PATTERN SEARCH METHODS FOR LINEARLY CONSTRAINED MINIMIZATION*

ROBERT MICHAEL LEWIS† AND VIRGINIA TORCZON‡

Abstract. We extend pattern search methods to linearly constrained minimization. We develop a general class of feasible point pattern search algorithms and prove global convergence to a Karush-Kuhn-Tucker point. As in the case of unconstrained minimization, pattern search methods for linearly constrained problems accomplish this without explicit recourse to the gradient or the directional derivative of the objective. Key to the analysis of the algorithms is the way in which the local search patterns conform to the geometry of the boundary of the feasible region.

Key words. Pattern search, linearly constrained minimization.

AMS subject classifications. 49M30, 65K05

1. Introduction.

This paper continues the line of development in [8, 9, 15] and extends pattern search algorithms to optimization problems with linear constraints:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \ell \le Ax \le u, \end{array} \tag{1.1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $x \in \mathbf{R}^n$, $A \in \mathbf{Q}^{m \times n}$, $\ell, u \in \mathbf{R}^m$, and $\ell \le u$. We allow the possibility that some of the variables are unbounded either above or below by permitting $\ell_i, u_i = \pm\infty$, $i \in \{1, \cdots, m\}$. We also admit equality constraints by allowing $\ell_i = u_i$.

We can guarantee that if the objective f is continuously differentiable, then a subsequence of the iterates produced by a pattern search method for linearly constrained minimization converges to a Karush-Kuhn-Tucker point of problem (1.1). As in the case of unconstrained minimization, pattern search methods for linearly constrained problems accomplish this without explicit recourse to the gradient or the directional derivative of the objective. We also do not attempt to estimate Lagrange multipliers.

As with pattern search methods for bound constrained minimization [8], when we are close to the boundary of the feasible region the pattern of points over which we search must conform to the geometry of the boundary. The general idea, which also applies to unconstrained minimization [9], is that the pattern must contain search directions that comprise a set of generators for the cone of feasible directions. We must be a bit more careful than this; we must also take into account the constraints that are almost binding in order to be able to take sufficiently long steps. In the bound constrained case this turns out to be simple to ensure (though in Section 8.3 we will

---

*Revision: 3.0 Date: 1999/08/18 22:40:42

†Institute for Computer Applications in Science and Engineering, Mail Stop 403, NASA Langley Research Center, Hampton, Virginia 23681–2199; `buckaroo@icase.edu`. This research was supported by the National Aeronautics and Space Administration under NASA Contracts No. NAS1-19480 and NAS1-97046 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, Virginia 23681–2199.

‡Department of Computer Science, College of William & Mary, P. O. Box 8795, Williamsburg, Virginia 23187–8795; `va@cs.wm.edu`. This research was supported by the Air Force Office of Scientific Research under Grant F49620–95–1–0210, by the National Science Foundation under Grant No. CCR-9734044, and the National Aeronautics and Space Administration under NASA Contracts No. NAS1-19480 and NAS1-97046 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, Virginia 23681–2199.

sharpen the results in [8]). In the case of general linear constraints the situation is more complicated.

Practically, we imagine pattern search methods being most applicable in the case where there are relatively few linear constraints besides simple bounds on the variables. This is true for the applications that motivated our investigation. Our analysis does not assume nondegeneracy, but the class of algorithms we propose will be most practical when the problem is nondegenerate.

2. Background. After we presented this work at the 16th International Symposium on Mathematical Programming in Lausanne, Robert Mifflin brought to our attention the work of Jerrold May in [11], which extended the derivative-free algorithm for unconstrained minimization in [12] to linearly constrained problems. May proves both global convergence and superlinear local convergence for his method. To the best of our knowledge, this is the only other provably convergent derivative-free method for linearly constrained minimization.

Both May's approach and the methods described here use only values of the objective at feasible points to conduct their searches. Moreover, the idea of using as search directions the generators of cones that are polar to cones generated by the normals of faces near the current iterate appears already in [11]. This is unavoidable if one wishes to be assured of not overlooking any possible feasible descent in f using only values of f at feasible points.

On the other hand, there are significant differences between May's work and the approach we discuss here. May's algorithm is more obviously akin to a finite-difference quasi-Newton method. Most significantly, May enforces a sufficient decrease condition; pattern search methods do not. Avoiding a sufficient decrease condition is useful in certain situations where the objective is prone to numerical error. The absence of a quantitative decrease condition also allows pattern search methods to be used in situations where only comparison (ranking) of objective values is possible.

May also assumes that the active constraints are never linearly dependent—i.e., nondegeneracy. Our analysis, which is based on the intrinsic geometry of the feasible region rather than its algebraic description, handles degeneracy (though from a practical perspective, degeneracy can make the calculation of the pattern expensive). On the other hand, we must place additional algebraic restrictions on the search directions since pattern search methods require their iterates to lie on a rational lattice. To do so, we require that the matrix of constraints A in (1.1) be rational. This mild restriction is a price paid for not enforcing a sufficient decrease condition.

May's algorithm also has a more elaborate way of sampling f than the general pattern search algorithm we discuss here. This, and the sufficient decrease condition he uses, enables May to prove local superlinear convergence, which is stronger than the purely global results we prove here.

In Section 3 we outline the general definition of pattern search methods for linearly constrained minimization. In Section 4 we present the convergence results. In Section 5 we review those results from the analysis for the unconstrained case upon which we rely for the analysis in the presence of linear constraints. In Section 6 we prove our main results. In Section 7 we discuss stopping criteria and the questions of identifying active constraints and estimating Lagrange multipliers. In Section 8 we outline practical implementations of pattern search methods for linearly constrained minimization. Section 9 contains some concluding remarks, while Section 10 contains essential, but rather technical, results concerning the geometry of polyhedra that are required for the proofs in Section 6.

Notation. We denote by $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{Z}$, and $\mathbf{N}$ the sets of real, rational, integer, and natural numbers, respectively. The ith standard basis vector will be denoted by e_i. Unless otherwise noted, norms and inner products are assumed to be the Euclidean norm and inner product. We will denote the gradient of the objective by $g(x)$.

We will use Ω to denote the feasible region for problem (1.1):

$$\Omega = \{\, x \in \mathbf{R}^n \;\mid\; \ell \leq Ax \leq u \,\}.$$

Given a convex cone $K \subset \mathbf{R}^n$ we denote its polar cone by K°; K° is the set of $v \in \mathbf{R}^n$ such that $(v, w) \leq 0$ for all $w \in K$, where (v, w) denotes the Euclidean inner product.

If Y is a matrix, $y \in Y$ means that the vector y is a column of Y.

3. Pattern search methods. We begin our discussion with a simple instance of a pattern search algorithm for unconstrained minimization: minimize $f(x)$. At iteration k, we have an iterate $x_k \in \mathbf{R}^n$ and a step-length parameter $\Delta_k > 0$. We successively look at the points $x_+ = x_k \pm \Delta_k e_i$, $i \in \{1, \ldots, n\}$, until we find x_+ for which $f(x_+) < f(x_k)$. Fig. 3.1 illustrates the set of points among which we search for x_+ for $n = 2$. This set of points is an instance of what we call a pattern, from which pattern search takes its name. If we find no x_+ such that $f(x_+) < f(x_k)$, then we reduce Δ_k by half and continue; otherwise, we leave the step-length parameter alone, setting $\Delta_{k+1} = \Delta_k$ and $x_{k+1} = x_+$. In the latter case we can also increase the step-length parameter, say, by a factor of 2, if we feel a longer step might be justified. We repeat the iteration just described until Δ_k is deemed sufficiently small.

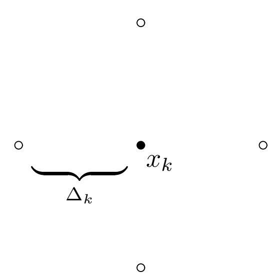

FIG. 3.1. *An illustration of pattern search for unconstrained minimization.*

One important feature of pattern search that plays a significant role in the global convergence analysis is that we do not need to have an estimate of the derivative of f at x_k so long as included in the search is a sufficient set of directions to form a positive spanning set for the cone of feasible directions, which in the unconstrained case is all of $\mathbf{R}^n$. In the unconstrained case the set $\{\, \pm e_i \;\mid\; i = 1, \ldots, n \,\}$ satisfies this condition, the purpose of which is to ensure that if the current iterate is not a stationary point of the problem, then we have at least one descent direction.

For the linearly constrained case we expand the notion of what constitutes a sufficient set of search directions. Now we must take into account explicit information about the problem: to wit, the geometry of the *nearby* linear constraints. We need to ensure that if we are not at a constrained stationary point, we have at least one feasible direction of descent. Moreover, we need a feasible direction of descent along which we will remain feasible for a sufficiently long distance to avoid taking too short a step. This is a crucial point since, just as in the unconstrained case, we will not enforce any notion of sufficient decrease. Practically, we must ensure that we have directions that allow us to move parallel to the constraints.

We modify the example given above by adding linear constraints near the current iterate x_k and show the effect this has on the choice of pattern. We add one further qualification to the essential logic of pattern search for the unconstrained case by noting that we are considering a feasible point method, so the initial iterate x_0 and all subsequent iterates must be feasible. To enforce this, we can introduce the simple rule of assigning an arbitrarily high function value (say $+\infty$) to any step that takes the search outside the feasible region defined by Ω. Otherwise, the logic of pattern search remains unchanged.

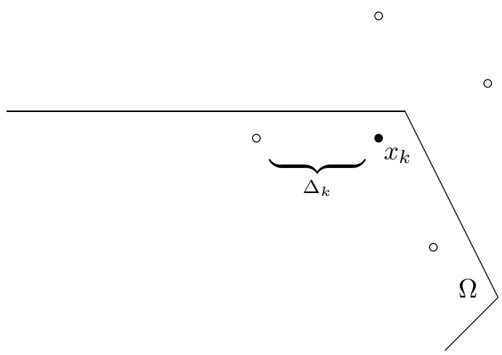

FIG. 3.2. *An illustration of pattern search for linearly constrained minimization.*

We now turn to the technical components of the general pattern search method for the linearly constrained problem (1.1). We borrow much of the machinery from the unconstrained case [15], modified in view of more recent developments in [8, 9]. We begin by describing how the pattern is specified and then used to generate subsequent iterates.

3.1. The pattern. The pattern for linearly constrained minimization is defined in a way that is a little less flexible than for patterns in the unconstrained case. In [15], at each iteration the pattern P_k is specified as the product $P_k = BC_k$ of two components, a fixed *basis matrix* B and a *generating matrix* C_k that can vary from iteration to iteration. This description of the pattern was introduced in the unconstrained case in order to unify the features of such disparate algorithms as the method of Hooke and Jeeves [7] and multidirectional search (MDS) [14]. In the case of bound constrained problems [8], we introduced restrictions on the pattern P_k itself rather than on B and C_k independently, but maintained the pretense of the independence of the choice of the basis and generating matrices.

