Pattern Search Methods for Linearly Constrained Minimization

Robert Michael Lewis
ICASE

Virginia Torczon
College of William and Mary

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, VA

Operated by Universities Space Research Association

January 1998

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. 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.

Subject classification. Applied and Numerical Mathematics

1. Introduction. This paper continues the line of development in [5, 6, 11] and extends pattern search algorithms to optimization problems with linear constraints:

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

where $f : I\!\!R^n \to I\!\!R$, $x \in I\!\!R^n$, $A \in \mathbf{Q}^{m \times n}$, $\ell, u \in I\!\!R^m$, 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. We also do not attempt to estimate Lagrange multipliers.

As with pattern search methods for bound constrained minimization [5], the pattern of points over which we must search in the worst case will, when we are close to the boundary, conform to the geometry of the boundary. The general idea, which also applies to unconstrained minimization [6], 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 §8.3 we will sharpen the results in [5]). 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 non-degeneracy, 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 [7], which extended the derivative-free algorithm for unconstrained minimization in [8] 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 [7]. As our analysis will indicate, this is unavoidable if one wishes to be assured of capturing any possible feasible descent in f using only values of f at feasible points (as would be the case in a derivative-free feasible point minimization method).

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 binding constraints are never linearly dependent — i.e., non-degeneracy. 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.

Notation. We denote by **R**, **Q**, **Z**, and **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 are assumed to be the Euclidean norm. We will denote the gradient of the objective by $g(x)$.

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

$$\Omega = \left\{\, x \in I\!\!R^n \mid \ell \leq Ax \leq u \,\right\}.$$

Given a convex cone K we denote its polar cone by K°; K° is the set of $v \in I\!\!R^n$ 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 by defining the general pattern search method for the linearly constrained problem (1.1).

3.1. The Pattern. The pattern for linearly constrained minimization is defined in a way that is only slightly less flexible than for patterns in the unconstrained case. In [11], 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 [4] and multidirectional search (MDS) [10]. In the case of bound constrained problems [5], we introduced restrictions on the pattern itself rather than on B and C_k independently, but maintained the artifice 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. 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 specify p_k in §3.5; for now we simply note that $p_k > n + 1$. There is no upper bound on p_k. We partition the generating matrix into components

$$(3.1) \qquad\qquad P_k = [\ \Gamma_k \quad L_k \quad].$$

We require that $\Gamma_k \in \mathbf{Z}^{n \times r_k}$ belongs to a finite set of matrices $\mathbf{\Gamma}$ with certain geometrical properties described in §3.5, and that $L_k \in \mathbf{Z}^{n \times (p_k - r_k)}$ contains at least one column, a column of zeroes. The inclusion of a column of zeroes is simply a formalism to allow for a zero step, i.e., $x_{k+1} = x_k$. Again, we will fully specify r_k in §3.5, but for now we note that $n + 1 \le r_k < p_k$.

Given $\Delta_k \in I\!\!R$, $\Delta_k > 0$, we define a *trial step* s_k^i to be any vector of the form $s_k^i = \Delta_k c_k^i$, where 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.

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. 3.1, 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. Note that the

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 $f(x_k + s_k) < f(x_k)$.

FIG. 3.1. *Hypotheses on the result of the linearly constrained exploratory moves.*

only change from the unconstrained case is the requirement that the iterates must be feasible.

3.3. The Generalized Pattern Search Method. Fig. 3.2 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.

3.4. The Updates. Fig. 3.3 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.*

3

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. 3.2. *The general pattern search method for linearly constrained problems.*

As is characteristic of pattern search methods, a step need only yield *simple* decrease, as opposed to *sufficient* decrease, in order to be acceptable.

Let $\tau \in \mathbf{Q}$, $\tau > 1$, and $\{w_0, w_1, \cdots, w_L\} \subset \mathbf{Z}$, $w_0 < 0$, and $w_i \geq 0$, $i = 1, \cdots, L$. Let $\theta = \tau^{w_0}$, and $\lambda_k \in \Lambda = \{\tau^{w_1}, \cdots, \tau^{w_L}\}$.

    a) If $f(x_k + s_k) \geq f(x_k)$, then $\Delta_{k+1} = \theta \Delta_k$.

    b) If $f(x_k + s_k) < f(x_k)$, then $\Delta_{k+1} = \lambda_k \Delta_k$.

FIG. 3.3. *Updating Δ_k.*

The conditions on θ and Λ ensure that $0 < \theta < 1$ and $\lambda_i \geq 1$ for all $\lambda_i \in \Lambda$. Thus, if an iteration is successful it may be possible to increase the step length parameter Δ_k, but Δ_k is not allowed to decrease. These conditions are identical to those for the unconstrained case.

