

ON THE CONVERGENCE OF GRID-BASED METHODS FOR UNCONSTRAINED OPTIMIZATION

I.D. COOPE AND C.J. PRICE *

Abstract. The convergence of direct search methods for unconstrained minimization is examined in the case where the underlying method can be interpreted as a grid or pattern search over successively refined meshes. An important aspect of the main convergence result is that translation, rotation, scaling and shearing of the successive grids are allowed.

Key words. Grid-based optimization, derivative free optimization, positive basis methods, convergence analysis, multidirectional search.

AMS subject classifications. 49M30, 65K05

1. Introduction. Recent survey papers, [1], [7], [10] report on significant renewed interest in algorithms for derivative-free unconstrained optimization. Much of this recent interest has been provoked by new convergence results, (see, for example, [1], [6], [8], [9]). Most of the current derivative-free algorithms for which convergence results have been established belong to one or more of three categories: line search methods, trust region methods or grid-based methods. In this paper, the convergence of derivative-free methods for unconstrained minimization is examined in the case where the underlying method can be interpreted as a grid or pattern search over successively refined meshes. Therefore, the methods discussed here are similar to those studied in [6], [8], [9], but permit greater freedom in the orientation and scaling of successive grids. Alternative approaches based on trust-regions or line searches can be found in [1], [7] and the references therein.

The properties of grid-based methods are explored and it is shown that convergence can be achieved for a quite general class of algorithm. An important aspect of the main convergence result is that successive grids may be arbitrarily translated, rotated, and sheared relative to one another, and each grid axis may be re-scaled independently of the others. This flexibility allows second-order information to be incorporated into the shape of successive grids, for example by aligning grid axes along conjugate directions or the principal axes of an approximating quadratic. The hope is to construct non-derivative algorithms that possess useful properties of conjugate direction or quasi-Newton algorithms, thus exploiting curvature information without assuming the existence of second derivatives or the availability of first derivatives.

We present two optimization frameworks for unconstrained optimization of continuously differentiable functions that are bounded below. For the first framework, in which finite searches are conducted along grid directions of descent, we establish convergence of a subsequence of iterates to a stationary point of the objective function. For the second framework, under the stronger assumption that the algorithm searches the grid direction of locally greatest descent at every iterate, we show that the entire sequence of iterates converges to a stationary point.

The restrictions on the grids in our framework are much less severe than for the pattern search methods of [6, 9] where a single set of grid axes is used, only rational scalings of grids are permitted, and arbitrary translations are not allowed.

*Department of Mathematics & Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. (i.coope@math.canterbury.ac.nz, c.price@math.canterbury.ac.nz).

Furthermore, the methods of [6, 9] do not allow scalings and realignments to reflect curvature information. A more complete discussion of differences from the pattern search methods of [6, 9] appears in Section 3.

The non-derivative method of [4] for bound-constrained optimization includes curvature information through interpolating quadratics, but uses only nested grids that are aligned with one another.

The great flexibility of the algorithm framework means that there is much work to be done in determining the best algorithms which conform to this framework. Extensive results for a specific algorithm conforming to the framework have not been presented in this paper; to do so would shift the focus of the paper away from the framework and onto that specific algorithm.

In the next section some properties of positive bases relevant to grid-based methods are introduced and in Section 3 an algorithmic framework for grid-based methods is described which allows considerable flexibility in the design of algorithms of this type. The main convergence results are established in Section 4 with further comments and discussion given in the final section.

2. Grid-based methods and positive bases. The algorithms under consideration seek a minimizer of the objective function $f : R^n \rightarrow R$ by examining f on a sequence of successively finer grids. Each grid $\mathcal{G}^{(m)}$ is defined by a set of n linearly independent basis vectors $\mathcal{V}^{(m)}$, where

$$\mathcal{V}^{(m)} = \left\{ v_i^{(m)} \in R^n : i = 1, \dots, n \right\}.$$

The points on the grid $\mathcal{G}^{(m)}$ are

$$\mathcal{G}^{(m)} = \left\{ x \in R^n : x = x_o^{(m)} + h^{(m)} \sum_{i=1}^n \eta_i v_i^{(m)} \right\}.$$

where $h^{(m)}$ is a positive scalar and each η_i is any integer. The parameter $h^{(m)}$ is referred to as the mesh size, and is adjusted as m is increased in order to ensure that the meshes become finer in a manner needed to establish convergence. The point $x_o^{(m)}$ allows the grids to be offset relative to one another. The basis vectors in $\mathcal{V}^{(m)}$ are parallel to the axes of the grid $\mathcal{G}^{(m)}$.