For linearly constrained problems, we will ignore the basis—i.e., we will take $B = I$—and work directly in terms of the pattern P_k (for many of the classical pattern search methods for unconstrained minimization, $B = I$). We do this because, as with bound constrained problems, we need to place restrictions on P_k itself and it is simplest just to ignore B.

A *pattern* P_k is a matrix $P_k \in \mathbf{Z}^{n \times p_k}$. We will place a lower bound on p_k, but it has no upper bound. To obtain the lower bound, we begin by partitioning the pattern matrix into components:

$$P_k \quad = \quad [\quad \Gamma_k \quad L_k \quad].$$

In Section 3.5 we will describe certain geometrical restrictions that $\Gamma_k \in \mathbf{Z}^{n \times r_k}$ must satisfy. For now we simply observe that $r_k \geq n+1$. We will have more to say on r_k in Section 8, in particular, how we may reasonably expect to arrange $r_k \leq 2n$. We also will have occasion to refer to Γ_k as the *core* pattern since it represents the set of sufficient directions required for the analysis. We require that $L_k \in \mathbf{Z}^{n \times (p_k - r_k)}$ contains at least one column, a column of zeroes; this is purely a convenience we will explain shortly. Additional columns of L_k may be present to allow algorithmic refinements but they play little active role in the analysis. Given these definitions of the components Γ_k and L_k of P_k, it should be clear that $p_k \geq r_k + 1 > n + 1$.

We define a *trial step* s_k^i to be any vector of the form $s_k^i = \Delta_k c_k^i$, where $\Delta_k \in \mathbf{R}$, $\Delta_k > 0$, and c_k^i denotes a column of $P_k = [c_k^1 \cdots c_k^{p_k}]$. We call a trial step s_k^i *feasible* if $(x_k + s_k^i) \in \Omega$. At iteration k, a *trial point* is any point of the form $x_k^i = x_k + s_k^i$, where x_k is the current iterate. We will accept a step s_k from among the trial steps s_k^i that have been considered to form the next iterate $x_{k+1} = x_k + s_k$. The inclusion of a column of zeroes in L_k allows for a zero step, i.e., $x_{k+1} = x_k$. Among other things, this ensures that if x_k is feasible, then the pattern P_k always contains at least one step—the zero step—that makes it possible to produce a feasible x_{k+1}.

3.2. The linearly constrained exploratory moves. Pattern search methods proceed by conducting a series of *exploratory moves* about the current iterate x_k to choose a new iterate $x_{k+1} = x_k + s_k$, for some feasible step s_k determined during the course of the exploratory moves. The hypotheses on the result of the linearly constrained exploratory moves, given in Fig. 3.3, allow a broad choice of exploratory moves while ensuring the properties required to prove convergence. In the analysis of pattern search methods, these hypotheses assume the role played by sufficient decrease conditions in quasi-Newton methods. The only change from the unconstrained case

1. $s_k \in \Delta_k P_k = \Delta_k \, [\Gamma_k \; L_k]$.
2. $(x_k + s_k) \in \Omega$.
3. If $\min \{ \, f(x_k + y) \;\mid\; y \in \Delta_k \Gamma_k \text{ and } (x_k + y) \in \Omega \, \} < f(x_k)$,
   then $f(x_k + s_k) < f(x_k)$.

FIG. 3.3. *Hypotheses on the result of the linearly constrained exploratory moves.*

is the requirement that the iterates must be feasible.

We also observe that the last of the hypotheses is not as restrictive as may first appear. Another way to state the condition is to say that the exploratory moves are allowed to return the zero step only if there is no feasible step $s_k \in \Delta_k \Gamma_k$ which yields improvement over $f(x_k)$. Otherwise, we may accept *any* feasible $s_k \in \Delta_k P_k$ for which $f(x_k + s_k) < f(x_k)$. Thus, in the unconstrained example depicted in Fig. 3.1, while we look successively in each of the directions defined by the unit basis vectors for x_+ for which $f(x_+) < f(x_k)$, we are free to abandon the search the moment we find such an x_+. This means that if we are lucky, we can get by with as few as one evaluation of $f(x)$ in an iteration. The same holds for the example with linear constraints. This economy is possible because we do not enforce a sufficient decrease condition on the improvement realized in the objective.

3.3. The generalized pattern search method for linearly constrained problems. Fig. 3.4 states the general pattern search method for minimization with linear constraints. To define a particular pattern search method, we must specify the pattern P_k, the linearly constrained exploratory moves to be used to produce a feasible step s_k, and the algorithms for updating P_k and Δ_k. We defer a discussion of stopping criteria to Section 7.

Let $x_0 \in \Omega$ and $\Delta_0 > 0$ be given.
For $k = 0, 1, \cdots,$
- a) Compute $f(x_k)$.
- b) Determine a step s_k using a linearly constrained exploratory moves algorithm.
- c) If $f(x_k + s_k) < f(x_k)$, then $x_{k+1} = x_k + s_k$. Otherwise $x_{k+1} = x_k$.
- d) Update P_k and Δ_k.

FIG. 3.4. *The generalized pattern search method for linearly constrained problems.*

3.4. The updates. Fig. 3.5 specifies the rules for updating Δ_k. The aim of the update of Δ_k is to force decrease in $f(x)$. An iteration with $f(x_k + s_k) < f(x_k)$ is *successful;* otherwise, the iteration is *unsuccessful.* As is characteristic of pattern search methods, a step need only yield *simple* decrease, as opposed to *sufficient* decrease, in order to be acceptable.

There are two possibilities:
- (a) If $f(x_k + s_k) \geq f(x_k)$ (i.e., the iteration is unsuccessful), then $\Delta_{k+1} = \theta_k \Delta_k$, where $\theta_k \in (0, 1)$.
- (b) If $f(x_k + s_k) < f(x_k)$ (i.e., the iteration is successful), then $\Delta_{k+1} = \lambda_k \Delta_k$, where $\lambda_k \in [1, +\infty)$.

The parameters θ_k and λ_k are not allowed to be arbitrary, but must be of the following particular form. Let $\tau \in \mathbf{Q}$, $\tau > 1$, and $\{w_0, \cdots, w_L\} \subset \mathbf{Z}$, $w_0 < 0$, $w_L \geq 0$, and $w_0 < w_1 < \cdots < w_L$, where $L > 1$ is independent of k. Then θ_k must be of the form τ^{w_i} for some $w_i \in \{w_0, \cdots, w_L\}$ such that $w_i < 0$, while λ_k must be of the form τ^{w_j} for some $w_j \in \{w_0, \cdots, w_L\}$ such that $w_j \geq 0$.

FIG. 3.5. *Updating* Δ_k.

We will sometimes refer to outcome (a) in Fig. 3.5, a reduction of Δ_k, as back-tracking, in a loose analogy to back-tracking in line-search methods. Note that part (3) in Fig. 3.3 prevents back-tracking, and thus shorter steps, unless we first sample $f(x)$ in a suitably large set of directions from x_k and find no improvement. This is at the heart of the global convergence analysis.

3.5. Geometrical restrictions on the pattern. In the case of linearly constrained minimization, the core pattern Γ_k must reflect the geometry of the feasible region when the iterates are near the boundary. Pattern search methods do not approximate the gradient of the objective, but instead rely on a sufficient sampling of $f(x)$ to ensure that feasible descent will not be overlooked if the pattern is sufficiently small. We now discuss the geometrical restrictions on the pattern that make this possible in the presence of linear constraints.

3.5.1. The geometry of the nearby boundary. We begin with the relevant features of the boundary of the feasible region near an iterate. Let a_i^T be the ith row of the constraint matrix A in (1.1), and define

$$A_{\ell_i} = \{\ x \ \mid \ a_i^T x = \ell_i\ \}$$
$$A_{u_i} = \{\ x \ \mid \ a_i^T x = u_i\ \}.$$

These are the boundaries of the half-spaces whose intersection defines Ω. Set

$$\partial\Omega_{\ell_i}(\varepsilon) = \{\ x \in \Omega \ \mid\ \mathrm{dist}(x, A_{\ell_i}) \leq \varepsilon\ \},$$
$$\partial\Omega_{u_i}(\varepsilon) = \{\ x \in \Omega \ \mid\ \mathrm{dist}(x, A_{u_i}) \leq \varepsilon\ \},$$

and

$$\partial\Omega(\varepsilon) = \bigcup_{i=1}^{m} \left(\partial\Omega_{\ell_i}(\varepsilon) \cup \partial\Omega_{u_i}(\varepsilon)\right).$$

Given $x \in \Omega$ and $\varepsilon \geq 0$ we define the index sets

$$I_\ell(x, \varepsilon) = \{\ i \ \mid\ x \in \partial\Omega_{\ell_i}(\varepsilon)\ \} \tag{3.1}$$
$$I_u(x, \varepsilon) = \{\ i \ \mid\ x \in \partial\Omega_{u_i}(\varepsilon)\ \}. \tag{3.2}$$

For $i \in I_\ell(x, \varepsilon)$ we define

$$\nu_{\ell_i}(x, \varepsilon) = -a_i \tag{3.3}$$

and for $i \in I_u(x, \varepsilon)$ we define

$$\nu_{u_i}(x, \varepsilon) = a_i. \tag{3.4}$$

These are the outward pointing normals to the corresponding faces of Ω.

Given $x \in \Omega$ we will define the cone $K(x, \varepsilon)$ to be the cone generated by the vectors $\nu_{\ell_i}(x, \varepsilon)$ for $i \in I_\ell(x, \varepsilon)$ and $\nu_{u_i}(x, \varepsilon)$ for $i \in I_u(x, \varepsilon)$. Recall that a convex cone K is called finitely generated if there exists a finite set of vectors $\{v_1, \cdots, v_r\}$ (the generators of K) such that

$$K = \{\ v \ \mid\ v = \sum_{i=1}^{r} \lambda_i v_i,\ \lambda_i \geq 0,\ i = 1, \cdots, r\ \}.$$

Finally, let $P_{K(x,\varepsilon)}$ and $P_{K^\circ(x,\varepsilon)}$ be the projections (in the Euclidean norm) onto $K(x, \varepsilon)$ and $K^\circ(x, \varepsilon)$, respectively. By convention, if $K(x, \varepsilon) = \emptyset$, then $K^\circ(x, \varepsilon) = \mathbf{R}^n$. Observe that $K(x, 0)$ is the cone of normals to Ω at x, while $K^\circ(x, 0)$ is the cone of tangents to Ω at x.

The cone $K(x, \varepsilon)$, illustrated in Fig. 3.6, is the cone generated by the normals to the faces of the boundary within distance ε of x. Its polar $K^\circ(x, \varepsilon)$ is important because if $\varepsilon > 0$ is sufficiently small, we can proceed from x along all directions in $K^\circ(x, \varepsilon)$ for a distance $\delta > 0$, depending only on ε, and still remain inside the feasible region. This is not the case for directions in the tangent cone of the feasible region at x, since the latter cone does not reflect the proximity of the boundary for points close to, but not on, the boundary.