3.5. Geometrical restrictions on the pattern. In the case of linearly constrained minimization the pattern P_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

$$A_{\ell_i} = \left\{ \; x \; | \; a_i^T x = \ell_i \; \right\}$$
$$A_{u_i} = \left\{ \; x \; | \; a_i^T x = u_i \; \right\}.$$

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

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

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

(3.2) $$I_\ell(x, \varepsilon) = \left\{ \; i \; | \; x \in \partial\Omega_{\ell_i}(\varepsilon) \; \right\}$$

(3.3) $$I_u(x, \varepsilon) = \left\{ \; i \; | \; x \in \partial\Omega_{u_i}(\varepsilon) \; \right\}.$$

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

(3.4)
$$\nu_{\ell_i}(x, \varepsilon) = -a_i$$

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

(3.5)
$$\nu_{u_i}(x, \varepsilon) = a_i.$$

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

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

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 $K^\circ(x, \varepsilon) = I\!\!R^n$. 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. 3.4, 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 $K^\circ(x_k, \varepsilon)$, $0 \leq \varepsilon < \varepsilon^*$, for some $\varepsilon^* > 0$ that is independent of k.

In §3.1 we required Γ_k to be one of a finite set of integral matrices Γ that is independent of k. Thus Γ will contain generators for all of the cones $K^\circ(x_k, \varepsilon)$, $0 \leq \varepsilon < \varepsilon^*$. 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 Γ 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 in §8. However, we will see that the construction is computationally tractable, and in many cases is not particularly difficult.

If x_k is "far" from the boundary in the sense that $K(x_k, \varepsilon) = \emptyset$, then $K^\circ(x_k, \varepsilon) = I\!\!R^n$ and a set of generators for $K^\circ(x_k, \varepsilon)$ is simply a positive basis for $I\!\!R^n$ [2, 6]. (A positive basis is a set of generators for a cone in the case that the cone is a vector space.) Thus, if the iterate is suitably in the interior of Ω, the algorithm will look like a pattern search algorithm for unconstrained minimization [6], 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 geometry of the boundary, as depicted in Fig. 3.4.

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 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.
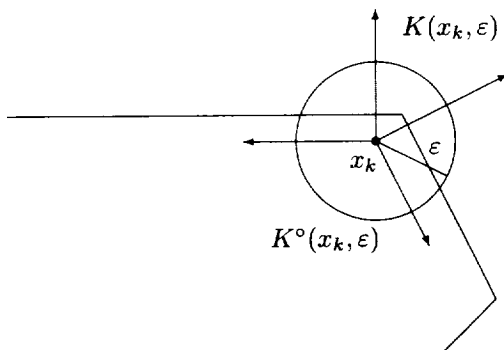
FIG. 3.4. *The situation near the boundary.*

4. Convergence analysis. The following hypotheses underlie our analysis:

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

The rationality of A is a simple way of ensuring that we can find a rational lattice that fits inside the feasible region in a suitable way.

HYPOTHESIS 2. *The set $L_\Omega(x_0) = \{\ x \in \Omega \mid f(x) \le f(x_0)\ \}$ is compact.*

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

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 will 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 §2 of [3], for instance.

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

$$\|\ q(x)\ \| \le \|\ g(x)\ \|$$

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 1–3 hold. Let $\{x_k\}$ be the sequence of iterates produced by the generalized pattern search method for linearly constrained minimization (Fig. 3.2). Then*

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

We can strengthen Theorem 4.2 in the same way that we do in the unconstrained and bound constrained cases [5, 11], by adding the following hypotheses.

HYPOTHESIS 4. *The columns of the generating matrix C_k remain bounded in norm, i.e., there exists $C > 0$ such that for all k, $C > \|c_k^i\|$, 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.*

HYPOTHESIS 6. *We have $\lim_{k \to +\infty} \Delta_k = 0$.*

6

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 $f(x_k + s_k) \leq \min \{ f(x_k + y) \mid y \in \Delta_k \Gamma_k \text{ and } (x_k + y) \in \Omega \}$.

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

Note that we do not require step lengths to be monotone non-increasing.

Then we obtain the following stronger result.

THEOREM 4.3. *Assume Hypotheses 1 6 hold. Then for the sequence of iterates $\{x_k\}$ produced by the generalized pattern search method for linearly constrained minimization (Fig. 3.2),*

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

5. Results from the standard theory.

We will need the following results from the analysis of pattern search methods in the unconstrained case. For the proofs, see [11]; 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. 3.2) can be expressed in the following form:*

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

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. 3.3),*
- r_{LB} *and r_{UB} are integers depending on N,*
- 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, $B = I$.

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) If there exists a constant $C > 0$ such that for all k, $C > \|c_k^i\|$, for all $i = 1, \cdots, p_k$, then 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 $\Delta_k \geq \psi_* \|s_k^i\|$.

6. Results concerning the geometry of polyhedra.

We need a number of results concerning the geometry of polyhedral and convex cones. We begin with a classical result on the structure of finitely generated cones.