The set $\mathcal{V}^{(m)}$ is used to form a positive basis $\mathcal{V}_+^{(m)}$. There are two requirements for a set \mathcal{V}_+ to form a positive basis:

- (i) every vector in R^n can be written as a nonnegative linear combination of the vectors in \mathcal{V}_+ ;
- (ii) no member of \mathcal{V}_+ is expressible as a nonnegative linear combination of the remaining members of \mathcal{V}_+ .

It is shown in [3] that the cardinality p of any positive basis for R^n satisfies $n + 1 \leq p \leq 2n$. For example, if $\{v_1, v_2, \dots, v_n\}$ is a basis for R^n , then

$$(2.1) \quad \left\{ v_1, v_2, \dots, v_n, -\sum_{i=1}^n v_i \right\}$$

is a positive basis with $n + 1$ elements. At the other extreme,

$$(2.2) \quad \left\{ v_1, v_2, \dots, v_n, -v_1, -v_2, \dots, -v_n \right\}$$

is a positive basis with $2n$ elements.

Let $p^{(m)}$ denote the cardinality of $\mathcal{V}_+^{(m)}$. We assume throughout that the first n members of $\mathcal{V}_+^{(m)}$ are those of $\mathcal{V}^{(m)}$, and that the remaining elements are given by an integer linear combination of the members of $\mathcal{V}^{(m)}$:

$$(2.3) \quad v_j^{(m)} = \sum_{i=1}^n \zeta_{ij}^{(m)} v_i^{(m)}, \quad j = n+1, \dots, p^{(m)}$$

where each $\zeta_{ij}^{(m)}$ must be an integer so that if $x \in \mathcal{G}^{(m)}$ and $v \in \mathcal{V}_+^{(m)}$, then $x + h^{(m)}v \in \mathcal{G}^{(m)}$. Equation (2.3) requires the members of $\mathcal{V}_+^{(m)}$ to assume a specific order, and positive bases satisfying (2.3) will be called *ordered* positive bases.

Use of ordered positive bases permits the formation of termination conditions for the search on each grid, via the following theorem.

THEOREM 2.1. *If the set of vectors \mathcal{V}_+ is a positive basis, then*

$$g^T v \geq 0 \quad \forall v \in \mathcal{V}_+ \quad \Rightarrow \quad g = 0.$$

Proof. Let the members of \mathcal{V}_+ be v_i for $i = 1, \dots, |\mathcal{V}_+|$. Then, for any $g \in R^n$

$$-g = \sum_{i=1}^{|\mathcal{V}_+|} \eta_i v_i \quad \text{where} \quad \eta_i \geq 0 \quad \forall i$$

therefore, if $v_i^T g \geq 0$, for $i = 1, \dots, |\mathcal{V}_+|$

$$0 \geq (-g)^T g = \sum_{i=1}^{|\mathcal{V}_+|} \eta_i v_i^T g \geq 0.$$

The only possibility is $g = 0$. \square

This theorem motivates the following definition

DEFINITION 2.2. Grid local minimizer. *A point x on the grid $\mathcal{G}^{(m)}$ is defined as a grid local minimizer with respect to the positive basis $\mathcal{V}_+^{(m)}$ if and only if*

$$f(x + h^{(m)}v_i) \geq f(x), \quad \forall v_i \in \mathcal{V}_+^{(m)}.$$

This definition is motivated by the fact that

$$v^T \nabla f \geq 0 \quad \forall v \in \mathcal{V}_+^{(m)} \quad \Rightarrow \quad \nabla f = 0.$$

The conditions which define a grid local minimizer are a finite difference approximation to this.

In order to establish convergence, some restrictions must be imposed on the form of the ordered positive bases used to define the grid local minimizers. The following definition allows these restrictions to be simply expressed as linear relationships between the members of each ordered positive basis.

DEFINITION 2.3. Structural equivalence. *Two ordered positive bases $\{v_1, \dots, v_p\}$ and $\{w_1, \dots, w_p\}$ are regarded as structurally equivalent iff*

$$\forall j > n, \quad v_j = \sum_{i=1}^n \zeta_{ij} v_i \quad \iff \quad w_j = \sum_{i=1}^n \zeta_{ij} w_i.$$

Structurally equivalent positive bases necessarily have the same cardinality. As a simple example, the following two ordered positive bases for R^3 are structurally equivalent.

$$\{e_1, e_2, e_3, -e_1, -2(e_2 + e_3)\}, \quad \{e_3, e_1, e_2, -e_3, -2(e_1 + e_2)\},$$

where e_i is the i^{th} unit co-ordinate vector.