3.5.2. Specifying the pattern. We now state the geometrical restriction on the pattern P_k. We require the core pattern Γ_k of P_k to include generators for all of the cones $K^\circ(x_k, \varepsilon)$, $0 \leq \varepsilon \leq \varepsilon^*$, for some $\varepsilon^* > 0$ that is independent of k.

We also require that the collection $\mathbf{\Gamma} = \cup_{k=0}^{\infty} \Gamma_k$ be finite. Thus (and this is the real point), $\mathbf{\Gamma}$ will contain a finite set of generators for all of the cones $K^\circ(x_k, \varepsilon)$, $0 \leq \varepsilon \leq \varepsilon^*$. Note that as ε varies from 0 to ε^* there is only a finite number of distinct cones $K(x_k, \varepsilon)$ since there is only a finite number of faces of Ω. This means that the finite cardinality of $\mathbf{\Gamma}$ is not an issue. There remains the question of constructing sets of generators that are also integral; we address the issue of constructing suitable patterns, by implicitly estimating ε^*, in Section 8. However, we will see that the construction is computationally tractable, and in many cases is not particularly difficult. We close by noting that the condition that Γ_k contains generators of $K^\circ(x_k, \varepsilon)$ implies that Γ_k contains generators for all tangent cones to Ω at all feasible points near x_k.

If x_k is "far" from the boundary in the sense that $K(x_k, \varepsilon) = \emptyset$, then $K^\circ(x_k, \varepsilon) = \mathbf{R}^n$ and a set of generators for $K^\circ(x_k, \varepsilon)$ is simply a positive spanning set for $\mathbf{R}^n$ [5, 9]. (A positive spanning set is a set of generators for a cone in the case that the cone is a vector space.) If the iterate is suitably in the interior of Ω, the algorithm will look like a pattern search algorithm for unconstrained minimization [9], as it ought. On the other hand, if x_k is near the boundary, $K(x_k, \varepsilon) \neq \emptyset$ and the pattern must conform to the local geometry of the boundary, as depicted in Figs. 3.2 and 3.6.

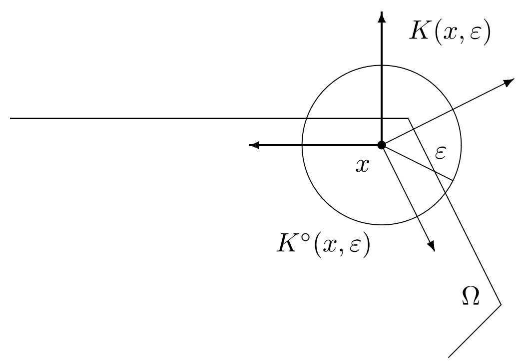

FIG. 3.6. *The situation near the boundary.*

The design of the pattern reflects the fundamental challenge in the development of constrained pattern search methods. We do not have an estimate of the gradient of the objective and consequently we have no idea which constraints locally limit feasible improvement in $f(x)$. In a projected gradient method one has the gradient and can detect the local interaction of the descent direction with the boundary by conducting a line-search along the projected gradient path. In derivative-free methods such as pattern search we must have a sufficiently rich set of directions in the pattern since any subset of the nearby faces may be the ones that limit the feasibility of the steepest descent direction, which is itself unavailable for use in the detection of the important nearby constraints. Nonetheless, in Section 4 we are able to outline the conditions for global convergence and in Section 8 we outline practical implementations of pattern search methods for linearly constrained minimization.

4. Convergence analysis. In this section we state the convergence results for pattern search methods for linearly constrained minimization. We defer the proofs of these results to Section 6, after reviewing existing results for pattern search methods in Section 5.

We first summarize features of the algorithm whose statements are scattered throughout Section 3.

HYPOTHESIS 0.

1. *The pattern $P_k = [\,\Gamma_k \; L_k\,] \in \mathbf{Z}^{n \times p_k}$, $p_k > n+1$, so that all search directions are integral vectors scaled by $\Delta_k \in \mathbf{R}^n$. All steps s_k are then required to be of the form $\Delta_k c_k^i$, where c_k^i denotes a column of $P_k = [c_k^1 \cdots c_k^{p_k}]$.*

2. *The core pattern $\Gamma_k \in \mathbf{Z}^{n \times r_k}$, $r_k \geq n+1$, belongs to $\mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is a finite set of integral matrices the columns of which include generators for all of the cones $K^\circ(x_k, \varepsilon)$, $0 \leq \varepsilon \leq \varepsilon^*$, for some $\varepsilon^* > 0$ that is independent of k.*
3. *The matrix $L_k \in \mathbf{Z}^{n \times (p_k - r_k)}$ contains at least one column, a column of zeroes.*
4. *The rules for updating Δ_k are as given in Fig. 3.5.*
5. *The exploratory moves algorithm returns steps that satisfy the conditions given in Fig. 3.3.*

We now add some additional hypotheses on the problem (1.1).

HYPOTHESIS 1. *The constraint matrix A is rational.*

Hypothesis 1 is a simple way of ensuring that we can find a rational lattice that fits inside the feasible region in a suitable way. In particular, the rationality of A ensures that we can construct Γ_k satisfying part 2 of Hypothesis 0, as discussed further in Section 8.

HYPOTHESIS 2. *The set $L_\Omega(x_0) = \{ \, x \in \Omega \;|\; f(x) \leq f(x_0) \, \}$ is compact.*

HYPOTHESIS 3. *The objective $f(x)$ is continuously differentiable on an open neighborhood D of $L_\Omega(x_0)$.*

We next remind the reader that unless otherwise noted, norms are assumed to be the Euclidean norm and that we denote by $g(x)$ the gradient of the objective f at x. Let P_Ω be the projection onto Ω. For feasible x, let

$$q(x) = P_\Omega(x - g(x)) - x.$$

Note that because the projection P_Ω is non-expansive, $q(x)$ is continuous on Ω. The following proposition summarizes properties of q that we need, particularly the fact that x is a constrained stationary point for (1.1) if and only if $q(x) = 0$. The results are classical; see Section 2 of [6], for instance.

PROPOSITION 4.1. *Let $x \in \Omega$. Then*

$$\| \, q(x) \, \| \leq \| \, g(x) \, \|$$

and x is a stationary point for problem (1.1) if and only if $q(x) = 0$.

We can now state the first convergence result for the general pattern search method for linearly constrained minimization.

THEOREM 4.2. *Assume Hypotheses 0–3 hold. Let $\{x_k\}$ be the sequence of iterates produced by the generalized pattern search method for linearly constrained minimization (Fig. 3.4). Then*

$$\liminf_{k \to +\infty} \| \, q(x_k) \, \| = 0 \, .$$

As an immediate corollary, we have

COROLLARY 4.3. *There exists a limit point of $\{x_k\}$ that is a constrained stationary point for (1.1).*

Note that Hypothesis 2 guarantees the existence of one such limit point.

We can strengthen Theorem 4.2, in the same way that we do in the unconstrained and bound constrained cases [8, 15], by adding the following hypotheses.

HYPOTHESIS 4. *The columns of the pattern matrix P_k remain bounded in norm, i.e., there exists $c_4 > 0$ such that for all k, $c_4 > \| \, c_k^i \, \|$, for all $i = 1, \cdots, p_k$.*

HYPOTHESIS 5. *The original hypotheses on the result of the linearly constrained exploratory moves are replaced with the stronger version given in Fig. 4.1.*

The third condition is stronger than the hypotheses on the result of the linearly constrained exploratory moves given in Fig. 3.3. Now we tie the amount of decrease

1. $s_k \in \Delta_k P_k = \Delta_k \, [\Gamma_k \; L_k]$.
2. $(x_k + s_k) \in \Omega$.
3. If $\min \{ \, f(x_k + y) \;\mid\; y \in \Delta_k \Gamma_k$ and $(x_k + y) \in \Omega \, \} < f(x_k)$,
   then $f(x_k + s_k) \leq \min \{ \, f(x_k + y) \;\mid\; y \in \Delta_k \Gamma_k$ and $(x_k + y) \in \Omega \, \}$.

FIG. 4.1. *Strong hypotheses on the result of the linearly constrained exploratory moves.*

in $f(x)$ that must be realized by the step s_k to the amount of decrease that could be realized were we to rely on the local behavior of the linearly constrained problem, as defined by the columns of Γ_k.

HYPOTHESIS 6. *We have* $\lim_{k \to +\infty} \Delta_k = 0$.
Note that we do not require Δ_k to be monotone non-increasing.

Then we obtain the following stronger results.

THEOREM 4.4. *Assume Hypotheses 0–6 hold. Then for the sequence of iterates $\{x_k\}$ produced by the generalized pattern search method for linearly constrained minimization (Fig. 3.4),*

$$\lim_{k \to +\infty} \| \, q(x_k) \, \| = 0 \, .$$

COROLLARY 4.5. *Every limit point of $\{x_k\}$ is a constrained stationary point for (1.1).*
Again, Hypothesis 2 guarantees the existence of at least one such limit point.

5. Results from the standard theory. We need the following results from the analysis of pattern search methods in the unconstrained case. For the proofs, see [15]; these results generalize to the linearly constrained case without change. Theorem 5.1 is central to the convergence analysis for pattern search methods; it allows us to prove convergence for these methods in the absence of any sufficient decrease condition.

THEOREM 5.1. *Any iterate x_N produced by a generalized pattern search method for linearly constrained problems (Fig. 3.4) can be expressed in the following form:*

$$x_N = x_0 + \left(\beta^{r_{LB}} \alpha^{-r_{UB}} \right) \Delta_0 B \sum_{k=0}^{N-1} z_k, \tag{5.1}$$

where

- x_0 *is the initial guess,*
- $\beta / \alpha \equiv \tau$*, with* $\alpha, \beta \in \mathbf{N}$ *and relatively prime, and* τ *is as defined in the rules for updating* Δ_k *(Fig. 3.5),*
- r_{LB} *and* r_{UB} *are integers depending on* N*, where* $r_{LB} \leq 0$ *and* $r_{UB} \geq 0$*,*
- Δ_0 *is the initial choice for the step length control parameter,*
- B *is the basis matrix, and*
- $z_k \in \mathbf{Z}^n$*,* $k = 0, \cdots, N-1$*.*

Recall that in the case of linearly constrained minimization we set $B = I$.

The quantity Δ_k regulates step length as indicated by the following.

LEMMA 5.2. *(i) There exists a constant $\zeta_* > 0$, independent of k, such that for any trial step $s_k^i \neq 0$ produced by a generalized pattern search method for linearly constrained problems we have $\| \, s_k^i \, \| \geq \zeta_* \Delta_k$.*

(ii) Under Hypothesis 4, there exists a constant $\psi_* > 0$, independent of k, such that for any trial step s_k^i produced by a generalized pattern search method for linearly constrained problems we have $\Delta_k \geq \psi_* \| \, s_k^i \, \|$.

In the case of pattern search for linearly constrained problems, P_k is integral. Since $s_k^i \in \Delta_k P_k$, we may take $\zeta_* = 1$.