THEOREM 6.1. *Let C be a finitely generated convex cone in $\mathbb{R}^n$. 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 [12]. □

COROLLARY 6.2. *Let C be a finitely generated convex cone in $\mathbb{R}^n$ with generators $\{v_1, \cdots, v_r\}$. Then there exists $c_{6.2} > 0$, 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 $\| \lambda \| \leq c_{6.2} \| z \|$.*

Proof. Theorem 6.1 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 $c_{6.2}$, independent of z. □

Let C be a closed convex cone in $\mathbb{R}^n$ 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 of z [9, 13] allows us to express z as

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

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

PROPOSITION 6.3. *Suppose the cone C is generated by $\{v_1, \cdots, v_r\}$. Given $\gamma > 0$, there exists $c_{6.3} > 0$, depending only on $\{v_1, \cdots, v_r\}$ and γ, such that if $\| z_C \| \geq \gamma \| z \|$, $z \neq 0$, then*

$$\max_{1 \leq i \leq r} \frac{z^T v_i}{\| z \| \| v_i \|} \geq c_{6.3}.$$

Proof. By Corollary 6.2, we have $c_{6.2} > 0$, 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 $\| \lambda \| \leq c_{6.2} \| z_C \|$ 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 $\| \lambda \| \leq c_{6.2} \| z_C \|$, we obtain

$$z^T v_i \geq \frac{1}{r} \frac{1}{c_{6.2}} \| z_C \|.$$

If we let

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

8

and apply the hypothesis $\| z_C \| \geq \gamma \| z \|$, we obtain

$$z^T v_i \geq \gamma \frac{1}{r} \frac{1}{c_{6.2}} \frac{1}{v^*} \| v_i \| \| z \|$$

and the desired result. $\square$

COROLLARY 6.4. *Given* $\gamma > 0$, *there exists* $c_{6.4} > 0$, *depending only on* A *and* γ, *for which the following hold. For any* $x \in \Omega$, $\varepsilon \geq 0$,

    1. If $\| z_{K(x,\varepsilon)} \| \geq \gamma \| z \|$, $z \neq 0$, *then*

$$\max_{1 \leq i \leq r} \frac{z^T v_i}{\| z \| \| v_i \|} \geq c_{6.4}.$$

    where $\{v_1, \cdots, v_r\}$ *are the generators of* $K(x,\varepsilon)$ *defined in (3.4) (3.5).*

    2. If $\| z_{K^\circ(x,\varepsilon)} \| \geq \gamma \| z \|$, $z \neq 0$, *then*

$$\max_{1 \leq i \leq r} \frac{z^T v_i}{\| z \| \| v_i \|} \geq c_{6.4}.$$

    where $\{v_1, \cdots, v_r\}$ *are the generators of* $K^\circ(x,\varepsilon)$ *required in §3.5.2 to be in the set* Γ.

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 6.3 to each of these cones in turn (with the generators (3.4) (3.5) for $K(x,\varepsilon)$ and the generators in Γ for $K^\circ(x,\varepsilon)$) and taking the minimum yields the corollary. $\square$

Let

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

We may assume, without loss of generality, that $a_* > 0$. The next proposition says that if $x \in \Omega$ is close to the boundary of Ω and a sufficiently long step in the direction w remains feasible, then w cannot be "too normal" to $\partial\Omega$ near x.

PROPOSITION 6.5. *Given* $\eta > 0$, *there exist* $R > 0$ *and* $c_{6.5} > 0$, *depending only on* A, *such that if* $0 \leq \varepsilon \leq R$, $x \in \partial\Omega(\varepsilon)$, $\| w \| \geq \eta$, *and* $(x + w) \in \Omega$, *then* $\| P_{K^\circ(x,\varepsilon)} w \| \geq c_{6.5} \| w \|$.

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

(6.1)
$$b - \varepsilon \| a_i \| \leq a_i^T x \leq b + \varepsilon \| a_i \|.$$

Also note that the distance from the affine subspace $a_i^T x = \ell_i$ to $a_i^T x = u_i$ is

$$\frac{u_i - \ell_i}{\| a_i \|}.$$

Now let $c_{6.4} > 0$ be the constant from Corollary 6.4. Set

$$R = \frac{c_{6.4} a_*}{a^*} \frac{\eta}{2}$$

and consider any ε such that $0 \leq \varepsilon \leq R$.