An appropriate framework for optimization algorithms using ordered positive bases is described and analysed in the following two sections.

3. The Algorithm Framework. The basic structure of the framework (listed immediately below) consists of two asynchronous loops. The outer loop (steps 1–3) selects each grid and checks the stopping conditions. The inner loop (step 2) conducts finite searches using each member of $\mathcal{V}_+^{(m)}$ in turn until $p^{(m)}$ consecutive searches fail to make progress. When this occurs a grid local minimizer has been found; the inner loop then terminates and the outer loop selects the new grid.

ALGORITHM FRAMEWORK A

Initialize $m = 1$, $k = 1$, and let $x_o^{(1)}$ be the initial point $x^{(1)}$.

while (stopping conditions do not hold) **do**

1. Choose $h^{(m)}$ and $\mathcal{V}_+^{(m)}$. Set $i = 1$ and $r = 0$. Set $p^{(m)} = |\mathcal{V}_+^{(m)}|$.

2. **while** $r < p^{(m)}$ **do**

(a) Calculate f at a finite number of points on the grid $\mathcal{G}^{(m)}$, including $x^{(k)} + h^{(m)}v_i^{(m)}$. If any points lower than $x^{(k)}$ are found, set $x^{(k+1)}$ equal to the lowest of these points, increment k , and let $r = 0$. Otherwise increment r .

(b) Set $i = i + 1$. If $i > p^{(m)}$, set $i = 1$.

end

3. Set $\hat{x}^{(m)} = x^{(k)}$. Execute any finite process, and let $x_o^{(m+1)}$ be the lowest known point. If this finite process yields descent set $x^{(k+1)} = x_o^{(m+1)}$ and increment k . Increment m .

end

In this framework r is the number of consecutive failed finite searches using the members of $\mathcal{V}_+^{(m)}$. When $r = p^{(m)}$ a grid local minimizer $\hat{x}^{(m)}$ has been found and the algorithm terminates the search over the current grid $\mathcal{G}^{(m)}$. The next grid $\mathcal{G}^{(m+1)}$ has its origin $x_o^{(m+1)}$ positioned at the lowest known point. The algorithm generates a new iterate $x^{(k)}$ only when a new lowest point (one with a strictly lower function value) is found. In contrast, it generates a new grid local minimizer $\hat{x}^{(m)}$ every time step 2 is completed. The sequence $\{x^{(k)}\}$ may have finitely many members, whereas the sequence $\{\hat{x}^{(m)}\}$ will always have infinitely many members (ignoring stopping conditions) and may contain repetitions of some members of $\{x^{(k)}\}$. However, every member of the sequence $\{\hat{x}^{(m)}\}$ is also a member of the sequence of iterates $\{x^{(k)}\}$. If, for example, $x^{(1)}$ is a global minimizer of f , then the sequence of iterates is the singleton set $\{x^{(1)}\}$, whereas the sequence of grid local minimizers is the set $\{x^{(1)}, x^{(1)}, x^{(1)}, \dots\}$.

The finite process in step 3 is arbitrary. Many possible choices exist, including a null process, or a finite ray search along an estimate of the direction of steepest descent or along a quasi-Newton direction. In proving convergence, finiteness is the only requirement for this process.

It is only necessary that the finite search in step 2(a) inspect one point, namely $x^{(k)} + h^{(m)}v_i^{(m)}$. No other point on $\mathcal{G}^{(m)}$ need be examined; however, it would normally be desirable to do so. For example, a search along the ray $x^{(k)} + \alpha h^{(m)}v_i^{(m)}$, $\alpha > 0$ could be implemented, and the lowest grid point in that search taken as $x^{(k+1)}$ provided descent is obtained. The option of examining a finite number of other grid points in step 2(a) admits the possibility of an arbitrarily long step to a grid point at each iteration of the inner while loop. For example, this could be exploited by examining the grid point closest to a quasi-Newton step, and accepting that grid point if it is sufficiently low. The intent behind such an approach would be to reduce the number of iterations of the while loop needed to locate a grid local minimum.

