinitely often in the subsequence, it must be a feasible direction at the  $\hat{x}$ , and thus lies in the tangent cone  $T_X(\hat{x})$ .

The key results of Audet and Dennis are now given. The first shows directional optimality conditions under the assumption of Lipschitz continuity, and is obtained by a very short and elegant proof (see [7]) using Clarke's [12] definition of the generalized directional derivative. Audet [4] provides an example to show that Lipschitz continuity (and even differentiability) is not sufficient to ensure convergence to a Clarke

stationary point (*i.e.*, where zero belongs to the Clarke generalized gradient). The second result, along with its corollary for unconstrained problems, shows convergence to a point satisfying first-order necessary conditions for optimality. The latter two results were originally proved by Torczon [27] and Lewis and Torczon [21, 22] under the assumption of continuous differentiability of  $f$  on the level set containing all of the iterates. Audet and Dennis [7] prove the same results, stated here, requiring only strict differentiability at the limit point.

**THEOREM 3.4.** *Let  $\hat{x}$  be a limit of a refining subsequence, and let  $d \in D$  be any refining direction of  $\hat{x}$ . Under Assumptions A1–A3, if  $f$  is Lipschitz continuous near  $\hat{x}$ , then the generalized directional derivative of  $f$  at  $\hat{x}$  in the direction  $d$  is nonnegative, *i.e.*,  $f^\circ(\hat{x}; d) \geq 0$ .*

**THEOREM 3.5.** *Under Assumptions A1–A3, if  $f$  is strictly differentiable at a limit point  $\hat{x}$  of a refining subsequence, then  $\nabla f(\hat{x})^T w \geq 0$  for all  $w \in T_X(\hat{x})$ , and  $-\nabla f(\hat{x}) \in N_X(\hat{x})$ . Thus,  $\hat{x}$  satisfies the KKT first-order necessary conditions for optimality.*

**COROLLARY 3.6.** *Under Assumption A1, if  $f$  is strictly differentiable at a limit point  $\hat{x}$  of a refining subsequence, and if  $X = \mathbb{R}^n$  or  $\hat{x} \in \text{int}(X)$ , then  $\nabla f(\hat{x}) = 0$ .*

Although GPS is a derivative-free method, its strong dependence on the set of mesh directions presents some advantages in terms of convergence results. For example, if  $f$  is only Lipschitz continuous at certain limit points  $x^*$ , Theorem 3.4 provides a measure of directional optimality there in terms of the Clarke generalized directional derivatives being nonnegative [7]. In the next two sections, we attempt to prove certain second order optimality conditions, given sufficient smoothness of the objective function  $f$ . Our goal is to quantify our belief that convergence of GPS to a point that is not a local minimizer is very rare.

**4. GPS and Local Maximizers.** We treat the possibility of convergence to a local maximizer separate from other stationary points because what we can prove requires far less stringent assumptions. We begin with an example, provided by Charles Audet, to show that it is indeed possible to converge to a maximizer, even when  $f$  is smooth.

**EXAMPLE 4.1.** *Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the continuously differentiable function defined by*

$$f(x, y) = -x^2 y^2.$$

*Choose  $(x_0, y_0) = (0, 0)$  as the initial point, and set  $D = [e_1, e_2, -e_1, -e_2]$ , where  $e_1$  and  $e_2$  are the standard coordinate directions. Now observe that, if the SEARCH phase is empty, then the iteration sequence begins at the global maximizer  $(0, 0)$ , but can never move off of that point because the directions in  $D$  are lines of constant function value. Thus the sequence  $x_k$  converges to the global maximizer  $(0, 0)$ .*

Example 4.1 is clearly pathological. Had we started at *any* other point or polled in any other direction (that is not a scalar multiple of a coordinate direction), the algorithm would not have stalled at the maximizer  $(0, 0)$ . However, it is clear that the method of steepest descent and even Newton’s method would also fail to move away from this point.

From this example, one can envision other cases (also pathological), in which convergence to a local maximizer is possible, but without starting there. However, we can actually characterize these rare situations, in which convergence to a maximizer can occur. Lemma 4.2 shows that convergence would be achieved after only a finite number of iterations. Under a slightly stronger assumption, Theorem 4.3 ensures

that convergence to a maximizer means that every refining direction is a direction of constant function value. This very restrictive condition is consistent with Example 4.1. Corollary 4.4 establishes the key result that, under appropriate conditions, convergence to a strict local maximizer cannot occur. This result does not hold for gradient-based methods, even when applied to smooth functions.

**LEMMA 4.2.** *Let  $\hat{x}$  be the limit of a refining subsequence. If  $f$  is lower semi-continuous at  $\hat{x}$ , and if  $\hat{x}$  is a local maximizer of  $f$  in  $X$ , then  $x_k = \hat{x}$  is achieved in a finite number of iterations.*

*Proof.* Since  $\hat{x} = \lim_{k \in K} x_k$  is a local maximizer for  $f$  in  $X$ , there exists an open ball  $B(\hat{x}, \epsilon)$  of radius  $\epsilon$ , centered at  $\hat{x}$ , for some  $\epsilon > 0$ , such that  $f(\hat{x}) \geq f(y)$  for all  $y \in B(\hat{x}, \epsilon) \cap X$ . Then for all sufficiently large  $k \in K$ ,  $x_k \in B(\hat{x}, \epsilon) \cap X$ , and thus  $f(\hat{x}) \geq f(x_k)$ . But since GPS generates a nonincreasing sequence of function values, and since  $f$  is lower semi-continuous at  $\hat{x}$ , it follows that

$$f(x_k) \leq f(\hat{x}) \leq f(x_{k+1}) \leq f(x_k), \quad (4.1)$$

and thus  $f(x_k) = f(\hat{x})$ , for all sufficiently large  $k \in K$ . But since GPS iterates satisfy  $x_{k+1} \neq x_k$  only when  $f(x_{k+1}) < f(x_k)$ , it follows that  $x_k = \hat{x}$  for all sufficiently large  $k$ . ■