By our convention, if $K(x, \varepsilon) = \emptyset$, then $K^\circ(x, \varepsilon) = I\!\!R^n$ and the proposition holds with $c_{6.5} = 1$ since the projection onto $K^\circ(x, \varepsilon)$ is the identity. Thus we need only consider the case where $K(x, \varepsilon) \neq \emptyset$. From (6.1), if $x \in \partial\Omega(\varepsilon)$ then for $i \in I_\ell(x, \varepsilon) \cup I_u(x, \varepsilon)$ we have either

$$\ell_i \leq a_i^T x \leq \ell_i + \varepsilon \parallel a_i \parallel$$

or

$$u_i - \varepsilon \parallel a_i \parallel \leq a_i^T x \leq u_i.$$

This gives us two cases to consider:

1. Suppose $\ell_i \leq a_i^T x \leq \ell_i + \varepsilon \parallel a_i \parallel$. Since $(x + w) \in \Omega$, we have

$$\ell_i \leq a_i^T x + a_i^T w_K + a_i^T w_{K^\circ},$$

where $w = w_K + w_{K^\circ}$ is the polar decomposition of w, whence

$$0 \leq a_i^T x - \ell_i + a_i^T w_K + a_i^T w_{K^\circ} \leq \varepsilon \parallel a_i \parallel + a_i^T w_K + a^* \parallel w_{K^\circ} \parallel.$$

Since in this case $\nu_{\ell_i}(x, \varepsilon) = -a_i$ (see (3.4)), we have

(6.2) $$\nu_{\ell_i}^T w_K \leq \varepsilon \parallel a_i \parallel + a^* \parallel w_{K^\circ} \parallel \leq a^*(\varepsilon + \parallel w_{K^\circ} \parallel).$$

2. If, on the other hand, $u_i - \varepsilon \parallel a_i \parallel \leq a_i^T x \leq u_i$, then, since $(x + w) \in \Omega$,

$$a_i^T x + a_i^T w_K + a_i^T w_{K^\circ} \leq u_i$$

yields

$$0 \leq u_i - a_i^T x - a_i^T w_K - a_i^T w_{K^\circ} \leq \varepsilon \parallel a_i \parallel - a_i^T w_K + a^* \parallel w_{K^\circ} \parallel.$$

In this case $\nu_{u_i}(x, \varepsilon) = a_i$ (see (3.5)), so

(6.3) $$\nu_{u_i}^T w_K \leq \varepsilon \parallel a_i \parallel + a^* \parallel w_{K^\circ} \parallel \leq a^*(\varepsilon + \parallel w_{K^\circ} \parallel).$$

Now consider the generators ν_{ℓ_i}, ν_{u_i} for $K(x, \varepsilon)$. If we apply Corollary 6.4 to $K(x, \varepsilon)$ and $z = w_K$ (with $\gamma = 1$), then for one of the generators, which we will simply call ν, we have $\nu^T w_K \geq c_{6.4} \parallel w_K \parallel \parallel \nu \parallel$. From (6.2) and (6.3) we obtain

$$c_{6.4} \parallel \nu \parallel \parallel w_K \parallel \leq \nu^T w_K \leq a^*(\varepsilon + \parallel w_{K^\circ} \parallel)$$

or

$$\parallel w_K \parallel \leq \frac{a^*}{c_{6.4} a_*} (\varepsilon + \parallel w_{K^\circ} \parallel).$$

Since $\parallel w \parallel \leq \parallel w_K \parallel + \parallel w_{K^\circ} \parallel$, we have

$$\parallel w \parallel \leq \frac{a^*}{c_{6.4} a_*} (\varepsilon + \parallel w_{K^\circ} \parallel) + \parallel w_{K^\circ} \parallel$$

By our choice of ε and the hypothesis $\parallel w \parallel \geq \eta$, we obtain

$$\parallel w \parallel \leq \frac{\eta}{2} + \left(1 + \frac{a^*}{c_{6.4} a_*}\right) \parallel w_{K^\circ} \parallel$$

$$\leq \frac{\parallel w \parallel}{2} + \left(1 + \frac{a^*}{c_{6.4} a_*}\right) \parallel w_{K^\circ} \parallel,$$

and finally

$$\| w \| \le 2 \left(1 + \frac{a^*}{c_{6.4} a_*} \right) \| w_{K^\circ} \|,$$

giving us the desired bound. □

The next proposition relates the global geometry of Ω and the local geometry of Ω near a feasible point x. It is an elaboration of the observation that a convex set lies on one side of a hyperplane tangent to its boundary. The result is also true without the restriction of x to a compact subset of Ω, but this assumption shortens the proof.

PROPOSITION 6.6. *Given $\gamma > 0$, $\eta > 0$, and a compact set $S \subseteq \Omega$, there exist $R > 0$ and $c_{6.6} > 0$ such that for all $x \in S$ and all ε, $0 \le \varepsilon \le R$, if $\| w \| \le \gamma$ and $\| P_\Omega(x + w) - x \| \ge \eta$, then*

$$\| P_{K^\circ(x,\varepsilon)} w \| \ge c_{6.6} \| P_{K^\circ(x,\varepsilon)} \left(P_\Omega(x + w) - x \right) \|.$$

Proof. Suppose the proposition is not true. Then, for all j, trying $c_{6.6} = R = 1/j$, we can find $x_j \in S$, ε_j for which $0 \le \varepsilon_j \le 1/j$, and w_j for which $\| w_j \| \le \gamma$ and $\| P_\Omega(x_j + w_j) - x_j \| \ge \eta$, such that

$$\| P_{K^\circ(x_j, \varepsilon_j)} w_j \| < \frac{1}{j} \| P_{K^\circ(x_j, \varepsilon_j)} \left(P_\Omega(x_j + w_j) - x_j \right) \|.$$

For convenience, let $z_j = P_\Omega(x_j + w_j) - x_j$; then $\| z_j \| \le \| w_j \|$ (see Proposition 4.1).