Framework A is not a special case of the analysis of pattern search methods in [6, 9]. In their notation, at each iteration [6, 9] examine a pattern of points $x^{(k)} + \Delta_k BC_k$, where Δ_k is a scale factor, B is a fixed matrix independent of the iteration number k , and C_k is an integer matrix. Three points should be noted:

- (i) Because B is independent of k , all grids are aligned with one another, and the grid alignment must be chosen at the start of optimization, before information from function evaluations is available;
- (ii) All grid axes at iteration k are scaled by the same factor Δ_k , and Δ_{k+1} must be a rational multiple of Δ_k ;
- (iii) The only way to scale directions is through the matrix C_k . But since the elements of C_k are integers, either the number of directions and scalings is small or else the grid may need to be much finer than the step sizes considered. Furthermore, pattern search methods cannot reach an arbitrary point or produce an arbitrary direction in a finite number of steps. For example, if the initial pattern is aligned with the x_1 and x_2 axes in two dimensions and $\Delta_0 = 1$, a pattern search method will require a very large number of iterations to reach a small neighborhood of the point $(0, \sqrt{2})$, and it cannot produce a direction at 30° to the x_1 axis. Thus pattern search methods do not possess the property of finite termination on convex quadratics that is standard with conjugate direction methods. By contrast, algorithms conforming to Framework A can possess this finite termination property; see [2] for such a method.

The analogues of Δ_k , B , and C_k in our framework are less restricted than in pattern search methods. For example, $h^{(k)}$ can be irrational, unlike Δ_k ; the bases $\mathcal{V}^{(m)}$ can be chosen independently from time to time, whereas B is fixed; and the ability to consider other grid points in step 2(a) of the framework offers the same level of freedom given by the matrices C_k . The techniques of [6, 9] are more general than Framework A in that they require C_k to contain only one of a finite number of integer positive bases, whereas we use exclusively the positive ordered basis $\mathcal{V}_+^{(m)}$. Framework A could be modified to use ordered positive bases other than $\mathcal{V}_+^{(m)}$ to explore the grid $\mathcal{G}^{(m)}$, but the flexibility in step 2(a), the choice of successive grids, and the arbitrary finite process in step 3 are likely to lessen the benefits of such an approach.

For Framework A, convergence can only be shown for subsequences of grid local minimizers. This is because the finite searches in step 2(a) are opportunistic; the first member encountered in $\mathcal{V}_+^{(m)}$ which gives descent leads to a new iterate. Convergence of the full sequence of iterates can be shown for a more restricted framework, Framework B (defined below), in which a thorough search is made along the member $s^{(k)}$

of $\mathcal{V}_+^{(m)}$ giving the “best drop” at $x^{(k)}$. The “best drop” member $s^{(k)}$ of $\mathcal{V}_+^{(m)}$ satisfies

$$(3.1) \quad f\left(x^{(k)} + h^{(m(k))}s^{(k)}\right) \leq f\left(x^{(k)} + h^{(m(k))}v\right) \quad \forall v \in \mathcal{V}_+^{(m(k))}$$

where $m(k)$ is the number of the grid on which $x^{(k)}$ is placed. This determination of $s^{(k)}$ requires $p^{(m(k))}$ function evaluations. The search along each $s^{(k)}$ must evaluate f at a sequence of points

$$(3.2) \quad \tilde{x}_i = x^{(k)} + \alpha_i h^{(m(k))}s^{(k)}, \quad i = 0, \dots,$$

where $\alpha_0 = 1$ and the integer sequence $\{\alpha_i\}, i \geq 1$, satisfies

$$\alpha_{i-1} + 1 \leq \alpha_i \leq \beta \alpha_{i-1} \quad \text{with} \quad \beta \geq 2.$$

The search may terminate only when an integer $\ell \geq 0$ is found such that $f(\tilde{x}_\ell) \leq f(\tilde{x}_{\ell+1})$.

ALGORITHM FRAMEWORK B

Initialize $m = 1$, $k = 1$, and let $x_o^{(1)}$ be the initial point $x^{(1)}$.

while (stopping conditions do not hold) **do**

1. Choose $h^{(m)}$ and $\mathcal{V}_+^{(m)}$. Set $i = 1$ and $r = 0$. Set $p^{(m)} = \left| \mathcal{V}_+^{(m)} \right|$.

2. **while** $x^{(k)}$ not a grid local minimizer **do**

(a) Calculate the “best drop” direction, $s^{(k)}$, satisfying (3.1). If $f^{(k)} \leq f(x^{(k)} + h^{(m)}s^{(k)})$, then exit step 2 as $x^{(k)}$ is a grid local minimizer.

(b) Starting with $\alpha_0 = 1$, choose successive integer values $\alpha_1, \alpha_2, \dots$ until

$$f\left(x^{(k)} + \alpha_{\ell+1}h^{(m)}s^{(k)}\right) \geq f\left(x^{(k)} + \alpha_\ell h^{(m)}s^{(k)}\right)$$

where $\alpha_{\ell+1} \in [\alpha_\ell + 1, \beta \alpha_\ell]$.