**THEOREM 4.3.** *Let  $\hat{x}$  be the limit of a refining subsequence. If  $f$  is lower semi-continuous in a neighborhood of  $\hat{x}$ , and if  $\hat{x}$  is a local maximizer of  $f$  in  $X$ , then every refining direction is a direction of constant function value.*

*Proof.* Let  $d \in D(\hat{x})$  be a refining direction. Since  $\hat{x}$  is a local maximizer, there exists  $\hat{\delta} > 0$  such that  $\hat{x} + td \in X$  and  $f(\hat{x}) \geq f(\hat{x} + td)$  for all  $t \in (0, \hat{\delta})$ . Now suppose that there exists  $\delta \in (0, \hat{\delta})$  such that  $f$  is continuous in  $B(\hat{x}, \delta)$  and  $f(\hat{x}) > f(\hat{x} + td)$  for all  $t \in (0, \delta)$ . Then  $f(\hat{x}) > f(\hat{x} + \Delta_k d)$  for  $\Delta_k \in (0, \delta)$ . But since Lemma 4.2 ensures convergence of GPS in a finite number of steps, we have the contradiction,  $f(\hat{x}) = f(x_k) \leq f(x_k + \Delta_k d) = f(\hat{x} + \Delta_k d)$  for all sufficiently large  $k$ . Therefore there must exist  $\delta > 0$  such that  $f(\hat{x}) = f(\hat{x} + td)$  for all  $t \in (0, \delta)$ . ■

**COROLLARY 4.4.** *The GPS algorithm cannot converge to any strict local maximizer of  $f$  at which  $f$  is lower semi-continuous.*

*Proof.* If  $\hat{x}$  is a strict local maximizer of  $f$  in  $X$ , then the first inequality of (4.1) is strict, yielding the contradiction,  $f(x_k) < f(\hat{x}) \leq f(x_k)$ . ■

The assumption that  $f$  is lower semi-continuous at  $\hat{x}$  is necessary for all three of these results to hold. As an example, consider the function  $f(x) = 1$  if  $x = 0$ , and  $x^2$  otherwise. This function has a strict local maximum at 0, and there are clearly no directions of constant function value. It is easy to see that any sequence of GPS iterates will converge to zero, and by choosing an appropriate starting point and mesh size, we can prevent convergence in a finite number of iterations. The theory is not violated because  $f$  is not lower semi-continuous there.

The additional assumption in Theorem 4.3 of lower semi-continuity in a neighborhood of the limit point (not just at the limit point) is needed to avoid other pathological examples, such as the function  $f(x) = 0$  if  $x \in \mathbb{Q}$  and  $-x^2$  if  $x \notin \mathbb{Q}$ . Continuity of  $f$  only holds at the local maximizer 0, and there are no directions of constant function value. A typical instance of GPS that uses rational arithmetic would stall at the starting point of 0.

**5. Second Order Theorems.** An interesting observation about Example 4.1 is that, even though  $(0, 0)$  is a local (and global) maximizer, the Hessian matrix is equal to the zero matrix there, meaning that it is actually positive semidefinite. This may

seem counterintuitive, but it is simply a case where the curvature of the function is described by Taylor series terms of higher than second order.

Thus an important question not yet answered is whether GPS can converge to a stationary point at which the Hessian is not positive semidefinite (given that the objective is twice continuously differentiable near the stationary point). The following simple example demonstrates that it is indeed possible, but once again, the algorithm does not move off of the starting point.

EXAMPLE 5.1. *Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the continuously differentiable function defined by*

$$f(x, y) = xy.$$

*Choose  $(x_0, y_0) = (0, 0)$  as the initial point, and set  $D = [e_1, e_2, -e_1, -e_2]$ , where  $e_1$  and  $e_2$  are the standard coordinate directions. Now observe that, if the SEARCH step is empty, then the iteration sequence begins at the saddle point  $(0, 0)$ , but can never move off of that point because the directions in  $D$  are lines of constant function value. Thus the sequence  $x_k$  converges to the saddle point. Furthermore, the Hessian of  $f$  at  $(0, 0)$  is given by*

$$\nabla^2 f(0, 0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

*which is indefinite, having eigenvalues of  $\pm 1$ .*

This result is actually not surprising, since many gradient-based methods have this same limitation. However, the results that follow provide conditions by which a pseudo-second order necessary condition is satisfied – one that is weaker than the traditional second order necessary condition, but stronger than the first-order condition that is all that can be guaranteed by most gradient-based methods.

We are now ready to present one of the main results of this paper. This will require the use of the Clarke [12] calculus in a manner similar to that of [7], but applied to  $f'$  instead of  $f$  itself. We will denote by  $f^{\circ\circ}(x; d_1, d_2)$ , the Clarke generalized directional derivative in the direction  $d_2$  of the directional derivative  $f'(x; d_1)$  of  $f$  at  $x$  in the fixed direction  $d_1$ . In other words, if  $g(x) = f'(x; d_1)$ , then  $f^{\circ\circ}(x; d_1, d_2) = g^{\circ}(x; d_2)$ . We should note that this is consistent with the concepts and notation given in [13] and [18]; however, we have endeavored to simplify the discussion for clarity. First, we give a general lemma that is independent of the GPS algorithm. The theorem and corollary that follow will be key to establishing a pseudo-second order result for GPS.