Applying the compactness of S and the boundedness of the sequences $\{w_j\}$, $\{z_j\}$, we may find a subsequence for which $x_{j_k} \to x_* \in S$, $w_{j_k} \to w_*$, and $z_{j_k} \to z_*$, where $\| w_* \| \le \gamma$, $\| z_* \| \ge \eta$, and $\| z_* \| \le \gamma$. Note $z_* = P_\Omega(x_* + w_*) - x_*$ by the continuity of the projection P_Ω.

Furthermore, since there are only a finite number of possible cones $K(x_{j_k}, \varepsilon_{j_k})$, there exists a cone K_* such that $K_* = K(x_{j_k}, \varepsilon_{j_k})$ infinitely often. By selecting this further subsequence, we may assume that $K_* = K(x_{j_k}, \varepsilon_{j_k})$ for all k.

Next we will show that $K_* \subseteq K(x_*, 0)$. By construction, K_* is generated by a subset of $\{\pm a_i\}$, $i \in \{1, \cdots, m\}$. If $-a_i \in K_* = K(x_{j_k}, \varepsilon_{j_k})$, then

$$\mathrm{dist}(x_{j_k}, A_{\ell_i}) \le \varepsilon_{j_k} \le 1/j_k$$

for all k, so

$$\begin{aligned}
\mathrm{dist}(x_*, A_{\ell_i}) &= \frac{| \ell_i - a_i^T x_* |}{\| a_i \|} \\
&\le \frac{| \ell_i - a_i^T x_{j_k} |}{\| a_i \|} + \frac{| a_i^T x_{j_k} - a_i^T x_* |}{\| a_i \|} \\
&= \mathrm{dist}(x_{j_k}, A_{\ell_i}) + \frac{| a_i^T x_{j_k} - a_i^T x_* |}{\| a_i \|}.
\end{aligned}$$

Taking the limit as $k \to \infty$ we see that $\mathrm{dist}(x_*, A_{\ell_i}) = 0$, so $-a_i \in K(x_*, 0)$. A similar argument, substituting u_i for ℓ_i, shows that if $a_i \in K_*$, then $a_i \in K(x_*, 0)$. Since $K(x_*, 0)$ contains the generators of K_*, it follows that $K_* \subseteq K(x_*, 0)$, as desired.

Consequently, $K_*^\circ \supseteq K^\circ(x_*, 0)$ and, since $K_*^\circ = K^\circ(x_{j_k}, \varepsilon_{j_k})$, we have

$$\| P_{K^\circ(x_*, 0)} w_* \| \le \| P_{K_*^\circ} w_* \| = \| P_{K^\circ(x_{j_k}, \varepsilon_{j_k})} w_* \|$$

11

$$\leq \left\| P_{K^{\circ}(x_{j_k}, \varepsilon_{j_k})}(w_* - w_{j_k}) \right\| + \left\| P_{K^{\circ}(x_{j_k}, \varepsilon_{j_k})} w_{j_k} \right\|$$

$$\leq \| w_* - w_{j_k} \| + \left\| P_{K^{\circ}(x_{j_k}, \varepsilon_{j_k})} z_{j_k} \right\| / j_k$$

$$\leq \| w_* - w_{j_k} \| + \| z_{j_k} \| / j_k$$

$$\leq \| w_* - w_{j_k} \| + \gamma / j_k,$$

from which we conclude that

$$\| P_{K^{\circ}(x_*, 0)} w_* \| = 0.$$

This means that w_* is normal to Ω at x_*. However, it would follow that $z_* = 0$, a contradiction. $\square$

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. Define

$$(6.4) \qquad\qquad h = \min_{\substack{1 \leq i \leq m \\ \ell_i \neq u_i}} \frac{u_i - \ell_i}{\| a_i \|}.$$

This is the minimum distance between the faces of Ω associated with the constraints that are not equality constraints.