(c) Calculate f at a finite number of grid points, and choose $x^{(k+1)}$ to be the lowest of these points and $x^{(k)} + \alpha_\ell h^{(m)}s^{(k)}$. Increment k .

end

3. Set $\hat{x}^{(m)} = x^{(k)}$. Execute any finite process, and let $x_o^{(m+1)}$ be the lowest known point. If this finite process yields descent set $x^{(k+1)} = x_o^{(m+1)}$ and increment k . Increment m .

end

Framework B is a specialization of Framework A.

4. Convergence analysis. The convergence results for the methods discussed in this paper are similar to those in [9], but the method of analysis is sufficiently flexible to allow a myriad of other possibilities that may be more appropriate in other cases. The first theorem proves convergence of the subsequence of grid local minimizers for algorithms conforming to Framework A. This theorem is also applicable to Framework B because any algorithm conforming to Framework B also conforms to Framework A.

For the purpose of establishing convergence, it is assumed in this section that the stopping conditions are never invoked. This permits examination of the full sequence of iterates and grid local minimizers and their asymptotic properties. From a practical point of view stopping conditions are essential which is why they are incorporated in the general frameworks.

THEOREM 4.1. *For any optimization algorithm conforming to Framework A, assume that*

- (a) *the sequence of iterates $\{x^{(k)}\}$ is bounded;*
- (b) *f is continuously differentiable;*
- (c) *there exist positive constants K and κ such that $|\det(v_1^{(m)} \dots v_n^{(m)})| \geq \kappa$ and $\|v_i^{(m)}\| \leq K$ for all m and i ;*
- (d) *$h^{(m)} \rightarrow 0$ as $m \rightarrow \infty$; and*
- (e) *there is a finite subset of \mathcal{B} such that each member of \mathcal{B} is structurally equivalent to some member of this finite subset, where \mathcal{B} denotes the sequence of ordered positive bases $\{\mathcal{V}_+^{(m)}\}_{m=1}^\infty$.*

Then the sequence $\{\hat{x}^{(m)}\}$ of grid local minimizers has infinitely many members, and each cluster point $\hat{x}^{(\infty)}$ of $\{\hat{x}^{(m)}\}$ is a stationary point of f .

Proof. The proof is in two parts. The first part shows that Step 2 terminates, and consequently the sequence of grid local minimizers, $\{\hat{x}^{(m)}\}$, is infinite. The main part of the theorem is then established.

Firstly, by condition (a) of the theorem, there is a compact set \mathcal{F} for which $\{x^{(k)}\} \subset \mathcal{F}$. Hence the set $\mathcal{F} \cap \mathcal{G}^{(m)}$ is finite. The sequence of function values is strictly decreasing so each iterate is distinct from all others, and so only a finite number of iterates is generated using $\mathcal{G}^{(m)}$. Hence the finiteness of step 2(a) means the algorithm generates only a finite number of points using each grid. The only way the algorithm can change from the grid $\mathcal{G}^{(m)}$ is if the last iterate generated using the grid $\mathcal{G}^{(m)}$ is a grid local minimizer. Hence $\{\hat{x}^{(m)}\}$ is infinite.

Next, choose a specific cluster point $\hat{x}^{(\infty)}$ of $\{\hat{x}^{(m)}\}$ and choose some $\mathcal{S} \subset \mathcal{B}$ such that \mathcal{S} is an infinite subsequence of structurally equivalent bases and the corresponding subsequence of $\{\hat{x}^{(m)}\}$ converges to $\hat{x}^{(\infty)}$. Condition (e) ensures that one or more subsequences like \mathcal{S} exist, and that all but a finite number of members of \mathcal{B} belong to a subsequence like \mathcal{S} . Now replace the sequence of iterates $\{x^{(k)}\}$, and all other sequences by the infinite subsequences of themselves which correspond to the subsequence \mathcal{S} . It then follows that

$$(4.1) \quad f\left(\hat{x}^{(m)} + h^{(m)}v_i^{(m)}\right) \geq f\left(\hat{x}^{(m)}\right) \quad \forall i \in 1, \dots, p$$

where $p = p^{(m)}$ for all m such that $\mathcal{V}_+^{(m)} \in \mathcal{S}$. Now