LEMMA 5.2. *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be continuously differentiable at  $x$ , and let  $f'(\cdot; \pm d)$  be Lipschitz near  $x$ . Then*

$$f^{\circ\circ}(x; d, d) = \limsup_{y \rightarrow x, t \downarrow 0} \frac{f(y + td) - 2f(y) + f(y - td)}{t^2}.$$

*Proof.* In general, we can apply the definition of the generalized directional deriva-



tive and the backward difference formula for directional derivatives to obtain

$$\begin{aligned}
f^{\circ\circ}(x; d, d) &= g^{\circ}(x; d) = \limsup_{y \rightarrow x, t \downarrow 0} \frac{g(y + td) - g(y)}{t} \\
&= \limsup_{y \rightarrow x, t \downarrow 0} \frac{f'(y + td; d) - f'(y; d)}{t} \\
&= \limsup_{y \rightarrow x, t \downarrow 0} \frac{1}{t} \left[ \lim_{s \rightarrow 0} \frac{f(y + td) - f(y + td - sd)}{s} - \lim_{s \rightarrow 0} \frac{f(y) - f(y - sd)}{s} \right] \\
&= \limsup_{y \rightarrow x, t \downarrow 0} \left[ \lim_{s \rightarrow 0} \frac{f(y + td) - f(y + (t - s)d) - f(y) + f(y - sd)}{ts} \right] \\
&= \limsup_{y \rightarrow x, t \downarrow 0} \left[ \frac{f(y + td) - 2f(y) + f(y - td)}{t^2} \right],
\end{aligned}$$

where the last equation follows from letting  $s$  approach zero as  $t$  does (which is allowable, since the limit as  $s \rightarrow 0$  exists and is independent of how it is approached).

THEOREM 5.3. *Let  $\hat{x}$  be the limit of a refining subsequence, and let  $D(\hat{x})$  be the set of refining directions for  $\hat{x}$ . Under Assumptions A1–A3, if  $f$  is continuously differentiable in a neighborhood of  $\hat{x}$ , then for every direction  $d \in D(\hat{x})$  such that  $\pm d \in D(\hat{x})$  and  $f'(\cdot; \pm d)$  is Lipschitz near  $\hat{x}$ ,  $f^{\circ\circ}(\hat{x}; d, d) \geq 0$ .*

*Proof.* From Lemma 5.2, it follows that

$$\begin{aligned}
f^{\circ\circ}(\hat{x}; d, d) &= \limsup_{y \rightarrow \hat{x}, t \downarrow 0} \frac{f(y + td) - 2f(y) + f(y - td)}{t^2} \\
&\geq \lim_{k \in K} \frac{f(x_k + \Delta_k d) - 2f(x_k) + f(x_k - \Delta_k d)}{\Delta_k^2} \\
&\geq 0,
\end{aligned}$$

since  $\pm d \in D(\hat{x})$  means that  $f(x_k) \leq f(x_k \pm \Delta_k d)$  for all  $k \in K$ . ■

COROLLARY 5.4. *Let  $\hat{x}$  be the limit of a refining subsequence, and let  $D(\hat{x})$  be the set of refining directions for  $\hat{x}$ . Under Assumptions A1–A3, if  $f$  is twice continuously differentiable at  $\hat{x}$ , then  $d^T \nabla^2 f(\hat{x}) d \geq 0$  for every direction  $d$  satisfying  $\pm d \in D(\hat{x})$ .*

*Proof.* This follows directly from Theorem 5.3 and the fact that, when  $\nabla^2 f(\hat{x})$  exists,  $d^T \nabla^2 f(\hat{x}) d = f^{\circ\circ}(x; d, d)$ . ■

The following example illustrates how a function can satisfy the hypotheses of Theorem 5.3, but not those of Corollary 5.4.

EXAMPLE 5.5. *Consider the strictly convex function  $f : \mathbb{R} \rightarrow \mathbb{R}$  defined by*

$$f(x) = \begin{cases} x^2, & \text{if } x \geq 0, \\ -x^3, & \text{if } x < 0. \end{cases}$$

*GPS will converge to the global minimizer at  $x = 0$  from any starting point. The derivative of  $f$  is given by*

$$f'(x) = \begin{cases} 2x, & \text{if } x \geq 0, \\ -3x^2, & \text{if } x < 0. \end{cases}$$

*Clearly,  $f'$  is (Lipschitz) continuous at all  $x \in \mathbb{R}$ , satisfying the hypotheses of Theorem 5.3. The second derivative of  $f$  is given by*

$$f''(x) = \begin{cases} 2, & \text{if } x > 0, \\ -6x, & \text{if } x < 0, \end{cases}$$

and it does not exist at  $x = 0$ . Thus, the hypotheses of Corollary 5.4 are violated. The conclusion of Theorem 5.3 can be verified by examining the Clarke derivatives of  $f'$  at  $x = 0$ :

$$\begin{aligned} f^{\circ\circ}(0; d, d) &= \limsup_{y \rightarrow 0, t \downarrow 0} \frac{f(y + td) - f(y) + f(y - td)}{t^2} \\ &\geq \limsup_{y \rightarrow 0, t \downarrow 0} \frac{(y + td)^2 - (2y)^2 + (y - td)^2}{t^2} \\ &= \limsup_{y \rightarrow 0, t \downarrow 0} \frac{y^2 + 2ytd + t^2d^2 - 2y^2 + y^2 - 2ytd + t^2d^2}{t^2} \\ &= 2d^2 \geq 0. \end{aligned}$$

**5.1. Results for Unconstrained Problems.** For unconstrained problems, recall that if  $f$  is twice continuously differentiable at a stationary point  $x^*$ , the second order necessary condition for optimality is that  $\nabla^2 f(x^*)$  is positive semi-definite; that is,  $v^T \nabla^2 f(x^*) v \geq 0$  for all  $v \in \mathbb{R}^n$ . The following definition gives a pseudo-second order necessary condition that is not as strong as the traditional one.