PROPOSITION 6.7. *Given $\varepsilon > 0$, $\varepsilon < h/2$, there exists $\delta_{6.7} > 0$ such that for any $x \in \Omega$, if $w \in K^{\circ}(x, \varepsilon)$ and $\| w \| \leq \delta_{6.7}$, then $(x + w) \in \Omega$.*

Proof. Let

$$\delta = \delta_{6.7} = \frac{\varepsilon}{2}$$

and 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

$$a_i^T x + a_i^T w \geq \ell_i + \varepsilon \| a_i \| - \| a_i \| \| w \| \geq \ell_i + (\varepsilon - \delta) \| a_i \| \geq \ell_i$$

and

$$a_i^T x + a_i^T w \leq u_i - \varepsilon \| a_i \| + \| a_i \| \| w \| \leq u_i - (\varepsilon - \delta) \| a_i \| \leq u_i.$$

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 $(w, -a_i) \leq 0$ and $(w, a_i) \leq 0$, 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 $(-a_i, w) \leq 0$. Applying (6.1) we obtain

$$\ell_i \leq a_i^T x + a_i^T w \leq \ell_i + \varepsilon \| a_i \| + \| a_i \| \| w \| \leq \ell_i + \| a_i \| (\varepsilon + \delta) \leq u_i.$$

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

$$u_i \geq a_i^T x + a_i^T w \geq u_i - \varepsilon \| a_i \| - \| a_i \| \| w \| \geq u_i - \| a_i \| (\varepsilon + \delta) \geq \ell_i.$$

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

7. Proof of Theorems 4.2 and 4.3. 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 [5]).

PROPOSITION 7.1. *Let $s \in \mathbb{R}^n$ and $x \in L_\Omega(x_0)$. Assume that $g(x) \neq 0$ and $g(x)^T s \leq -\varepsilon \| s \|$. Then, if $\| s \| < \omega(x, \frac{\varepsilon}{2})$,*

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

The next result is the crux of the convergence analysis. All the results of §6 are brought to bear to 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 Γ, where Γ is as in §3.1 and §3.5. If $\Delta < \delta / \Gamma^*$, then $\| s_k^i \| < \delta$ for all $s_k^i \in \Gamma_k$.

PROPOSITION 7.2. *Let $\eta > 0$. Then there exists $c_{7.2} > 0$ and $\delta_{7.2} > 0$ such that if $\| q_k \| \geq \eta$ and $\Delta_k < \delta_{7.2}$ then there is a trial step s_k^i defined by a column of $\Delta_k \Gamma_k$ for which $(x_k + s_k^i) \in \Omega$ and*

$$-g_k^T s_k^i \geq c_{7.2} \| q_k \| \| s_k^i \|.$$

Proof. By hypothesis, $\| g_k \|$ is bounded on $L_\Omega(x_0)$; let g^* be an upper bound for $\| g_k \|$. Consider $\varepsilon > 0$ sufficiently small that

1. $\varepsilon < \min(\varepsilon^*, h/2)$, where ε^* was introduced in §3.5.2 and h is given by (6.4), and
2. both Proposition 6.5 and Proposition 6.6 hold,

where we invoke Proposition 6.6 with $\gamma = g^*$ since we intend to apply it to $w = -g_k$.

If $x_k \in \partial\Omega(\varepsilon)$, then, since $(x_k + q_k) \in \Omega$ and $\| q_k \| \geq \eta$, by Proposition 6.5 we have

$$(7.1) \qquad \| P_{K^\circ(x_k, \varepsilon)} q_k \| \geq c_{6.5} \| q_k \|.$$

This bound also holds if $x \notin \partial\Omega(\varepsilon)$, since then $K(x, \varepsilon) = \emptyset$ and $K^\circ(x, \varepsilon) = \mathbb{R}^n$, and necessarily $c_{6.5} \leq 1$.

Meanwhile, Proposition 6.6 says that

$$\| P_{K^\circ(x_k, \varepsilon)}(-g_k) \| \geq c_{6.6} \| P_{K^\circ(x_k, \varepsilon)} q_k \|,$$

so from (7.1) we obtain

$$(7.2) \qquad \| P_{K^\circ(x_k, \varepsilon)}(-g_k) \| \geq c_{6.5} c_{6.6} \| q_k \|$$

$$(7.3) \qquad \geq c_{6.5} c_{6.6} \eta \| g_k \| / g^*.$$

We require the core pattern Γ_k of P_k to include generators for all of the cones $K^\circ(x_k, \varepsilon)$, $\varepsilon < \varepsilon^*$. Therefore some subset of the core pattern steps s_k^i forms a set of generators for $K^\circ(x_k, \varepsilon)$. The lower bound (7.3) allows us to apply Corollary 6.4 with $z = -g_k$ and $K^\circ(x_k, \varepsilon)$. If we do so and apply (7.2), we see that for some $s_k^i \in \Delta_k \Gamma_k$,

$$(-g_k, s_k^i) \geq c_{6.4} \| -g_k \| \| s_k^i \|$$

$$\geq c_{6.4} \| P_{K^\circ(x_k, \varepsilon)}(-g_k) \| \| s_k^i \|$$

$$\geq c_{6.4} c_{6.5} c_{6.6} \| q_k \| \| s_k^i \|$$

13

Thus we are assured of a descent direction inside the pattern.