6. Proof of Theorems 4.2 and 4.4. We now proceed with the proofs of the two main results stated in Section 4. Essential to our arguments are some results concerning the geometry of polyhedra. We defer the treatment of these technical details to Section 10.

Given an iterate x_k, let $g_k = g(x_k)$ and $q_k = P_\Omega(x_k - g_k) - x_k$. Let $B(x, \delta)$ be the ball with center x and radius δ, and let ω denote the following modulus of continuity of g: given $x \in L_\Omega(x_0)$ and $\varepsilon > 0$,

$$\omega(x, \varepsilon) = \sup \{ \ \delta > 0 \ \mid \ B(x, \delta) \subset D \text{ and } \| \ g(y) - g(x) \ \| < \varepsilon \text{ for all } y \in B(x, \delta) \ \} .$$

Then we have this elementary proposition concerning descent directions, whose proof we omit (see [8]).

Proposition 6.1. *Let $s \in \mathbf{R}^n$ and $x \in L_\Omega(x_0)$. Assume that $g(x) \neq 0$ and $g(x)^T s \leq -\varepsilon \| \ s \ \|$ for some $\varepsilon > 0$. Then, if $\| \ s \ \| < \omega(x, \varepsilon/2)$,*

$$f(x+s) - f(x) \leq -\frac{\varepsilon}{2} \| \ s \ \|.$$

The next result is the crux of the convergence analysis. Using the results in Section 10, we show that if we are not at a constrained stationary point, then the pattern always contains a descent direction along which we remain feasible for a sufficiently long distance.

Let Γ^* be the maximum norm of any column of the matrices in the set $\mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is as in Section 3.1 and Section 3.5. If $\Delta_k \leq \delta / \Gamma^*$, then $\| \ s_k^i \ \| \leq \delta$ for all $s_k^i \in \Delta_k \Gamma_k$. Also define

$$h = \min_{\substack{1 \leq i \leq m \\ \ell_i \neq u_i}} \frac{u_i - \ell_i}{\| \ a_i \ \|}. \tag{6.1}$$

This is the minimum distance between the faces of Ω associated with the constraints that are not equality constraints. Finally, $\| \ g_k \ \|$ is bounded on $L_\Omega(x_0)$ by hypothesis; let g^* be an upper bound for $\| \ g_k \ \|$.

Proposition 6.2. *There exist $r_{6.2} > 0$ and $c_{6.2} > 0$ such that if $\eta > 0$, $\| \ q_k \ \| \geq \eta$, and $\Delta_k \leq r_{6.2} \eta^2$, then there is a trial step s_k^i defined by a column of $\Delta_k \Gamma_k$ for which, given $x_k \in \Omega$, $(x_k + s_k^i) \in \Omega$ and*

$$-g_k^T s_k^i \geq c_{6.2} \| \ q_k \ \| \| \ s_k^i \ \|.$$

Proof. Let

$$r = \min(\varepsilon^* / (g^*)^2, \ h / (2 (g^*)^2), \ r_{10.7}),$$

where ε^* is the constant introduced in Section 3.5.2, h is given by (6.1), and $r_{10.7}$ is the constant that appears in Proposition 10.7.

Now consider $\varepsilon = r \eta^2$. From Proposition 4.1, $\| \ q_k \ \| \leq \| \ g_k \ \| \leq g^*$, so our choice of r ensures that ε is sufficiently small that

1. $\varepsilon \leq \varepsilon^*$,
2. $\varepsilon \leq h/2$, and

3. $\varepsilon \le r_{10.7}\eta^2$.

Because of this last fact, (3), we may apply Proposition 10.7 to $w = -g_k$ with $x = x_k$ and $\gamma = g^*$ to obtain

$$\| P_{K^\circ(x_k,\varepsilon)}(-g_k) \| \ge c_{10.7}\| q_k \|. \tag{6.2}$$

Meanwhile, since we require the core pattern Γ_k of P_k to include generators for all of the cones $K^\circ(x_k,\delta)$, $\delta \le \varepsilon^*$, then, because $\varepsilon \le \varepsilon^*$, some subset of the core pattern steps s_k^i forms a set of generators for $K^\circ(x_k,\varepsilon)$. Consequently, by virtue of (6.2) we may invoke Corollary 10.4: for some $s_k^i \in \Delta_k\Gamma_k$ we have

$$-g_k^T s_k^i \ge c_{10.4} \ \| P_{K^\circ(x_k,\varepsilon)}(-g_k) \| \ \| s_k^i \|. \tag{6.3}$$

From (6.3) we then obtain

$$-g_k^T s_k^i \ge c_{10.4} \ c_{10.7} \ \| q_k \| \ \| s_k^i \| = c_{6.2} \ \| q_k \| \ \| s_k^i \|,$$

where $c_{6.2} = c_{10.4} \ c_{10.7}$. Thus we are assured of a descent direction inside the pattern.

Now we must show that we can take a sufficiently long step along this descent direction and remain feasible. Define

$$r_{6.2} = r/(2\Gamma^*)$$

and consider what happens when $\Delta_k \le r_{6.2}\eta^2$. We have $\Delta_k \le \varepsilon/(2\Gamma^*)$, and since $s_k^i \in \Delta_k\Gamma_k$, we have $\| s_k^i \| \le \varepsilon/2$. Since $s_k^i \in K^\circ(x_k,\varepsilon)$, and $\varepsilon \le h/2$ by (2) above, we can apply Proposition 10.8 to $w = s_k^i$ to conclude that $(x_k + s_k^i) \in \Omega$. □

We now show that if we are not at a constrained stationary point, we can always find a step in the pattern which is both feasible and yields improvement in the objective.

Proposition 6.3. *Given any $\eta > 0$, there exists $r_{6.3} > 0$, independent of k, such that if $\Delta_k \le r_{6.3}\eta^2$ and $\| q_k \| \ge \eta$, the pattern search method for linearly constrained minimization will find an acceptable step s_k; i.e., $f(x_{k+1}) < f(x_k)$ and $x_{k+1} = (x_k + s_k) \in \Omega$.*

If, in addition, the columns of the generating matrix remain bounded in norm and we enforce the strong hypotheses on the results of the linearly constrained exploratory moves (Hypotheses 4 and 5), then, given any $\eta > 0$, there exists $\sigma > 0$, independent of k, such that if $\Delta_k < r_{6.3}\eta^2$ and $\| q_k \| \ge \eta$, then

$$f(x_{k+1}) \le f(x_k) - \sigma\| q_k \|\| s_k \|.$$

Proof. Proposition 6.2 assures us of the existence of $r_{6.2}$ and a step s_k^i defined by a column of $\Delta_k\Gamma_k$ such that $(x_k + s_k^i) \in \Omega$ and

$$g_k^T s_k^i \le -c_{6.2}\| q_k \|\| s_k^i \|,$$

provided $\Delta_k \le r_{6.2}\eta^2$. Also, since $g(x)$ is uniformly continuous on $L_\Omega(x_0)$ and $L_\Omega(x_0)$ is a compact subset of the open set D on which $f(x)$ is continuously differentiable, there exists $\omega_* > 0$ such that

$$\omega\left(x_k, \frac{c_{6.2}}{2}\eta\right) \ge \omega_*$$

for all k for which $\| q_k \| \ge \eta$.

Now define

$$r_{6.3} = \min\left(r_{6.2}, \omega_*/(\Gamma^*(g^*)^2)\right)$$

and suppose $\| q_k \| \geq \eta$ and $\Delta_k \leq r_{6.3}\eta^2$. We have

$$\| s_k^i \| \leq \Delta_k \Gamma^* \leq \omega_* \leq \omega\left(x_k, \frac{c_{6.2}}{2}\| q_k \|\right).$$

Hence, by Proposition 6.1,

$$f(x_k + s_k^i) - f(x_k) \leq -\frac{c_{6.2}}{2}\| q_k \|\| s_k^i \|.$$

Thus, when $\Delta_k \leq r_{6.3}\eta^2$, $f(x_k + s_k^i) < f(x_k)$ for at least one feasible $s_k^i \in \Delta_k\Gamma_k$. The hypotheses on linearly constrained exploratory moves guarantee that if

$$\min\{\ f(x_k + y)\ \mid\ y \in \Delta_k\Gamma_k,\ (x_k + y) \in \Omega\ \} < f(x_k),$$

then $f(x_k+s_k) < f(x_k)$ and $(x_k+s_k) \in \Omega$. This proves the first part of the proposition.

If, in addition, we enforce the strong hypotheses on the result of the linearly constrained exploratory moves, then we actually have

$$f(x_{k+1}) - f(x_k) \leq -\frac{c_{6.2}}{2}\| q_k \|\| s_k^i \|.$$

Part *(i)* of Lemma 5.2 then ensures that

$$f(x_{k+1}) \leq f(x_k) - \frac{c_{6.2}}{2}\zeta_*\Delta_k\| q_k \|.$$

Applying part *(ii)* of Lemma 5.2, we arrive at

$$f(x_{k+1}) \leq f(x_k) - \frac{c_{6.2}}{2}\zeta_*\psi_*\| q_k \|\| s_k \|.$$

This yields the second part of the proposition with $\sigma = (c_{6.2}/2)\zeta_*\psi_*$. ◻

COROLLARY 6.4. *If* $\liminf_{k\to+\infty} \| q_k \| \neq 0$*, then there exists a constant* $\Delta_* > 0$ *such that for all* k*,* $\Delta_k \geq \Delta_*$.

Proof. By hypothesis, there exists N and $\eta > 0$ such that for all $k > N$, $\| q_k \| \geq \eta$. By Proposition 6.3, we can find $\delta = r_{6.3}\eta^2$ such that if $k > N$ and $\Delta_k < \delta$, then we will find an acceptable step. In view of the update of Δ_k given in Fig. 3.5, we are assured that for all $k > N$, $\Delta_k \geq \min(\Delta_N, \tau^{w_0}\delta)$. We may then take $\Delta_* = \min\{\Delta_0, \cdots, \Delta_N, \tau^{w_0}\delta\}$. ◻

The next theorem combines the strict algebraic structure of the iterates with the simple decrease condition of the generalized pattern search algorithm for linearly constrained problems, along with the rules for updating Δ_k, to tell us the limiting behavior of Δ_k.

THEOREM 6.5. *Under Hypotheses 0–3,* $\liminf_{k\to+\infty} \Delta_k = 0$.

Proof. The proof is like that of Theorem 3.3 in [15]. Suppose $0 < \Delta_{LB} \leq \Delta_k$ for all k. Using the rules for updating Δ_k, found in Fig. 3.5, it is possible to write Δ_k as $\Delta_k = \tau^{r_k}\Delta_0$, where $r_k \in \mathbf{Z}$.