**DEFINITION 5.6.** *Suppose that  $x^*$  is a stationary point of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  that is twice continuously differentiable at  $x^*$ . Then  $f$  is said to satisfy a pseudo-second order necessary condition at  $x$  for an orthonormal basis  $V \subset \mathbb{R}^n$  if*

$$v^T \nabla^2 f(x^*) v \geq 0 \quad \forall v \in V. \quad (5.1)$$

We note that (5.1) holds for  $-V$  as well; therefore, satisfying this condition means that it holds for a set of “evenly distributed” vectors in  $\mathbb{R}^n$ .

Now recall that a symmetric matrix is positive semidefinite if and only if it has nonnegative real eigenvalues. The following theorem gives an analogous result for matrices that are positive semidefinite with respect to only an orthonormal basis. We note that this general linear algebra result is independent of the convergence results presented in this paper.

**THEOREM 5.7.** *Let  $B \in \mathbb{R}^{n \times n}$  be symmetric, and let  $V$  be an orthonormal basis for  $\mathbb{R}^n$ . If  $B$  satisfies  $v^T B v \geq 0$  for all  $v \in V$ , then the sum of its eigenvalues is nonnegative. If  $B$  also satisfies  $v^T B v > 0$  for at least one  $v \in V$ , then this sum is positive.*

*Proof.* Since  $B$  is symmetric, its Schur decomposition can be expressed as  $B = Q \Lambda Q^T$ , where  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  and  $Q \in \mathbb{R}^{n \times n}$  is an orthogonal matrix whose columns  $q_i, i = 1, 2, \dots, n$ , are the orthonormal eigenvectors corresponding to the real eigenvalues  $\lambda_i, i = 1, 2, \dots, n$ . Then for each  $v_i \in V, i = 1, 2, \dots, n$ ,

$$0 \leq v_i^T B v_i = v_i^T Q \Lambda Q^T v_i = \sum_{j=1}^n \lambda_j (Q^T v_i)_j^2 = \sum_{j=1}^n \lambda_j (q_j^T v_i)^2, \quad (5.2)$$

and, since  $\{q_j\}_{j=1}^n$  and  $\{v_i\}_{i=1}^n$  are both orthonormal bases for  $\mathbb{R}^n$ , it follows that

$$0 \leq \sum_{i=1}^n v_i^T B v_i = \sum_{i=1}^n \sum_{j=1}^n \lambda_j (q_j^T v_i)^2 = \sum_{j=1}^n \lambda_j \sum_{i=1}^n (v_i^T q_j)^2 = \sum_{j=1}^n \lambda_j \|q_j\|_2^2 = \sum_{j=1}^n \lambda_j. \quad (5.3)$$

To obtain the final result, observe that making just one of the inequalities in (5.2) strict yields a similar strict inequality in (5.3), and the result is proved. ■

It is easy to see from this proof, that if  $V$  happens to be the set of eigenvectors  $Q$  of  $B$ , then  $B$  is positive (semi-) definite, since in this case, (5.2) yields  $q_j^T v_i = q_j^T q_i = \delta_{ij}$ , which means that  $\lambda_i \geq (>)0$ .

We now establish pseudo-second order results for GPS by the following two theorems. The first theorem requires convergence in a finite number of steps, while the second necessitates the use of a more specific set of positive spanning directions.

**THEOREM 5.8.** *Let  $V$  be an orthonormal basis in  $\mathbb{R}^n$ . Let  $\hat{x}$  be the limit of a refining subsequence, and let  $D(\hat{x})$  be the set of refining directions for  $\hat{x}$ . Under Assumption A1, if  $f$  is twice continuously differentiable at  $\hat{x}$ ,  $D_k \supseteq V$  infinitely often in the subsequence, and  $x_k = \hat{x}$  for all sufficiently large  $k$ , then  $f$  satisfies a pseudo-second order necessary condition for  $V$  at  $\hat{x}$ .*

*Proof.* For all  $k \in K$  and  $d \in D(\hat{x})$ , we have  $f(x_k + \Delta_k d) \geq f(x_k)$ . Furthermore, for all sufficiently large  $k \in K$ , since  $x_k = \hat{x}$ , a simple substitution yields  $f(\hat{x} + \Delta_k d) \geq f(\hat{x})$  for all  $d \in D(\hat{x})$ . For each  $d \in D(\hat{x})$ , Taylor's Theorem yields

$$f(\hat{x} + \Delta_k d) = f(\hat{x}) + \Delta_k d^T \nabla f(\hat{x}) + \frac{1}{2} \Delta_k^2 d^T \nabla^2 f(\hat{x}) d + \mathcal{O}(\Delta_k^3).$$

Since Corollary 3.6 ensures that  $\nabla f(\hat{x}) = 0$ , we have

$$0 \leq f(\hat{x} + \Delta_k d) - f(\hat{x}) = \frac{1}{2} \Delta_k^2 d^T \nabla^2 f(\hat{x}) d + \mathcal{O}(\Delta_k^3),$$

or  $d^T \nabla^2 f(\hat{x}) d \geq \mathcal{O}(\Delta_k)$  for all  $d \in D(\hat{x})$  and for all sufficiently large  $k \in K$ . The result is obtained by taking limits of both sides (in  $K$ ) and noting that  $D(\hat{x})$  must contain  $V$ . ■

**THEOREM 5.9.** *Let  $V$  be an orthonormal basis in  $\mathbb{R}^n$ . Let  $\hat{x}$  be the limit of a refining subsequence, and let  $D(\hat{x})$  be the set of refining directions for  $\hat{x}$ . Under Assumption A1, if  $f$  is twice continuously differentiable at  $\hat{x}$  and  $D_k \supseteq V \cup -V$  infinitely often in the subsequence, then  $f$  satisfies a pseudo-second order necessary condition for  $V$  at  $\hat{x}$ . Furthermore, the sum of the eigenvalues of  $\nabla^2 f(\hat{x})$  must be nonnegative.*

*Proof.* Since  $D(\hat{x}) \subset D$  is finite, it must contain  $V \cup -V$ , and the result follows directly from Corollary 5.4 and Definition 5.6. The final result follows directly from the symmetry of  $\nabla^2 f(\hat{x})$  and Theorem 5.7. ■

The significance of Theorem 5.9 is that, if  $f$  is sufficiently smooth, then the choice of orthonormal mesh directions at each iteration will ensure that the pseudo-second order necessary condition is satisfied, and that the sum of the eigenvalues of  $\nabla^2 f(\hat{x})$  will be nonnegative. Thus, under the assumptions, GPS cannot converge to any saddle point whose Hessian has eigenvalues that sum to less than zero.

These saddle points (to which GPS cannot converge) are those which have sufficiently large regions (cones) of negative curvature. To see this, consider the contrapositive of Theorem 5.7 applied to the Hessian at the limit point; namely, if the sum of the eigenvalues of  $\nabla^2 f(\hat{x})$  is negative, then for *any* orthonormal basis  $V \in \mathbb{R}^n$ , at least one vector  $v \in V$  must lie in a cone of negative curvature (*i.e.*,  $v^T \nabla^2 f(\hat{x}) v < 0$ ). Since the angle between any two of these orthogonal directions is 90 degrees, there must be a cone of negative curvature with an angle greater than 90 degrees.

The following example shows that, even for orthonormal mesh directions, it is still possible to converge to a saddle point – even when not starting there. It also illustrates our assertion about cones of negative curvature.

EXAMPLE 5.10. Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the twice continuously differentiable function defined by

$$f(x, y) = 99x^2 - 20xy + y^2 = (9x - y)(11x - y). \quad (5.4)$$

Choose  $(x_0, y_0) = (1, 1)$  as the initial point, and set  $D = \{e_1, e_2, -e_1, -e_2\}$ , where  $e_1$  and  $e_2$  are the standard coordinate directions. Now observe that, at the saddle point  $(0, 0)$ , directions of negative curvature lie only in between the lines  $y = 9x$  and  $y = 11x$ . Thus, to avoid the saddle point, the GPS sequence would have to include a point inside the narrow cone formed by these two lines, when sufficiently close to the origin. If the SEARCH step is empty, and the polling directions in  $D$  are chosen consecutively in the POLL step (i.e., we poll in the order  $e_1, e_2, -e_1, -e_2$ ), then the iteration sequence arrives exactly at the origin after 10 iterations and remains there because none of the directions in  $D$  point inside of a cone of negative curvature. Figure 5.1 shows the cones of negative curvature for  $f$  near the saddle point. Note that these cones, depicted in the shaded areas, are very narrow compared to those of positive curvature. Thus, for the search directions in  $D$ , it will be difficult to yield a trial point inside one of these cones.

On the other hand, if our objective function were  $-f$ , then the cones of negative curvature would be depicted by the non-shaded areas. In this case, Theorem 5.7 ensures that GPS cannot converge to the saddle point, since any set of  $2n$  orthonormal directions would generate a trial point inside one of these cones, and thus a lower function value than that of the saddle point.

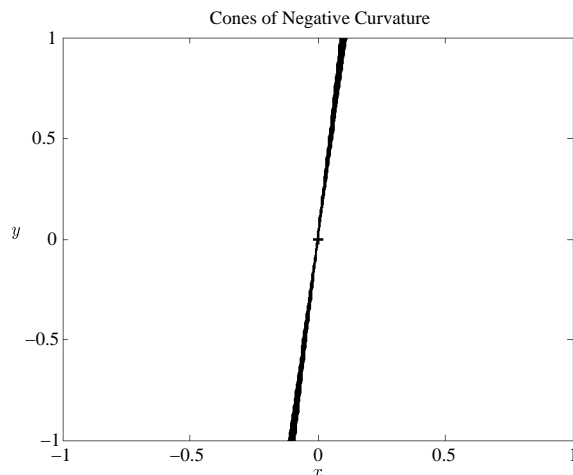


FIG. 5.1. For  $f(x, y) = (9x - y)(11x - y)$ , the cones of negative curvature at the saddle point  $(0, 0)$  are shown in the shaded area between the lines  $y = 9x$  and  $y = 11x$ .

**5.2. Results for Linearly Constrained Problems.** We now treat the linear constrained problem given in (1.1). At this point, we note that there are two equivalent formulations for the classical Karush-Kuhn-Tucker (KKT) first-order necessary conditions for optimality, one of which is imbedded in Theorem 3.5. It states that a point  $x^*$  satisfies the first-order necessary conditions if  $\nabla f(x^*)^T w \geq 0$  for all directions  $w$  in the tangent cone  $T_X(x^*)$  at  $x^*$ , and  $-\nabla f(x^*)$  lies in the normal cone  $N_X(x^*)$  at  $x^*$ . However, since we do not have such a straightforward description of

a second order necessary condition in this form, we now give a more traditional form of the KKT necessary conditions, from which we will be able to establish a sensible pseudo-second order condition. The following lemma, given without proof, is taken from a well-known textbook [26].

LEMMA 5.11. *If  $x^*$  is a local solution of (1.1), then for some vector  $\lambda$  of Lagrange multipliers,*

1.  $\nabla f(x^*) = A^T \lambda$ , or equivalently,  $W^T \nabla f(x^*) = 0$ ,
2.  $\lambda \geq 0$ ,
3.  $\lambda^T (Ax^* - b) = 0$ ,
4.  $W^T \nabla^2 f(x^*) W$  is positive semi-definite,

where the columns of  $W$  form a basis for the null-space of the active constraints at  $x^*$ .

The first three conditions of Lemma 5.11 are generally referred to as first-order necessary conditions, while the last is the second order necessary condition. Convergence of a subsequence of GPS iterates to a point satisfying first-order conditions has been proved previously [7, 22] and is summarized in Theorem 3.5. Based on the second order condition, we now provide a pseudo-second order necessary condition for linearly constrained problems that is analogous to that given in Definition 5.6 for unconstrained problems.

DEFINITION 5.12. *For the optimization problem given in (1.1), let  $W$  be an orthonormal basis for the null space of the binding constraints at  $x^*$ , where  $x^*$  satisfies the KKT first-order necessary optimality conditions, and  $f$  is twice continuously differentiable at  $x^*$ . Then  $f$  is said to satisfy a pseudo-second order necessary condition for  $W$  at  $x^*$  if*

$$w^T \nabla^2 f(x^*) w \geq 0 \quad \forall w \in W. \quad (5.5)$$

The following theorem shows that the condition given in (5.5) has an equivalent reduced Hessian formulation similar to Definition 5.6. It is formulated to be a general linear algebra result, independent of the GPS algorithm.

THEOREM 5.13. *Let  $B \in \mathbb{R}^{n \times n}$  be symmetric, and let  $W \in \mathbb{R}^{n \times p}$  be a matrix with orthonormal columns  $\{w_i\}_{i=1}^p$ , where  $p \leq n$ . Then the following two statements are equivalent.*

1.  $w_i^T B w_i \geq 0$ ,  $i = 1, 2, \dots, p$ ,
2. *There exists a matrix  $Y$  whose columns  $\{y_i\}_{i=1}^p$  form an orthonormal basis for  $\mathbb{R}^p$  such that  $y_j^T W^T B W y_j \geq 0$ ,  $j = 1, 2, \dots, p$ .*

*Proof.* Suppose  $w_i^T B w_i \geq 0$ ,  $i = 1, 2, \dots, p$ . Then  $e_i^T B e_i \geq 0$ ,  $i = 1, 2, \dots, p$ , and the result holds since  $\{e_i\}_{i=1}^p$  are orthonormal.

Conversely, suppose there exists  $Y \in \mathbb{R}^{p \times p}$  such that  $y_j^T W^T B W y_j \geq 0$ ,  $j = 1, 2, \dots, p$ . Let  $Z = WY$  with columns  $\{z_i\}_{i=1}^p$ . Then for  $i = 1, 2, \dots, p$ , we have  $z_i^T B z_i = y_i^T W^T B W y_i \geq 0$ . Furthermore, the columns of  $Z$  are orthonormal, since  $z_i^T z_j = (W y_i)^T (W y_j) = y_i^T W^T W y_j = y_i^T y_j = \delta_{ij}$  (the last step by the orthogonality of  $Y$ ). ■

Assumptions A2–A3 ensure that the GPS algorithm chooses directions that conform to  $X$  [7, 21, 22]. This means that the finite set  $T$  of all tangent cone generators for all points  $x \in X$  must be a subset of  $D$ , and that if an iterate  $x_k$  is within  $\epsilon > 0$  of a constraint boundary, then certain directions in  $T$  must be included in  $D_k$ . An algorithm that identifies these directions  $T_k \subseteq T$ , in the non-degenerate case, is given in [22], where it is noted that  $T_k$  is chosen so as to contain a (positive) basis for

the null-space of the  $\epsilon$ -active constraints at  $x_k$ . Thus, the set of refining directions  $D(\hat{x})$  will always contain tangent cone generators at  $\hat{x}$ , a subset of which forms a basis for null-space of the active constraints at  $\hat{x}$ . We will denote this null-space by  $\mathcal{N}(\hat{A})$ , where  $\hat{A}$  is the matrix obtained by deleting the rows of  $A$  corresponding to the non-active constraints at  $\hat{x}$ .

However, in order to exploit the theory presented here we require the following additional assumption so that  $D(\hat{x})$  will always contain an orthonormal basis for  $\mathcal{N}(\hat{A})$ .

**A4:** The algorithm that computes the tangent cone generators at each iteration includes an orthonormal basis for the null-space of the  $\epsilon$ -active constraints at each iterate.

Furthermore, since  $\mathcal{N}(\hat{A})$  contains the negative of any vector in the space, we can prove convergence to a point satisfying a pseudo-second order necessary condition by including  $T_k \cup -T_k$  in each set of directions  $D_k$ .

The next theorem establishes convergence of a subsequence of GPS iterates to a point satisfying a pseudo-second order necessary condition, similar to that of Theorem 5.8 under the fairly strong condition that convergence occurs in a finite number of steps.

**THEOREM 5.14.** *Let  $V$  be an orthonormal basis for  $\mathbb{R}^n$ . Let  $\hat{x}$  be the limit of a refining subsequence, and let  $D(\hat{x})$  be the set of refining directions for  $\hat{x}$ . Under Assumptions A1–A4, if  $f$  is twice continuously differentiable at  $\hat{x}$ , and for all sufficiently large  $k$ ,  $D_k \subset V$  and  $x_k = \hat{x}$ , then  $f$  satisfies a pseudo-second order necessary condition for  $V$  at  $\hat{x}$ .*

*Proof.* For all  $k \in K$  and  $d \in D(\hat{x})$ , we have  $f(x_k + \Delta_k d) \geq f(x_k)$ . Furthermore, for all sufficiently large  $k \in K$ , since  $x_k = \hat{x}$ , a simple substitution yields  $f(\hat{x} + \Delta_k d) \geq f(\hat{x})$  for all  $d \in D(\hat{x})$ . For each  $d \in D(\hat{x})$ , Taylor's Theorem yields

$$f(\hat{x} + \Delta_k d) = f(\hat{x}) + \Delta_k d^T \nabla f(\hat{x}) + \frac{1}{2} \Delta_k^2 d^T \nabla^2 f(\hat{x}) d + \mathcal{O}(\Delta_k^3).$$

For  $d \in \mathcal{N}(\hat{A})$ , Lemma 5.11 ensures that  $d^T \nabla f(\hat{x}) = 0$ , and thus

$$0 \leq f(\hat{x} + \Delta_k d) - f(\hat{x}) = \frac{1}{2} \Delta_k^2 d^T \nabla^2 f(\hat{x}) d + \mathcal{O}(\Delta_k^3),$$

or  $d^T \nabla^2 f(\hat{x}) d \geq \mathcal{O}(\Delta_k)$  for all  $d \in D(\hat{x}) \cap \mathcal{N}(\hat{A})$  and for all sufficiently large  $k \in K$ . The result is obtained by taking limits of both sides (in  $K$ ), since  $D(\hat{x})$  must contain an orthonormal basis for  $\mathcal{N}(\hat{A})$ .  $\blacksquare$

In the theorem that follows, we show that, given sufficient smoothness of  $f$ , if mesh directions are chosen in a fairly standard way, a subsequence of GPS iterates converges to a point satisfying a pseudo-second order necessary condition. The theorem is similar to Theorem 5.9. Once again, the corollary to this theorem identifies an entire class of saddle points to which GPS cannot converge.

**THEOREM 5.15.** *Let  $V$  be an orthonormal basis for  $\mathbb{R}^n$ . Let  $\hat{x}$  be the limit of a refining subsequence, and let  $D(\hat{x})$  be the set of refining directions for  $\hat{x}$ . Under Assumptions A1–A4, if  $f$  is twice continuously differentiable at  $\hat{x}$  and  $D_k \supseteq V \cup -V \cup T_k \cup -T_k$  infinitely often in the subsequence, then  $f$  satisfies a pseudo-second order necessary condition for  $V$  at  $\hat{x}$ .*

*Proof.* From the discussion following Theorem 5.13,  $D(\hat{x})$  contains an orthonormal basis  $W$  for  $\mathcal{N}(\hat{A})$ . Since  $D$  is finite, for infinitely many  $k$ , we have  $-W \subseteq -T_k \subset D_k$ ,

which means that  $-W \subseteq D(\hat{x})$ . Thus,  $D(\hat{x}) \supseteq W \cup -W$ , where  $W$  is an orthonormal basis for  $\mathcal{N}(\hat{A})$ , and the result follows from Corollary 5.4 and Definition 5.12. ■

**COROLLARY 5.16.** *If hypotheses of Theorem 5.15 hold, then the sum of the eigenvalues of the reduced Hessian  $W^T \nabla^2 f(\hat{x}) W$  is nonnegative, where the columns of  $W$  form a basis for the null space of the active constraints at  $\hat{x}$ .*

*Proof.* Theorem 5.15 ensures that the pseudo-second order condition holds; *i.e.*,  $w^T \nabla^2 f(\hat{x}) w \geq 0$  for all  $w \in W$ . Then for  $i = 1, 2, \dots, |W|$ ,  $e_i^T W^T \nabla^2 f(\hat{x}) W e_i \geq 0$ , where  $e_i$  denotes the  $i^{\text{th}}$  coordinate vector in  $\mathbb{R}^{|W|}$ . Since  $W^T \nabla^2 f(\hat{x}) W$  is symmetric and  $\{e_i\}_{i=1}^{|W|}$  forms an orthonormal basis for  $\mathbb{R}^{|W|}$ , the result follows from Theorem 5.7. ■

**6. Concluding Remarks.** Clearly, the class of GPS algorithms can never be guaranteed to converge to a point satisfying classical second order necessary conditions for optimality. However, we have been able to show the following important results, which are surprisingly stronger than what has been proved for many gradient-based (and some Newton-based) methods:

- Under mild assumptions, GPS can only converge to a local maximizer if it does so in a finite number of steps, and if all the directions used infinitely often are directions of constant function value at the maximizer (Lemma 4.2, Theorem 4.3).
- Under mild assumptions, GPS cannot converge to or stall at a strict local maximizer (Corollary 4.4).
- If  $f$  is sufficiently smooth and mesh directions contain an orthonormal basis and its negatives, then a subsequence of GPS iterates converges to a point satisfying a pseudo-second order necessary condition for optimality (Theorems 5.9 and 5.15).
- If  $f$  is sufficiently smooth and mesh directions contain an orthonormal basis and its negatives, then GPS cannot converge to a saddle point at which the sum of the eigenvalues of the Hessian (or reduced Hessian) are negative (Theorem 5.9 and Corollary 5.16).

Thus an important characteristic of GPS is that, given reasonable assumptions, the likelihood of converging to a point that does not satisfy second order necessary conditions is small.

**Acknowledgements.** The author wishes to thank John Dennis and Charles Audet for several suggestions and useful discussions, along with two anonymous referees for their suggestions, all of which have improved the presentation.

The views expressed in this paper are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or United States Government.

#### REFERENCES

- [1] M. A. ABRAMSON, *Pattern Search Algorithms for Mixed Variable General Constrained Optimization Problems*, PhD thesis, Rice University, Department of Computational and Applied Mathematics (CAAM), Houston, TX, August 2002. Also available as CAAM Technical Report TR02-11.
- [2] ———, *Mixed variable optimization of a load-bearing thermal insulation system using a filter pattern search algorithm*, *Optimization and Engineering*, 5 (2004), pp. 157–177.
- [3] M. A. ABRAMSON, C. AUDET, AND J. E. DENNIS, JR., *Generalized pattern searches with derivative information*, *Mathematical Programming*, 100 (2004).
- [4] C. AUDET, *Convergence results for pattern search algorithms are tight*, *Optimization and Engineering*, 5 (2004), pp. 101–122.

- [5] C. AUDET, A. J. BOOKER, J. E. DENNIS, JR., P. D. FRANK, AND D. W. MOORE, *A surrogate-model-based method for constrained optimization*, in AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, vol. 2000-4891, AIAA, 2000.
- [6] C. AUDET AND J. E. DENNIS, JR., *Pattern search algorithms for mixed variable programming*, SIAM J. Optim., 11 (2001), pp. 573–594.
- [7] ———, *Analysis of generalized pattern searches*, SIAM J. Optim., 13 (2003), pp. 889–903.
- [8] ———, *Mesh adaptive direct search algorithms for constrained optimization*, Tech. Report TR04-02, Department of Computational and Applied Mathematics, Rice University, Houston Texas, 2004.
- [9] ———, *A pattern search filter method for nonlinear programming without derivatives*, SIAM J. Optim., 14 (2004), pp. 980–1010.
- [10] A. BOOKER, J. E. DENNIS, JR., P. FRANK, D. SERAFINI, AND V. TORCZON, *Optimization using surrogate objectives on a helicopter test example*, in Optimal Design, J. Burns and E. Cliff, eds., Philadelphia, 1998, SIAM.
- [11] 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 function by surrogates*, Structural Optimization, 17 (1999), pp. 1–13.
- [12] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, SIAM Classics in Applied Mathematics, 5, SIAM, 1990.
- [13] R. COMINETTI AND R. CORREA, *A generalized second-order derivative in nonsmooth optimization*, SIAM J. Control Optim., 28 (1990), pp. 789–809.
- [14] A. R. CONN, N. I. M. GOULD, AND P. 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.
- [15] 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.
- [16] C. DAVIS, *Theory of positive linear dependence*, American Journal of Mathematics, 76 (1954), pp. 733–746.
- [17] R. FLETCHER AND S. LEYFFER, *Nonlinear programming without a penalty function*, Mathematical Programming, 91 (2002), pp. 239–269.
- [18] J.-B. HIRIART-URRUTY, J.-J. STRODIOT, AND V. H. NGUYEN, *Generalized Hessian matrix and second-order optimality conditions for problems with  $C^{1,1}$  data*, Appl. Math. Optim., 11 (1984), pp. 43–56.
- [19] 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*, Optimization and Engineering, 2 (2001), pp. 5–29.
- [20] R. M. LEWIS AND V. TORCZON, *Rank ordering and positive basis in pattern search algorithms*, Tech. Report TR 96-71, ICASE NASA Langley Research Center, 1996.
- [21] ———, *Pattern search algorithms for bound constrained minimization*, SIAM J. Optim., 9 (1999), pp. 1082–1099.
- [22] ———, *Pattern search algorithms for linearly constrained minimization*, SIAM J. Optim., 10 (2000), pp. 917–941.
- [23] ———, *A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, SIAM J. Optim., 12 (2002), pp. 1075–1089.
- [24] S. LUCIDI, M. SCIANDRONE, AND P. TSENG, *Objective-derivative-free methods for constrained optimization*, Mathematical Programming, 92 (2002), pp. 37–59.
- [25] A. L. MARSDEN, M. WANG, J. E. DENNIS, JR., AND P. MOIN, *Optimal aeroelastic shape design using the surrogate management framework*, Optimization and Engineering, 5 (2004), pp. 235–262.
- [26] S. G. NASH AND A. SOFER, *Linear and Nonlinear Programming*, McGraw-Hill, New York, 1996.
- [27] V. TORCZON, *On the convergence of pattern search algorithms*, SIAM J. Optim., 7 (1997), pp. 1–25.