Now we must show that we can take a sufficiently long step in the direction of this descent direction and remain feasible. Proposition 6.7 allows us to do this; given $\varepsilon > 0$ we can find δ such that once $\Delta < \delta/\Gamma^*$, then, since $s_k^i \in K^\circ(x_k, \varepsilon)$ and $\| s_k^i \| < \delta$, we have $(x_k + s_k^i) \in \Omega$. $\square$

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

PROPOSITION 7.3. *Given any $\eta > 0$, there exists $\delta > 0$, independent of k, such that if $\Delta_k < \delta$ and $\| q_k \| > \eta$, the pattern search method for linearly constrained minimization will find an acceptable step s_k; i.e., $f(x_k + s_k) < f(x_k)$ and $(x_k + s_k) \in \Omega$.*

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 exist $\delta_{7.3} > 0$ and $\sigma > 0$, independent of k, such that if $\Delta_k < \delta_{7.3}$ and $\| q_k \| > \eta$, then

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

Proof. 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

$$\omega \left(x_k, \frac{c_{7.2}}{2} \eta \right) \geq \omega_*$$

for all k for which $\| q_k \| > \eta$. Also define

$$\delta = \delta_{7.3} = \min \left(\delta_{7.2}, \omega_*/\Gamma^* \right).$$

Now suppose $\| q_k \| > \eta$ and $\Delta_k < \delta$. Since $\Delta_k < \delta_{7.2}$, Proposition 7.2 assures us of the existence of a step s_k^i defined by a column of $\Delta_k \Gamma_k$ such that $(x_k + s_k^i) \in \Omega$ and

$$g_k^T s_k^i \leq -c_{7.2} \| q_k \| \| s_k^i \|.$$

At the same time, we also have

$$\| s_k^i \| \leq \Delta_k \Gamma^* \leq \omega_* \leq \omega \left(x_k, \frac{c_{7.2}}{2} \| q_k \| \right).$$

Hence, by Proposition 7.1,

$$f(x_k + s_k^i) - f(x_k) \leq -\frac{c_{7.2}}{2} \| q_k \| \| s_k^i \|.$$

Thus, when $\Delta_k < \delta$, $f(x_k + s_k^i) < f(x_k)$ 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

$$f(x_{k+1}) - f(x_k) \leq -\frac{c_{7.2}}{2} \| q_k \| \| s_k^i \|.$$

14

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

$$f(x_{k+1}) \leq f(x_k) - \frac{c_{7.2}}{2} \zeta_* \Delta_k \parallel q_k \parallel.$$

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

$$f(x_{k+1}) \leq f(x_k) - \frac{c_{7.2}}{2} \zeta_* \psi_* \parallel q_k \parallel \parallel s_k \parallel.$$

This yields the second part of the proposition with $\sigma = (c_{7.2}/2)\zeta_*\psi_*$. $\square$

COROLLARY 7.4. *If* $\liminf_{k \to +\infty} \parallel q_k \parallel \neq 0$, *then there exists a constant* $\Delta_* > 0$ *such that for all* k, $\Delta_k > \Delta_*$.

Proof. By hypothesis, there exists N and $\eta > 0$ such that for all $k > N$, $\parallel q_k \parallel > \eta$. By Proposition 7.3, we can find δ 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. 3.3, we are assured that for all $k > N$, $\Delta_k > \theta\delta$. We may then take $\Delta_* = \min\{\Delta_0, \cdots, \Delta_N, \theta\delta\}$. $\square$

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 7.5. *Under Hypotheses 1-3,* $\liminf_{k \to +\infty} \Delta_k = 0$.

Proof. The proof is like that of Theorem 3.3 in [11]. Suppose $0 < \Delta_{LB} \leq \Delta_k$ for all k. Using the rules for updating Δ_k, found in Fig. 3.3, 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)* Lemma 5.2 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 [11]. In the context of Theorem 5.1,

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

If, in the matter at hand, we let

(7.4) $$r_{LB} = \min_{0 \leq k < +\infty}\{r_k\} \qquad r_{UB} = \max_{0 \leq k < +\infty}\{r_k\},$$

then (5.1) holds for the bounds given in (7.4), 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. 3.2, which guarantees that an iterate 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 $k \geq N$, $x_k = x_*$, which implies $f(x_k) = f(x_k + s_k)$.

We now appeal to the algorithm for updating Δ_k (part (a) in Fig. 3.3) to see that $\Delta_k \to 0$, thus leading to a contradiction. $\square$

7.1. The Proof of Theorem 4.2. The proof is like that of Theorem 3.5 in [11]. Suppose that $\liminf_{k \to +\infty} \| q_k \| \neq 0$. Then Corollary 7.4 tells us that there exists $\Delta_* > 0$ such that for all k, $\Delta_k > \Delta_*$. But this contradicts Theorem 7.5.