The hypothesis that $\Delta_{LB} \leq \Delta_k$ for all k means that the sequence $\{\tau^{r_k}\}$ is bounded away from zero. Meanwhile, we also know that the sequence $\{\Delta_k\}$ is bounded above because all the iterates x_k must lie inside the set $L_\Omega(x_0) = \{\ x \in \Omega\ \mid\ f(x) \leq f(x_0)\ \}$

and the latter set is compact; part *(i)* of Lemma 5.2 (which is a consequence of the rules for updating Δ_k) then guarantees an upper bound Δ_{UB} for $\{\Delta_k\}$. This, in turn, means that the sequence $\{\tau^{r_k}\}$ is bounded above. Consequently, the sequence $\{\tau^{r_k}\}$ is a finite set. Equivalently, the sequence $\{r_k\}$ is bounded above and below.

Next we recall the exact identity of the quantities r_{LB} and r_{UB} in Theorem 5.1; the details are found in the proof of Theorem 3.3 in [15]. In the context of Theorem 5.1,

$$r_{LB} = \min_{0 \le k < N} \{r_k\} \qquad r_{UB} = \max_{0 \le k < N} \{r_k\}.$$

If, in the matter at hand, we let

$$r_{LB} = \min_{0 \le k < +\infty} \{r_k\} \qquad r_{UB} = \max_{0 \le k < +\infty} \{r_k\}, \tag{6.4}$$

then (5.1) holds for the bounds given in (6.4), and we see that for all k, x_k lies in the translated integer lattice G generated by x_0 and the columns of $\beta^{r_{LB}} \alpha^{-r_{UB}} \Delta_0 I$.

The intersection of the compact set $L_\Omega(x_0)$ with the lattice G is finite. Thus, there must exist at least one point x_* in the lattice for which $x_k = x_*$ for infinitely many k.

We now appeal to the simple decrease condition in part (c) of Fig. 3.4, which guarantees that a lattice point cannot be revisited infinitely many times since we accept a new step s_k if and only if $f(x_k) > f(x_k + s_k)$ and $(x_k + s_k) \in \Omega$. Thus there exists an N such that for all $k \ge N$, $x_k = x_*$, which implies $f(x_k) = f(x_k + s_k)$.

We now appeal to the algorithm for updating Δ_k (part (a) in Fig. 3.5) to see that $\Delta_k \to 0$, thus leading to a contradiction. $\square$

6.1. The proof of Theorem 4.2. The proof is like that of Theorem 3.5 in [15]. Suppose that $\liminf_{k \to +\infty} \| q_k \| \ne 0$. Then Corollary 6.4 tells us that there exists $\Delta_* > 0$ such that for all k, $\Delta_k \ge \Delta_*$. But this contradicts Theorem 6.5.

6.2. The Proof of Theorem 4.4. The proof, also by contradiction, follows that of Theorem 3.7 in [15]. Suppose $\limsup_{k \to +\infty} \| q_k \| \ne 0$. Let $\varepsilon > 0$ be such that there exists a subsequence $\| q(x_{m_i}) \| \ge \varepsilon$. Since

$$\liminf_{k \to +\infty} \| q_k \| = 0,$$

given any $0 < \eta < \varepsilon$, there exists an associated subsequence l_i such that

$$\| q_k \| \quad \ge \quad \eta \qquad \text{for} \qquad m_i \le k < l_i, \qquad\qquad \| q(x_{l_i}) \| \quad < \quad \eta.$$

Since $\Delta_k \to 0$, we can appeal to Proposition 6.3 to obtain for $m_i \le k < l_i$, i sufficiently large,

$$f(x_k) - f(x_{k+1}) \quad \ge \quad \sigma \| q_k \| \| s_k \| \quad \ge \quad \sigma\eta \| s_k \|,$$

where $\sigma > 0$. Summation then yields

$$f(x_{m_i}) - f(x_{l_i}) \quad \ge \quad \textstyle\sum_{k=m_i}^{l_i} \sigma\eta \| s_k \| \quad \ge \quad \sigma\eta \| x_{m_i} - x_{l_i} \|.$$

Since f is bounded below on the set $L_\Omega(x_0)$, we know that $f(x_{m_i}) - f(x_{l_i}) \to 0$ as $i \to +\infty$, so $\| x_{m_i} - x_{l_i} \| \to 0$ as $i \to +\infty$. Then, because q is uniformly continuous, $\| q(x_{m_i}) - q(x_{l_i}) \| < \eta$, for i sufficiently large. However,

$$\| q(x_{m_i}) \| \quad \le \quad \| q(x_{m_i}) - q(x_{l_i}) \| + \| q(x_{l_i}) \| \quad \le \quad 2\eta. \tag{6.5}$$

Since (6.5) must hold for any η, $0 < \eta < \varepsilon$, we have a contradiction (e.g., try $\eta = \frac{\varepsilon}{4}$).

7. Comments on the algorithm. We next discuss some practical aspects of pattern search algorithms for linearly constrained problems. In this section we propose some stopping criteria for these algorithms, as well as examine the questions of estimating Lagrange multipliers and identifying the constraints active at a solution.

7.1. Stopping criteria. The stopping criterion that seems most natural to us is to halt the algorithm once Δ_k falls below some prescribed tolerance Δ_*. Equivalently, one can halt once the absolute length of the steps in the core pattern falls below some prescribed tolerance δ_*.

The following proposition concerning the correlation of stationarity and the size of Δ_k lends support to this choice of a stopping criterion. The result relates $\| q_k \|$ and the Δ_k at those steps where Δ_k is reduced (i.e., where back-tracking occurs); if we terminate the algorithm at such an iterate, then, if Δ_k is sufficiently small, $\| q_k \|$ will also be small. In the case of bound constraints, a similar result allows one to establish convergence for a pattern search algorithm for general nonlinearly constrained problems via inexact bound constrained minimization of the augmented Lagrangian [10]. For convenience, we assume that $\nabla f(x)$ is Lipschitz continuous. However, if we assume only that $\nabla f(x)$ is uniformly continuous on $L_\Omega(x_0)$, we can still establish a correlation between stationarity and the size of Δ_k.

PROPOSITION 7.1. *Suppose $\nabla f(x)$ is Lipschitz continuous on $L_\Omega(x_0)$ with Lipschitz constant C. There exists $c_{7.1} > 0$ for which the following holds. If x_k is an iterate at which there is an unsuccessful iteration, then*

$$\| q_k \|^2 \le c_{7.1} \Delta_k. \tag{7.1}$$

Proof. We need only consider the situation where $\eta = \| q_k \| > 0$. There are two cases to consider. First suppose $r_{6.2} \| q_k \|^2 \le \Delta_k$, where $r_{6.2}$ is the constant of the same name in Proposition 6.2. Then we immediately have

$$\| q_k \|^2 \le \Delta_k / r_{6.2}. \tag{7.2}$$

On the other hand, suppose $r_{6.2} \| q_k \|^2 > \Delta_k$. By Proposition 6.2, there exists $s_k^i \in \Delta_k \Gamma_k$ such that $(x_k + s_k^i) \in \Omega$ and

$$-g_k^T s_k^i \ge c_{6.2} \| q_k \| \| s_k^i \|. \tag{7.3}$$

Since iteration k is unsuccessful, it follows from Fig. 3.3 that $f(x_k + s_k^i) - f(x_k) \ge 0$ for all feasible $s_k^i \in \Delta_k \Gamma_k$. By the mean-value theorem, for some ξ in the line segment connecting x_k and $x_k + s_k^i$ we have

$$\begin{aligned} 0 &\le f(x_k + s_k^i) - f(x_k) \\ &= \nabla f(x_k)^T s_k^i + (\nabla f(\xi) - \nabla f(x_k))^T s_k^i \\ &\le -c_{6.2} \| q_k \| \| s_k^i \| + \| \nabla f(\xi) - \nabla f(x_k) \| \| s_k^i \|, \end{aligned}$$

where s_k^i is the step for which (7.3) holds. Thus

$$c_{6.2} \| q_k \| \le \| \nabla f(\xi) - \nabla f(x_k) \|.$$

Using the Lipschitz constant C for $\nabla f(x)$, we obtain

$$c_{6.2} \| q_k \| \le C \| \xi - x_k \| \le C \Gamma^* \Delta_k,$$

where Γ^* is the maximum norm of any column of the matrices in the set $\mathbf{\Gamma}$. Thus

$$c_{6.2} \| q_k \|^2 \leq g^* C \Gamma^* \Delta_k, \tag{7.4}$$

where g^* is the upper bound on $\nabla f(x)$. The proposition then follows from (7.2) and (7.4). $\square$

Remark. We conjecture that one can establish the estimate $\| q_k \| \leq c\Delta_k$ at unsuccessful steps. The appearance of $\| q_k \|^2$ rather than $\| q_k \|$ in (7.1) is a consequence of the appearance of η^2 in the hypotheses of Proposition 10.7, which in turn derives from the limitations of the way in which the latter proposition is proved.

May's algorithm [11], which is based on a difference approximation of feasible directions of descent, uses a difference approximation of local feasible descent in its stopping criterion. In connection with pattern search one could also attempt to do something similar, estimating $\nabla f(x)$ either by a difference approximation or a regression fit, and using this information in a stopping test. However, depending on the application, the simpler stopping criterion $\Delta_k < \Delta_*$ may be preferable; for instance, if the objective is believed to be untrustworthy in its accuracy, or if $f(x)$ is not available as a numerical value and only comparison of objective values is possible.

7.2. Identifying active constraints. Another practical issue is that of identifying active constraints, as in [2, 3, 4]. A desirable feature of an algorithm for linearly constrained minimization is the identification of active constraints in a finite number of iterations, that is, if the sequence $\{x_k\}$ converges to a stationary point x_*, then in a finite number of iterations the iterates x_k land on the constraints active at x_* and remain thereafter on those constraints.

As discussed in [8] for the case of bound constraints, there are several impediments to proving such results for pattern search algorithms and showing that ultimately the iterates will land on the active constraints and remain there. For algorithms such as those considered in [2, 3, 4], this is not a problem because the explicit use of the gradient impels the iterates to do so in the neighborhood of a constrained stationary point. However, pattern search methods do not have this information, and at this point it is not clear how to avoid the possibility that these algorithms take a purely interior approach to a point on the boundary. On the other hand, the kinship of pattern search methods and gradient projection methods makes us hopeful that we may be able to devise a suitable mechanism to ensure pattern search methods also identify the active constraints in a finite number of iterations.

7.3. Estimating Lagrange multipliers. Similar limitations pertain to estimating Lagrange multipliers as do to identifying active constraints. Pattern search methods do not use an explicit estimate of $\nabla f(x)$, and one does not obtain an estimate of the Lagrange multipliers for (1.1) from the usual workings of the algorithm. Some manner of post-optimality sensitivity analysis would be required to obtain estimates of the multipliers; again, either through difference estimates or regression estimates of $\nabla f(x)$.

By way of comparison, in May's algorithm one looks at both the cones tangent and polar to the nearby constraints and computes directional derivatives along generators for both of these cones. The directional derivatives associated with the normal cone yield multiplier estimates. (The authors are indebted to one of the referees for pointing this out and for suggesting that the same idea could be applied to our algorithm.)

In the algorithm we propose, we restrict attention to the behavior of $f(x)$ solely in feasible directions, ignoring the behavior of $f(x)$ in infeasible directions, which is precisely the information needed to compute multipliers. However, one could estimate multipliers along the lines of the calculations in May's algorithm, by estimating the active constraints at a stopping point x_k and computing directional derivatives in the directions of the generators of the associated estimate of the cone normal to the feasible region. (The computation of the generators of the requisite cones is discussed in the following section.)

For another way in which one can obtain information about multipliers from pattern search methods, see the augmented Lagrangian approach in [10].

8. Constructing patterns for problems with linear constraints. In this section we outline practical implementations of pattern search methods for linearly constrained minimization. The details will be the subject of future work. In the process we also show that under the assumption that A is rational, one can actually construct patterns with both the algebraic properties required in Section 3.1 and the geometric properties required in Section 3.5.

8.1. Remarks on the general case. We begin by showing that in general it is possible to find rational generators for the cones $K^\circ(x, \varepsilon)$. By clearing denominators we then obtain the integral vectors for $\mathbf{\Gamma}$ as required in Section 3.1. The construction is an elaboration of the proof that polyhedral cones are finitely generated (see [16], for instance). The proof outlines an algorithm for the construction of generators of cones. Given a cone K we will use V to denote a matrix whose columns are generators of K:

$$K = \{ \ x \ \mid \ x = V\lambda, \ \lambda \geq 0 \ \} .$$

PROPOSITION 8.1. *Suppose K is a cone with rational generators V. Then there exists a set of rational generators for K°.*

Proof. Suppose $w \in K^\circ$; then $(w, v) \leq 0$ for all $v \in K$. Let $v = V\lambda$, $\lambda \geq 0$. Then

$$(w \ , \ v) = \left(P_{\mathcal{N}(V^T)} w + P_{\mathcal{N}(V^T)^\perp} w \ , \ V\lambda \right) \leq 0,$$

where $P_{\mathcal{N}(V^T)}$ and $P_{(\mathcal{N}(V^T))^\perp}$ are the projections onto the nullspace $\mathcal{N}(V^T)$ of V^T and its orthogonal complement $\mathcal{N}(V^T)^\perp$, respectively. Since $\mathcal{N}(V^T)^\perp$ is the same as the range $\mathcal{R}(V)$ of V, we have

$$(w \ , \ v) = \left(P_{\mathcal{R}(V)} w \ , \ V\lambda \right) \leq 0.$$

Let N and R be rational bases for $\mathcal{N}(V^T)$ and $\mathcal{R}(V)$ respectively; these can be constructed, for instance, via reduction to row echelon form since V is rational.

Let $\{p_1, \cdots, p_t\}$ be a rational positive basis for $\mathcal{N}(V^T)$. Such a positive basis can be constructed as follows. If N is $n \times r$ then if Π is a rational positive basis (with t elements) for $\mathbf{R}^r$ (e.g., $\Pi = [I \ \ -I]$), then $N\Pi$ is a rational positive basis for $\mathcal{N}(V^T)$.

Meanwhile, if R is a rational basis for $\mathcal{R}(V)$, then for some z we have

$$P_{\mathcal{R}(V)} w = Rz,$$

whence

$$(w \ , \ v) = (Rz \ , \ V\lambda) \leq 0.$$

Since $\lambda^T V^T R z \leq 0$ for all $\lambda \geq 0$, it follows that $V^T R z \leq 0$. Let $e = (1, \cdots, 1)^T$ and consider

$$C = \left\{ \ z \ \mid \ V^T R z \leq 0, \ e^T V^T R z \geq -1 \ \right\} .$$

Since C is convex and compact, it is the convex hull of its extreme points $\{c_1, \cdots, c_s\}$. Furthermore, note that the extreme points of C will define a set of generators for the cone $\left\{ z \mid V^T R z \leq 0 \right\}$. The extreme points of C are also rational since $V^T R$ is rational; the extreme points will be solutions to systems of equations with rational coefficients. These extreme points, which are the vertices of the polyhedron C, can be computed by any number of vertex enumeration techniques (e.g., see [1] and the references cited therein).

Returning to $w \in K^\circ$, we see that we can express w as a positive linear combination of the vectors $\{p_1, \cdots, p_t, c_1, \cdots, c_s\}$. Moreover, by construction the latter vectors are rational. □

8.2. The nondegenerate case. As we have seen, the construction of sets of generators for cones is non-trivial and is related to the enumeration of vertices of polyhedra. However, in the case of nondegeneracy—the absence of any point on the boundary at which the set of binding constraints is linearly dependent—we can compute the required generators in a straightforward way. This case is handled in [11] by using the QR factorization to derive the search directions. Because we require rational search directions, we use the LU factorization (reduction to row echelon form, to be more precise) since the latter can be done in rational arithmetic.

The following proposition shows that once we have identified a cone $K(x_k, \delta)$ with a linearly independent set of generators, we can construct generators for all the cones $K(x_k, \varepsilon)$, $0 \leq \varepsilon \leq \delta$.

PROPOSITION 8.2. *Suppose that for some δ, $K(x, \delta)$ has a linearly independent set of rational generators V. Let N be a rational positive basis for the nullspace of V^T.*

Then for any ε, $0 \leq \varepsilon \leq \delta$, a set of rational generators for $K^\circ(x, \varepsilon)$ can be found among the columns of N, $V(V^T V)^{-1}$, and $-V(V^T V)^{-1}$.

Proof. Given $x \in \Omega$ and $\delta > 0$, let $K = K(x, \delta)$. Suppose $w \in K^\circ$; then $(w, v) \leq 0$ for all $v \in K$. Let $v = V\lambda$, $\lambda \geq 0$. Since V has full column rank, we have

$$(w, v) = ((I - V(V^T V)^{-1} V^T) w + V(V^T V)^{-1} V^T w, V\lambda) \leq 0$$

or $(V^T w, \lambda) \leq 0$ for all $\lambda \geq 0$. Let $\xi = V^T w$; then we have $(\xi, \lambda) \leq 0$ for all $\lambda \geq 0$, so $\xi \leq 0$.

The matrix N is a positive basis for the range of $I - V(V^T V)^{-1} V^T$, since the latter subspace is the same as the nullspace of V^T. Then any $w \in K^\circ$ can be written in the form

$$w = N\zeta - V(V^T V)^{-1} \xi$$

where $\zeta \geq 0$ and $\xi \geq 0$. Thus the columns of N and $-V(V^T V)^{-1}$ are a set of generators for K°.

Moreover, for $\varepsilon < \delta$ we obtain $\tilde{K} = K(x, \varepsilon)$ by dropping generators from V. Without loss of generality we will assume that we drop the first r columns of V, where V has p columns. Then consider $w \in \tilde{K}^\circ$. Proceeding as before, we obtain $(V^T w, \lambda) \leq 0$ for all $\lambda \geq 0$, $\lambda_1, \cdots, \lambda_r = 0$. If we once again define $\xi = V^T w$, then we see that $\xi_{r+1}, \cdots, \xi_p \leq 0$, while $\xi_1, \cdots, \xi_r$ are unrestricted in sign. Hence we obtain a set of generators for $\tilde{K}^\circ$ from the columns of N, the first r columns of $V(V^T V)^{-1}$ and their negatives, and the last $p - r$ columns of $-V(V^T V)^{-1}$. □

Proposition 8.2 leads to the following construction of patterns for linearly constrained minimization. Under the assumption of nondegeneracy, we know there exists

ε^* such that if $0 \le \varepsilon \le \varepsilon^*$, then $K(x,\varepsilon)$ has a linearly independent set of generators. If we knew this ε^*, it would be a convenient choice for the ε^* required in Section 3.5. The following algorithm implicitly estimates ε^*: it conducts what amounts to a safeguarded backtracking on ε at each iteration to find a value of ε_k for which $K(x_k,\varepsilon_k)$ has a linearly independent set of generators.

Given ε_* independent of k, choose $\varepsilon_k \ge \varepsilon_*$. Then

1. Define the cone $K(x_k,\varepsilon_k)$ as in Section 3.5.
2. Let V represent the matrix whose columns are the generators $\nu_i^\ell(x_k,\varepsilon_k)$ and $\nu_i^u(x_k,\varepsilon_k)$ of $K(x_k,\varepsilon_k)$ (defined in (3.3)–(3.4)). Determine whether or not V has full column rank. If so, go to Step 3. Otherwise, reduce ε_k just until $|I_\ell(x_k,\varepsilon_k)| + |I_u(x_k,\varepsilon_k)|$ is decreased. Return to Step 1.
3. Construct a rational positive basis N for the range of $I - V(V^TV)^{-1}V^T$. This can be done via reduction to row echelon form, or simply by taking the columns of the matrices $\pm\left(I - V(V^TV)^{-1}V^T\right)$.
4. Form the matrix $\Gamma_k = [\, N \;\; V(V^TV)^{-1} \;\; -V(V^TV)^{-1}\,]$.

Under the assumption of nondegeneracy, ε_k will remain bounded away from 0 as a function of k, implicitly giving us the ε^* introduced in Section 3.5.2.

This construction also shows that we may reasonably expect to arrange that r_k, the number of columns of Γ_k defined in Section 3.1, to be at most $2n$. Suppose V has rank r. Then the nullspace of V has dimension $n-r$, so we can find a positive basis N for the nullspace with as few as $n-r+1$ elements (or 0 elements, if $n=r$). At the same time, $V(V^TV)^{-1}$ has r columns, so we can arrange Γ_k to have as few as $(n-r+1)+2r = n+r-1$ columns, if $r<n$, or $2r$ elements, if $r=n$. In either case Γ_k has at most $2n$ columns.

8.3. The case of bound constraints. Matters simplify enormously in the case of bound constraints, previously considered in [8]. We will briefly discuss the specialization to bound constrained minimization and in the process sharpen the results in [8].

In the case of bound constraints we have

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & l \le x \le u. \end{array}$$

Again, we allow the possibility that some of the variables are unbounded either above or below by permitting $\ell_j, u_j = \pm\infty$, $j \in \{1,\cdots,n\}$.

In the case of bound constraints we know *a priori* the possible generators of the cones $K(x,\varepsilon)$ and $K^\circ(x,\varepsilon)$. For any $x \in \Omega$ and any $\varepsilon > 0$ the cone $K(x,\varepsilon)$ is generated by some subset of the coordinate vectors $\pm e_i$. If $K(x,\varepsilon)$ is generated by $\nu_{i_1},\cdots,\nu_{i_r}$, where $\nu_{i_j} \in \{e_{i_j}, -e_{i_j}\}$, then $K^\circ(x,\varepsilon)$ is generated by the set $-\nu_{i_1},\cdots,-\nu_{i_r}$ together with a positive basis for the orthogonal complement of the space spanned by $\nu_{i_1},\cdots,\nu_{i_r}$. This orthogonal complement simply corresponds to the remaining coordinate directions.

This simplicity allows us to prescribe in advance patterns that work for all $K(x,\varepsilon)$. In [8] we gave the prescription $\Gamma_k = [I \; -I]$. This choice, independent of k, includes generators for all possible $K^\circ(x,\varepsilon)$. However, if not all the variables are bounded, then one can make a choice of Γ_k that is independent of k but more parsimonious in the number of directions. Let $x_{i_1},\cdots,x_{i_r}$ be the variables with either a lower or upper bound; then Γ_k should include the coordinate vectors $\pm e_{i_1},\cdots,\pm e_{i_r}$ together with a positive basis for the orthogonal complement of the linear span of $e_{i_1},\cdots,e_{i_r}$; a positive basis for the orthogonal complement can have as few as $(n-r)+1$ elements.

The choice of $\Gamma_k = [I - I]$ in [8] requires, in the worst case, $2n$ objective evaluations per iteration. The more detailed analysis given here leads to a reduction in this cost if not all the variables are bounded. If only $r < n$ variables are bounded, then we can find an acceptable pattern containing as few as $2r + ((n-r)+1) = n + r + 1$ points.

Finally, note that if general linear constraints are present but A has full row rank (i.e., there are no more than n constraints and they are all linearly independent), then one can carry out a construction similar to that for bound constraints.

9. Conclusions. We have introduced pattern search algorithms for solving problems with general linear constraints. We have shown that under mild assumptions we can guarantee global convergence of pattern search methods for linearly constrained problems to a Karush-Kuhn-Tucker point. As in the case of unconstrained minimization, pattern search methods for linearly constrained problems accomplish this without explicit recourse to the gradient or the directional derivative. In addition, we have outlined particular instances of such algorithms and shown how the general approach can be greatly simplified when the only constraints are bounds on the variables. The effectiveness of these techniques will be the subject of future work.

10. Appendix: results concerning the geometry of polyhedra. We need a number of results concerning the geometry of polyhedra for the proofs of Section 6. We begin with a classical result on the structure of finitely generated cones.

THEOREM 10.1. *Let C be a finitely generated convex cone in $\mathbf{R}^n$. Then C is the union of finitely many finitely generated convex cones each having a linearly independent set of generators chosen from the generators of C.*

Proof. See Theorem 4.17 in [16]. □

COROLLARY 10.2. *Let C be a finitely generated convex cone in $\mathbf{R}^n$ with generators $\{v_1, \cdots, v_r\}$. Then there exists $c_{10.2} > 0$, depending only on $\{v_1, \cdots, v_r\}$, such that any $z \in C$ can be written in the form $z = \sum_{i=1}^{r} \lambda_i v_i$ with $\lambda \geq 0$ and $\| \lambda \| \leq c_{10.2} \| z \|$.*

Proof. Theorem 10.1 says that we can write z in the form $z = \sum_{j=1}^{r_z} \lambda_{i_j} v_{i_j}$ where $r_z \leq r$, $\lambda_{i_j} \geq 0$, and the matrix $V_z = [v_{i_1} \cdots v_{i_{r_z}}]$ has full column rank. The full column rank of V_z means that the induced linear transformation is one-to-one, so if V_z^+ is the pseudoinverse of V_z, then $(\lambda_{i_1}, \cdots, \lambda_{i_{r_z}})^T = V_z^+ z$. If we define λ via

$$\lambda_i = \begin{cases} \lambda_{i_j} & \text{if } i = i_j, \\ 0 & \text{otherwise,} \end{cases}$$

then $\lambda \geq 0$, $z = V\lambda$, and $\| \lambda \| \leq \| V_z^+ \| \| z \|$. Since the matrix V_z is drawn from a finite set of possibilities (e.g., the set of all subsets of $\{v_1, \cdots, v_r\}$), we can find the desired constant $c_{10.2}$, independent of z. □

Let C be a closed convex cone in $\mathbf{R}^n$ with vertex at the origin and let C° be its polar. Given any vector z, we will denote by z_C and z_{C° the projections of z onto the cones C and C°, respectively. The classical polar decomposition [13, 17] allows us to express z as

$$z = z_C + z_{C^\circ},$$

where $(z_C, z_{C^\circ}) = 0$.

PROPOSITION 10.3. *Suppose the cone C is generated by $\{v_1, \cdots, v_r\}$. Then there exists $c_{10.3} > 0$, depending only on $\{v_1, \cdots, v_r\}$, such that for any z for which $z_C \neq 0$,*

$$\max_{1 \leq i \leq r} \frac{z^T v_i}{\| v_i \|} \geq c_{10.3} \| z_C \|.$$

Proof. By Corollary 10.2, we have $c_{10.2} > 0$, depending only on $\{v_1, \cdots, v_r\}$, such that we can write z_C as $z_C = \sum_{i=1}^r \lambda_i v_i$, with $\| \lambda \| \leq c_{10.2} \| z_C \|$ and $\lambda \geq 0$. Then

$$z^T z_C = \sum_{i=1}^r \lambda_i z^T v_i,$$

so for some i we must have

$$\lambda_i z^T v_i \geq \frac{1}{r} z^T z_C = \frac{1}{r} \| z_C \|^2.$$

Since $\| \lambda \| \leq c_{10.2} \| z_C \|$ and $\| z_C \| \neq 0$, we obtain

$$z^T v_i \geq \frac{1}{r} \frac{1}{c_{10.2}} \| z_C \|.$$

If we let

$$v^* = \max_{1 \leq i \leq r} \| v_i \|$$

we obtain

$$z^T v_i \geq \frac{1}{r} \frac{1}{c_{10.2}} \frac{1}{v^*} \| v_i \| \| z_C \|$$

and the desired result, with $c_{10.3} = (r c_{10.2} v^*)^{-1}$. □

For the polyhedron defining the feasible region of (1.1), we have the following.

COROLLARY 10.4. *There exists $c_{10.4} > 0$, depending only on A, for which the following holds. For any $x \in \Omega$ and $\varepsilon \geq 0$, let $K = K(x, \varepsilon)$. Then for any z for which $z_{K^\circ} \neq 0$,*

$$\max_{1 \leq i \leq r} \frac{z^T v_i}{\| v_i \|} \geq c_{10.4} \| z_{K^\circ} \|,$$

where $\{v_1, \cdots, v_r\}$ are the generators of $K^\circ(x, \varepsilon)$ required in Section 3.5.2 to be in $\mathbf{\Gamma}$.

Proof. The corollary follows from the observation that since $K(x, \varepsilon)$ is generated by subsets of the rows of A, $K(x, \varepsilon)$ can be one of only a finite number of possible cones. Consequently $K^\circ(x, \varepsilon)$ will also be one of only a finite number of possible cones. Applying Proposition 10.3 to each of these latter cones in turn (with the generators in $\mathbf{\Gamma}$ for $K^\circ(x, \varepsilon)$) and taking the minimum yields the corollary. □

Let

$$a^* = \max_{1 \leq i \leq m} \{\| a_i \|\}.$$

Then we have the following straightforward proposition.

PROPOSITION 10.5. *For any $x \in \Omega$ and $\varepsilon \geq 0$, we have*

$$\ell_i \;\leq\; a_i^T x \;\leq\; \ell_i + \varepsilon \| a_i \| \;\leq\; \ell_i + \varepsilon a^* \quad \text{for } i \in I_\ell(x, \varepsilon), \tag{10.1}$$

$$u_i - \varepsilon a^* \;\leq\; u_i - \varepsilon \| a_i \| \;\leq\; a_i^T x \;\leq\; u_i \quad \text{for } i \in I_u(x, \varepsilon), \tag{10.2}$$

where $I_\ell(x,\varepsilon)$ and $I_u(x,\varepsilon)$ are the index sets defined in (3.1)–(3.2).

Proof. A simple calculation shows that the distance from any point x to the affine subspace defined by $a_i^T z = b$ is $\left| b - a_i^T x \right| / \| \, a_i \, \|$. Thus, if the distance from x to $a_i^T z = b$ is no more than ε, then

$$b - \varepsilon \| \, a_i \, \| \leq a_i^T x \leq b + \varepsilon \| \, a_i \, \|.$$

Then (10.1) and (10.2) follow from the fact that $x \in \Omega$ and the definitions of $I_\ell(x,\varepsilon)$ and $I_u(x,\varepsilon)$. $\square$

Despite its unpromising appearance, the following result is extremely useful, as it relates the local geometry of Ω (as manifest in $K(x,\varepsilon)$) to the global geometry of Ω (as manifest in the projection P_Ω).

PROPOSITION 10.6. *There exists $c_{10.6} > 0$ such that for any $x \in \Omega$, $\varepsilon \geq 0$, and $w \in \mathbf{R}^n$,*

$$\| \, (x+w) - P_\Omega(x+w) \, \|^2 \geq \| \, P_{K(x,\varepsilon)} w \, \|^2 - c_{10.6} \; \varepsilon \; \| \, P_{K(x,\varepsilon)} w \, \|.$$

Proof. $P_\Omega(x+w)$ is the solution y of the convex quadratic program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} \| \, y - (x+w) \, \|^2 \\ \text{subject to} & \ell \leq Ay \leq u. \end{array} \tag{10.3}$$

The dual of (10.3) is the following program in (z, μ_1, μ_2):

$$\begin{array}{ll} \text{maximize} & \frac{1}{2} \| \, z - (x+w) \, \|^2 - \mu_1^T (u - Az) - \mu_2^T (Az - \ell) \\ \text{subject to} & z - (x+w) + A^T \mu_1 - A^T \mu_2 = 0 \\ & \mu_1, \mu_2 \geq 0. \end{array} \tag{10.4}$$

The proposition will follow from a felicitous choice of (z, μ_1, μ_2) for the dual.

Given $x \in \Omega$ and $\varepsilon \geq 0$, let $K = K(x,\varepsilon)$ and consider the polar decomposition $w = w_K + w_{K^\circ}$. We can write

$$w_K = A^T \mu_1 - A^T \mu_2$$

where $\mu_1, \mu_2 \geq 0$, and the only non-zero components of μ_1, μ_2 correspond to the generators of $K(x,\varepsilon)$, which are the outward pointing normals to the constraints within distance ε of x. More precisely,

$$\mu_1^i \neq 0 \quad \text{only if } i \in I_u(x,\varepsilon), \qquad \mu_2^i \neq 0 \quad \text{only if } i \in I_\ell(x,\varepsilon). \tag{10.5}$$

Furthermore, by Corollary 10.2 we can choose μ_1, μ_2 in such a way that there exists $c_{10.2} > 0$, depending only on A, such that

$$\| \, \mu_1 \, \| + \| \, \mu_2 \, \| \leq c_{10.2} \| \, w_K \, \|. \tag{10.6}$$

Meanwhile, let $z = x + w_{K^\circ}$. Then

$$w = w_K + w_{K^\circ} = z - x + A^T \mu_1 - A^T \mu_2,$$

so (z, μ_1, μ_2) is feasible for the dual (10.4). Since $y = P_\Omega(x+w)$ is feasible for the primal (10.3), by duality we have

$$\begin{aligned} & \tfrac{1}{2} \| \, (x+w) - P_\Omega(x+w) \, \|^2 \\ & \geq \tfrac{1}{2} \| \, (x+w) - z \, \|^2 - \mu_1^T (u - Az) - \mu_2^T (Az - \ell) \\ & = \tfrac{1}{2} \| \, w_K \, \|^2 - \mu_1^T (u - Ax) - \mu_2^T (Ax - \ell) + (A^T \mu_1 - A^T \mu_2)^T w_{K^\circ}. \end{aligned}$$

Since $w_K = A^T\mu_1 - A^T\mu_2$ and $(w_K, w_{K^\circ}) = 0$, the latter expression reduces to

$$(10.7) \quad \tfrac{1}{2}\| \, (x+w) - P_\Omega(x+w) \, \|^2 \geq \tfrac{1}{2}\| \, w_K \, \|^2 - \mu_1^T(u - Ax) - \mu_2^T(Ax - \ell).$$

Now, in light of (10.5) and Proposition 10.5 we have

$$\mu_1^T(u - Ax) + \mu_2^T(Ax - \ell) \leq a^* \, \varepsilon \, \| \, \mu_1 \, \|_1 + a^* \, \varepsilon \, \| \, \mu_2 \, \|_1 \leq a^*\sqrt{n} \, \varepsilon \, (\| \, \mu_1 \, \| + \| \, \mu_2 \, \|).$$

Applying (10.6) we obtain

$$\mu_1^T(u - Ax) + \mu_2^T(Ax - \ell) \leq c_{10.2} \; a^*\sqrt{n} \, \varepsilon \, \| \, w_K \, \|.$$

Substituting this into (10.7) yields

$$\tfrac{1}{2}\| \, (x+w) - P_\Omega(x+w) \, \|^2 \geq \tfrac{1}{2}\| \, w_K \, \|^2 - c_{10.2} \; a^*\sqrt{n} \, \varepsilon \, \| \, w_K \, \|,$$

which is the desired result, with $c_{10.6} = 2c_{10.2}a^*\sqrt{n}$. □

The consequence of Proposition 10.6 of utility to us is the following. It says that if $x \in \Omega$ is close to $\partial\Omega$ and the step from x to $P_\Omega(x+w)$ is sufficiently long, then w cannot be "too normal" to $\partial\Omega$ near x.

Proposition 10.7. *Given $\gamma > 0$, there exist $r_{10.7} > 0$ and $c_{10.7} > 0$, depending only on A and γ, such that if $\eta > 0$, $x \in \Omega$, $0 \leq \varepsilon \leq r_{10.7}\eta^2$, $\| \, w \, \| \leq \gamma$, and $\| \, P_\Omega(x+w) - x \, \| \geq \eta$, then*

$$\| \, P_{K^\circ(x,\varepsilon)} w \, \| \geq c_{10.7} \; \| \, P_\Omega(x+w) - x \, \|.$$

Proof. Given $x \in \Omega$ and $\varepsilon \geq 0$, let $K = K(x, \varepsilon)$ and consider the polar decomposition $w = w_K + w_{K^\circ}$. Let $q = P_\Omega(x+w) - x$. We have

$$\| \, w \, \|^2 = \| \, w_K \, \|^2 + \| \, w_{K^\circ} \, \|^2 = \| \, (w - q) + q \, \|^2 = \| \, w - q \, \|^2 + 2\,(w - q \; , \; q) + \| \, q \, \|^2.$$

We know that $(z - P_\Omega(z) \; , \; P_\Omega(z) - y) \geq 0$ for all $y \in \Omega$ from the properties of the projection P_Ω [17]. Choosing $z = x + w$ and $y = x$ we obtain $(w - q \; , \; q) \geq 0$, so

$$\| \, w_K \, \|^2 + \| \, w_{K^\circ} \, \|^2 \geq \| \, w - q \, \|^2 + \| \, q \, \|^2.$$

From Proposition 10.6 we obtain

$$\| \, w_K \, \|^2 + \| \, w_{K^\circ} \, \|^2 \geq \| \, w_K \, \|^2 - c_{10.6} \; \varepsilon \, \| \, w_K \, \| + \| \, q \, \|^2.$$

Using the hypothesis that $\| \, w \, \| \leq \gamma$, we obtain

$$\| \, w_{K^\circ} \, \|^2 \geq -c_{10.6} \; \varepsilon\gamma + \| \, q \, \|^2.$$

Let

$$r_{10.7} = \frac{3}{4}\frac{1}{\gamma}\frac{1}{c_{10.6}}.$$

Then, if $\varepsilon \geq 0$ satisfies $\varepsilon \leq r_{10.7}\eta^2$, we have

$$\| \, w_{K^\circ} \, \|^2 \geq \| \, q \, \|^2/4.$$

Taking square roots yields the proposition, with $c_{10.7} = 1/2$ □

As we noted at the introduction of $K^\circ(x,\varepsilon)$, we can proceed from x along all directions in $K^\circ(x,\varepsilon)$ for a distance $\delta > 0$, depending only on ε, and still remain inside the feasible region. The following proposition is the formal statement of this observation.

PROPOSITION 10.8. *Suppose $\varepsilon > 0$ satisfies $\varepsilon \le h/2$, where h is defined by (6.1). Then for any $x \in \Omega$, if $w \in K^\circ(x,\varepsilon)$ and $\| w \| \le \varepsilon/2$, then $(x+w) \in \Omega$.*

Proof. Consider any index $i \in \{1, \cdots, m\}$. We will show that $x + w$ is feasible with respect to the ith constraint.

If $x \notin \partial\Omega_{\ell_i}(\varepsilon) \cup \partial\Omega_{u_i}(\varepsilon)$, then $\ell_i + \varepsilon\| a_i \| < a_i^T x < u_i - \varepsilon\| a_i \|$, so

$$a_i^T x + a_i^T w \ge \ell_i + \varepsilon\| a_i \| - \| a_i \|\| w \| \ge \ell_i + (\varepsilon/2)\| a_i \| \ge \ell_i$$

and

$$a_i^T x + a_i^T w \le u_i - \varepsilon\| a_i \| + \| a_i \|\| w \| \le u_i - (\varepsilon/2)\| a_i \| \le u_i.$$

On the other hand, suppose $x \in \partial\Omega_{\ell_i}(\varepsilon) \cup \partial\Omega_{u_i}(\varepsilon)$. There are three cases to consider. First suppose $x \in \Omega_{\ell_i}(\varepsilon)$ and $x \in \partial\Omega_{u_i}(\varepsilon)$. Since $\varepsilon < h/2$, this means that $\ell_i = u_i$ (i.e., the constraint is an equality constraint). Then, if $w \in K^\circ(x,\varepsilon)$, we have both $(w\ ,\ -a_i) \le 0$ and $(w\ ,\ a_i) \le 0$, so $(w\ ,\ a_i) = 0$. Thus

$$\ell_i = a_i^T x + a_i^T w = u_i.$$

Next suppose $x \in \partial\Omega_{\ell_i}(\varepsilon)$ but $x \notin \partial\Omega_{u_i}(\varepsilon)$. If $w \in K^\circ(x,\varepsilon)$, we have $(-a_i, w) \le 0$. Applying Proposition 10.5 we obtain

$$\ell_i \le a_i^T x + a_i^T w \le \ell_i + \varepsilon\| a_i \| + \| a_i \|\| w \| \le \ell_i + (3\varepsilon/2)\| a_i \| \le u_i.$$

Finally, if $x \in \partial\Omega_{u_i}(\varepsilon)$ but $x \notin \partial\Omega_{\ell_i}(\varepsilon)$, then, if $w \in K^\circ(x,\varepsilon)$, $(a_i, w) \le 0$, so

$$u_i \ge a_i^T x + a_i^T w \ge u_i - \varepsilon\| a_i \| - \| a_i \|\| w \| \ge u_i - (3\varepsilon/2)\| a_i \| \ge \ell_i.$$

Thus $(x+w)$ satisfies the constraints for all $i \in \{1, \cdots, m\}$, so $(x+w) \in \Omega$. ∎

Acknowledgments. The authors thank Robert Mifflin for bringing to our attention the work cited in Section 2 and for providing us with a copy of Jerrold May's thesis. We also thank Bob Bixby and Bill Cook for directing us to the literature on vertex enumeration, to which we refer in Section 8, and David Avis and David Bremner for explaining to us the practical issues in vertex enumeration.

Finally, the authors thank the referees both for their careful reading of earlier drafts of this paper and for their many helpful comments. The results are stronger and the presentation is both clearer and more accurate as a consequence of their questions and suggestions.

REFERENCES

[1] D. M. Avis and K. Fukuda, *A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra*, Discrete and Computational Geometry, 8 (1992), pp. 295–313.
[2] J. V. Burke and J. J. Moré, *On the identification of active constraints*, SIAM Journal on Numerical Analysis, 25 (1988), pp. 1197–1211.
[3] ———, *Exposing constraints*, SIAM Journal on Optimization, 4 (1994), pp. 573–595.
[4] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Mathematical Programming, 39 (1987), pp. 93–116.

[5] C. Davis, *Theory of positive linear dependence*, American Journal of Mathematics, 76 (1954), pp. 733–746.

[6] J. C. Dunn, *Global and asymptotic convergence rate estimates for a class of projected gradient processes*, SIAM Journal on Control and Optimization, 19 (1981), pp. 368–400.

[7] R. Hooke and T. A. Jeeves, *Direct search solution of numerical and statistical problems*, Journal of the Association for Computing Machinery, 8 (1961), pp. 212–229.

[8] R. M. Lewis and V. J. Torczon, *Pattern search algorithms for bound constrained minimization*, Tech. Rep. 96–20, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, Virginia 23681–2199, March 1996. To appear in SIAM Journal on Optimization.

[9] ———, *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.

[10] ———, *A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, Tech. Rep. 98–31, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, VA 23681–2199, July 1998. In revision for SIAM Journal on Optimization. Available at `http://www.icase.edu/library/reports/rdp/1998.html#98-31`.

[11] J. H. May, *Linearly Constrained Nonlinear Programming: A Solution Method That Does Not Require Analytic Derivatives*, PhD thesis, Yale University, December 1974.

[12] R. Mifflin, *A superlinearly convergent algorithm for minimization without evaluating derivatives*, Mathematical Programming, 9 (1975), pp. 100–117.

[13] J. J. Moreau, *Décomposition orthgonale d'un espace hilbertien selon deux cônes mutuellement polaires*, Comptes Rendus de l'Académie des Sciences de Paris, 255 (1962), pp. 238–240.

[14] V. Torczon, *On the convergence of the multidirectional search algorithm*, SIAM Journal on Optimization, 1 (1991), pp. 123–145.

[15] ———, *On the convergence of pattern search algorithms*, SIAM Journal on Optimization, 7 (1997), pp. 1–25.

[16] J. van Tiel, *Convex Analysis: An Introductory Text*, John Wiley & Sons, 1984.

[17] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, 1971, pp. 237–424.