$$\begin{aligned} f\left(\hat{x}^{(m)} + h^{(m)}v_i^{(m)}\right) &= f\left(\hat{x}^{(m)}\right) + \int_{t=0}^{h^{(m)}} \left[g\left(\hat{x}^{(m)} + tv_i^{(m)}\right) - \hat{g}^{(m)} + \hat{g}^{(m)}\right]^T v_i^{(m)} dt \\ &= f\left(\hat{x}^{(m)}\right) + h^{(m)}\left(\hat{g}^{(m)}\right)^T v_i^{(m)} + E_i^{(m)} \end{aligned}$$

where $g(x) \equiv \nabla f(x)$, where $\hat{g}^{(m)} = g(\hat{x}^{(m)})$, and where

$$E_i^{(m)} = \int_{t=0}^{h^{(m)}} \left[g\left(\hat{x}^{(m)} + tv_i^{(m)}\right) - \hat{g}^{(m)}\right]^T v_i^{(m)} dt.$$

The bound on $\|v_i^{(m)}\|$ in (c) yields

$$\left|E_i^{(m)}\right| \leq \int_{t=0}^{h^{(m)}} KM_i^{(m)} dt = h^{(m)}KM_i^{(m)}$$

where

$$M_i^{(m)} = \max \left\{ \left\| g \left(\hat{x}^{(m)} + tv_i^{(m)} \right) - \hat{g}^{(m)} \right\| : t \in \left[0, h^{(m)} \right] \right\}.$$

The continuity of g and the compactness of \mathcal{F} imply that g is also uniformly continuous on \mathcal{F} . The bound on $\|v_i^{(m)}\|$ in (c) then ensures that $M_i^{(m)} \rightarrow 0$ as $m \rightarrow \infty$ by condition (d).

Now $\hat{x}^{(\infty)}$ is a cluster point of the sequence $\{\hat{x}^{(m)}\}$ of grid local minimizers, and $\|v_i^{(m)}\| \leq K$ for all m and $i = 1, \dots, n$, so there is a subsequence $\{\bar{x}^{(m)}\}$ of $\{\hat{x}^{(m)}\}$, and corresponding subsequences $\{\bar{v}_i^{(m)}\}$ of $\{v_i^{(m)}\}$ for $i = 1, \dots, n$ which have unique limits $\hat{x}^{(\infty)}$ and $\bar{v}_i^{(\infty)}$ for $i = 1, \dots, n$. The structural equivalence of all members of \mathcal{S} implies that

$$(4.2) \quad \lim_{m \rightarrow \infty} \bar{v}_i^{(m)} = \bar{v}_i^{(\infty)} \quad \forall i = 1, \dots, p.$$

Condition (c) implies that $\bar{v}_1^{(\infty)}, \dots, \bar{v}_n^{(\infty)}$ are linearly independent and $\bar{v}_1^{(\infty)}, \dots, \bar{v}_p^{(\infty)}$ are bounded in norm by K . Hence $\{\bar{v}_1^{(\infty)}, \dots, \bar{v}_p^{(\infty)}\}$ is an ordered positive basis which is structurally equivalent to all members of \mathcal{S} .

Now (4.1) implies

$$\bar{h}^{(m)} \left(\bar{g}^{(m)} \right)^T \bar{v}_i^{(m)} + \bar{h}^{(m)} K \bar{M}_i^{(m)} \geq 0 \quad \forall i = 1, \dots, p.$$

In the limit as $m \rightarrow \infty$, condition (d) implies

$$\left(\hat{g}^{(\infty)} \right)^T \bar{v}_i^{(\infty)} \geq 0 \quad \forall i = 1, \dots, p$$

and so $\hat{g}^{(\infty)} = \nabla f(\hat{x}^{(\infty)}) = 0$ by Theorem 2.1. The choices of \mathcal{S} and of the cluster point of the sequence of grid local minimizers were arbitrary, so every cluster point of the sequence of grid local minimizers is a stationary point of the objective function.

□

Theorem 4.1 makes very few assumptions about how the sequence of grid local minimizers is generated; all that is required is that this sequence be bounded and have infinitely many members. Assumption (a) on the full sequence of iterates is only needed to establish these two properties. This assumption is automatically satisfied if, for example, the level set $\{x : f(x) \leq f(x^{(1)})\}$ is bounded, an assumption frequently made in convergence analyses, however, it may also be valid under much less restrictive conditions. Assumption (c) is easily satisfied by choosing each $\mathcal{V}_+^{(m)}$ appropriately. A very simple way to satisfy (d) is to halve h every time a grid local minimizer is found. An example of a more complex scheme is given in [2]. Assumption (e) is one of practicality and is easily enforced. We expect that most useful algorithms will use only one or two types of structurally equivalent bases (corresponding to ordered positive bases such as (2.1) or (2.2)); the proof is valid, however, for any finite number. Theorem 4.1 is, therefore, applicable to a wide range of algorithms.

If stopping conditions are never invoked and the sequence of iterates $x^{(k)}$ has finitely many members then its last member is necessarily a stationary point of f . In the more usual case, when an infinite sequence of iterates is generated, we now show convergence of the full sequence of iterates under the stricter Framework B.

THEOREM 4.2. *If the conditions of Theorem 4.1 hold then algorithms conforming to Framework B generate sequences of iterates which converge to one or more stationary points of f .*

Proof. The proof is by contradiction. Let $x^{(\infty)}$ be a cluster point of the sequence $\{x^{(k)}\}$ for which $g_\infty = \nabla f(x^{(\infty)})$ is non-zero. Following the proof of Theorem 4.1, replace $\{x^{(k)}\}$ by a subsequence of itself with $x^{(\infty)}$ as its unique limit, for which the corresponding subsequence of ‘best drop’ directions $\{s^{(k)}\}$ (defined by (3.1)) has a unique limit $s^{(\infty)}$, and for which all ordered positive bases are structurally equivalent and have a unique limit $\mathcal{V}_+^{(\infty)}$. The structural equivalence of these ordered positive bases, condition (c) of Theorem 4.1, and (4.2) show that $\mathcal{V}_+^{(\infty)}$ is also an ordered positive basis.

First it is shown that $g_\infty^T s^{(\infty)} < 0$. Now

$$\forall k, \forall v_i \in \mathcal{V}_+^{(m(k))}, \quad f\left(x^{(k)} + h^{(m(k))} v_i\right) - f^{(k)} \geq f\left(x^{(k)} + h^{(m(k))} s^{(k)}\right) - f^{(k)}$$

where $m(k)$ is the number of the grid on which $x^{(k)}$ is placed. Taylor’s series expansions on both sides yields

$$v_i^T g^{(k)} + L_k \geq \left(g^{(k)}\right)^T s^{(k)} - L_k \quad \forall v_i \in \mathcal{V}_+^{(m(k))}$$

where L_k is defined as

$$L_k = K \max \left\{ \|g(x) - g^{(k)}\| : x \in \mathcal{F} \text{ and } \|x - x^{(k)}\| \leq Kh^{(m(k))} \right\}.$$

In the limit as $k \rightarrow \infty$, $L_k \rightarrow 0$ by the uniform continuity of g on \mathcal{F} . Hence

$$(4.3) \quad g_\infty^T v_i^{(\infty)} \geq g_\infty^T s^{(\infty)} \quad \forall i = 1, \dots, p.$$

However, $g_\infty \neq 0$ so there exists a $v \in \mathcal{V}_+^{(\infty)}$ such that $g_\infty^T v < 0$. Clearly $g_\infty^T s^{(\infty)} < 0$ by inequality (4.3).

Next, define the closed ball B_ϵ about $x^{(\infty)}$ via

$$B_\epsilon = \left\{ x : \|x - x^{(\infty)}\| \leq \epsilon \right\}$$

and similarly for B_δ , where $\delta < \epsilon$. The continuity of $g(x)$ and the convergence of $s^{(k)}$ to $s^{(\infty)}$ imply

$$\exists N > 0, \exists \epsilon > 0, \text{ such that } \forall k > N, \forall x \in B_\epsilon,$$

$$(4.4) \quad g^T(x) s^{(k)} \leq \frac{1}{2} g_\infty^T s^{(\infty)} < 0 \quad \text{and} \quad \|s^{(k)}\| \leq 2 \|s^{(\infty)}\|.$$

Now choose a specific $k > N$ such that $x^{(k)} \in B_\delta$. The C^1 continuity of f implies

$$(4.5) \quad f^{(k)} \leq f^{(\infty)} + M\delta$$

where M is an upper bound for $\|g(x)\|$ over B_ϵ . The first inequality in (4.4) implies that f is strictly descending along the line segment

$$x^{(k)} + \alpha h^{(m(k))} s^{(k)} \quad \text{for } 0 \leq \alpha \leq \frac{(\epsilon - \delta)}{h^{(m(k))} \|s^{(k)}\|}.$$

The restrictions on successive α values in the ray searches mean that the final α value for the ray $x^{(k)} + \alpha h^{(m(k))} s^{(k)}$, $\alpha > 0$, is at least $(\epsilon - \delta)/(\beta h^{(m(k))} \|s^{(k)}\|)$. Hence

$$\begin{aligned} f^{(k+1)} &= f^{(k)} + \int_{t=0}^{\alpha} h^{(m(k))} \left(s^{(k)} \right)^T g \left(x^{(k)} + t h^{(m(k))} s^{(k)} \right) dt \\ &\leq f^{(k)} + \frac{\alpha h^{(m(k))}}{2} g_{\infty}^T s^{(\infty)} \\ &\leq f^{(k)} - \frac{\epsilon - \delta}{2\beta \|s^{(k)}\|} \left| g_{\infty}^T s^{(\infty)} \right|. \end{aligned}$$

Then the last inequality in (4.4), and (4.5) imply

$$f^{(k+1)} \leq f^{(\infty)} + M\delta - \frac{\epsilon - \delta}{4\beta \|s^{(\infty)}\|} \left| g_{\infty}^T s^{(\infty)} \right|.$$

As $k \rightarrow \infty$, δ can be made arbitrarily small, implying that $f^{(k)} < f^{(\infty)}$ for k large. The continuity of f and the monotonicity of $\{f^{(k)}\}$ imply that $x^{(\infty)}$ cannot be a cluster point of the sequence of iterates. \square

An example of an existing method which conforms to Framework A is that of Hooke and Jeeves [5]. Their method does not explicitly impose an upper bound on the step length, which appears to be at odds with Theorem 4.1. The applicability of Theorem 4.1 follows on noting that if condition (a) of the theorem holds (the sequence of iterates is bounded) then the maximum step length must also be bounded. Satisfaction of condition (a) of Theorem 4.1 is ensured if at least one iterate lies within a bounded level set. Although the method of Hooke and Jeeves conforms to Framework A, it makes little use of the flexibility afforded by that framework. An algorithm that exploits the flexibility allowed by Framework A is presented in [2], where numerical results are given for standard test functions; the authors expect further improvements to follow with more research into algorithms conforming to this framework.

5. Summary. We have presented two general algorithmic frameworks for unconstrained optimization methods based only on function values, and have shown that, under mild conditions, such algorithms generate sequences of grid local minimizers that are guaranteed to converge to stationary points. There is much scope for improving efficiency within Framework A through the choice of the ordered positive basis $\mathcal{V}_+^{(m)}$ and the finite process in step 3, which could, for example, allow a quasi-Newton step or a step to the minimizer of an interpolating quadratic. An efficient algorithm that uses this flexibility to align grid axes along conjugate directions is described in [2].

The authors believe that Theorem 4.1 is applicable to many effective grid search methods, and much work remains to be done in examining the properties of these algorithms.

Acknowledgements: The authors would like to acknowledge the contribution of John Dennis for providing the initial stimulus for this work, Margaret Wright for her many helpful comments and suggestions, and two anonymous referees.

REFERENCES

- [1] A. CONN, K. SCHEINBERG, AND P. L. TOINT, *On the convergence of derivative-free methods for unconstrained optimization*, in Approximation theory and optimization, M. D. Buhmann and A. Iserles, eds., Cambridge, 1997, Cambridge University Press, pp. 83–108.
- [2] I. D. COOPE AND C. J. PRICE, *A direct search conjugate directions algorithm for unconstrained minimization*, research report no. 188 (Nov. 1999), Department of Mathematics, Canterbury University, Christchurch, New Zealand.
- [3] C. DAVIS, *Theory of positive linear dependence*, American Journal of Mathematics, (1954), pp. 733–746.
- [4] C. ELSTER AND A. NEUMAIER, *A grid algorithm for bound-constrained optimization of noisy functions*, IMA Journal of Numerical Analysis, 15 (1995), pp. 585–608.
- [5] R. HOOKE AND T. A. JEEVES, *Direct search solution of numerical and statistical problems*, Journal of the Association for Computing Machinery (ACM), 8 (1961), pp. 212–219.
- [6] R. M. LEWIS AND V. TORCZON, *Rank ordering and positive bases in pattern search algorithms*, Tech. Rep. 96-71, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, Virginia 23681-2199, 1996. In revision for Mathematical Programming.
- [7] M. J. D. POWELL, *Direct search algorithms for optimization calculations*, Acta Numerica, 7 (1998), pp. 287–336.
- [8] V. TORCZON, *On the convergence of the multidirectional search algorithm*, SIAM Journal on Optimization, 1 (1991), pp. 123–145.
- [9] ———, *On the convergence of pattern search algorithms*, SIAM Journal on Optimization, 7 (1997), pp. 1–25.
- [10] M. H. WRIGHT, *Direct search methods: once scorned now respectable*, in Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis, Addison-Wesley, Reading, MA and Longman, Harlow, UK, 1996.