7.2. The Proof of Theorem 4.3. The proof, also by contradiction, follows that of Theorem 3.7 in [11]. Suppose $\limsup_{k \to +\infty} \| q_k \| \neq 0$. Let $\varepsilon > 0$ be such that there exists a subsequence $\| q(x_{m_i}) \| \geq \varepsilon$. Since

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

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

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

Since $\Delta_k \to 0$, we can appeal to Proposition 7.3 to obtain for $m_i \leq k < l_i$, i sufficiently large,

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

where $\sigma > 0$. Summation then yields

$$f(x_{m_i}) - f(x_{l_i}) \; \geq \; \sum_{k=m_i}^{l_i} \sigma \eta \| s_k \| \; \geq \; c' \| x_{m_i} - x_{l_i} \|.$$

Since f is bounded below, $f(x_{m_i}) - f(x_{l_i}) \to 0$ 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,

$$(7.5) \qquad \| q(x_{m_i}) \| \; \leq \; \| q(x_{m_i}) - q(x_{l_i}) \| + \| q(x_{l_i}) \| \; \leq \; 2\eta.$$

Since (7.5) must hold for any η, $0 < \eta < \varepsilon$, we have a contradiction (e.g., try $\eta = \frac{\varepsilon}{3}$).

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 §3.1 and the geometric properties required in §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 Γ as required in §3.1. The construction is an elaboration of the proof that polyhedral cones are finitely generated (see [12], 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:

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

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

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

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 $\left\{ z \mid V^T Rz \leq 0 \right\}$. 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. $\square$

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 non-degeneracy — 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 [7] 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. 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 $-V(V^TV)^{-1}$ 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 $(V^Tw, \lambda) \leq 0$ for all $\lambda \geq 0$, $\lambda_1, \cdots, \lambda_r = 0$. If we once again define $\xi = V^Tw$, 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 $V(V^TV)^{-1}$ and their negatives, and the last $p - r$ columns of $-V(V^TV)^{-1}$. $\square$

Proposition 8.2 leads to the following construction of patterns for linearly constrained minimization. Under the assumption of non-degeneracy, we know there exists ε^* such that if $0 \leq \varepsilon \leq \varepsilon^*$, 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 §3.5. The following algorithm implicitly estimates ε^*: it conducts what amounts to a safe-guarded 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 $\varepsilon_k \geq \varepsilon_*$. Then

1. Define the cone $K(x_k, \varepsilon_k)$ as in §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 (3.4) (3.5)). 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 $\pm (I - V(V^TV)^{-1}V^T)$.
4. Form the matrix $\Gamma_k = [\, N \quad V(V^TV)^{-1} \quad -V(V^TV)^{-1}\,]$.

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

8.3. The case of bound constraints. Matters simplify enormously in the case of bound constraints, previously considered in [5]. We will briefly discuss the specialization to bound constrained minimization and in the process sharpen the results in [5].

In the case of bound constraints we have

$$
\begin{aligned}
&\text{minimize} && f(x) \\
&\text{subject to} && l \leq x \leq u.
\end{aligned}
$$
(8.1)

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 [5] 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 [5] 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.

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] C. DAVIS, *Theory of positive linear dependence*, American Journal of Mathematics, (1954), pp. 733 746.

[3] 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.

[4] 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. 212 229.

[5] R. M. LEWIS AND V. J. TORCZON, *Pattern search algorithms for bound constrained minimization*, Tech. Rep. 96-20, ICASE, Mail Stop 403, NASA Langley Research Center, Hampton, Virginia 23681-0001, 1996. To appear in SIAM J. on Optimization.

[6] ———, *Rank ordering and positive bases in pattern search algorithms*, Tech. Rep. 96-71, ICASE, Mail Stop 403, NASA Langley Research Center, Hampton, Virginia 23681-0001, 1996. Submitted to Mathematical Programming.

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

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

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

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

19

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

[12] J. VAN TIEL, *Convex Analysis: An Introductory Text*, John Wiley & Sons, 1984.

[13] 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.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY(Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1998	Contractor Report

4. TITLE AND SUBTITLE

Pattern Search Methods for Linearly Constrained Minimization

5. FUNDING NUMBERS

C NAS1-19480
WU 505-90-52-01

6. AUTHOR(S)

Robert Michael Lewis
Virginia Torczon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Computer Applications in Science and Engineering
Mail Stop 403, NASA Langley Research Center
Hampton, VA 23681-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

ICASE Report No. 98-3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23681-2199

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA/CR-1998-206904
ICASE Report No. 98-3

11. SUPPLEMENTARY NOTES

Langley Technical Monitor: Dennis M. Bushnell
Final Report
Submitted to SIAM Journal on Optimization

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 64
Distribution: Nonstandard
Availability: NASA-CASI (301)621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

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. 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.

14. SUBJECT TERMS

pattern search, linearly constrained minimization

15. NUMBER OF PAGES

25

16. PRICE CODE

A03

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified		