FRAMES AND GRIDS IN UNCONSTRAINED AND LINEARLY
CONSTRAINED OPTIMIZATION: A NONSMOOTH APPROACH∗

C. J. PRICE† AND I. D. COOPE†

Abstract. This paper describes a class of frame-based direct search methods for unconstrained and linearly constrained optimization. A template is described and analyzed using Clarke’s nonsmooth calculus. This provides a unified and simple approach to earlier results for grid- and frame-based methods, and also provides partial convergence results when the objective function is not smooth, undefined in some places, or both. The template also covers many new methods which combine elements of previous ideas using frames and grids. These new methods include grid-based simple descent algorithms which allow moving to points off the grid at every iteration and can automatically control the grid size, provided function values are available. The concept of a grid is also generalized to that of an admissible set, which allows sets, for example, with circular symmetries. The method is applied to linearly constrained problems using a simple barrier approach.

Key words. derivative-free optimization, positive basis methods, nonsmooth convergence analysis, frame-based methods

AMS subject classifications. 49M30, 65K05

DOI. 10.1137/S1052623402407084

1. Introduction. This paper discusses the use of frames and grids in derivative-free optimization. The unconstrained optimization problem is examined first and analyzed using Clarke’s nonsmooth calculus [3]. This is extended to linearly constrained problems by aligning the frames and grids with appropriate subsets of the linear constraints. Herein a grid is the set of points in \( R^n \) which contains a given origin point and all points which differ from this origin by an integer combination of members of a basis for \( R^n \).

In 1997, Torczon [17] showed that many existing direct search methods conform to a common structure called generalized pattern search (GPS), which restricts attention to a sequence of interrelated meshes. A mesh is defined in the same way as a grid, except that only nonnegative integer combinations are used, and the basis is replaced by a set of vectors whose nonnegative combinations (with real coefficients) contain \( R^n \). In [17] GPS was shown to converge under mild conditions, including continuous differentiability of the objective function. Since then a number of generalizations and modifications of GPS have been proposed. Amongst them is the work of Lewis and Torczon [12, 13] in extending GPS to bound and linearly constrained problems. A simple barrier approach is used where the objective function is declared to be infinite at any point which violates one or more constraints. The barrier approach aligns the set of interrelated meshes with the constraints. This allows the set of search steps to adequately reflect each possible cone of feasible directions. More recently, Audet and Dennis [1] have simplified the analysis in [13, 17] and extended it to nonsmooth functions by using Clarke’s generalized derivatives [3].

In GPS the meshes are related because each mesh is a subset of some member of a sequence of nested grids. Coope and Price [6] have shown that for unconstrained
optimization problems, the grids do not have to be related to one another. In [6] a step to any lower point (not necessarily a grid point) is permitted each time a new grid is selected. The orientation and shape of the new grid can be chosen independently from those of previous grids. This permits grids to be chosen to reflect information gathered during previous iterations. In contrast to GPS, the algorithm in [6] must force the grids to become arbitrarily fine. As shall be shown later there are convenient methods for doing this.

A disadvantage of the grid-based template in [6] is that steps to arbitrary lower points can occur only when there is a change of grid, and such changes can be infrequent. In [4] a frame-based method which allows steps to arbitrary lower points is described, and convergence is shown under mild conditions. This method uses a sufficient descent condition to enforce convergence. A similar set of methods is described by García-Palomares and Rodríguez [10]. These methods [10] are not explicitly formulated in terms of frames and restrict themselves to a fixed set of search directions for all iterations. Indeed, it can be shown that the implemented algorithms SDSA and NSDSA in [10] are special cases of the framework presented in [4] and also of the template presented herein. It is shown in section 4 that the prototype sequential algorithms presented in [10] also conform to a simple extension of the template presented herein. Without the extension, our work uses a single sufficient descent condition for all directions, whereas [10] uses a different condition for each direction. An explicit sufficient descent condition is used in [4], which is the only reason why the prototype sequential algorithms in [10] do not conform to the framework in [4]. The convergence results in [10] are similar to the ones in this paper and exceed those in [4] as the latter restricts attention to $C^1$ functions. Unlike [10], we do not consider the case when $f$ is locally convex.

This paper looks at the use of grids and frames in unconstrained and linearly constrained derivative-free optimization. The optimization problem may be concisely expressed as

$$\min_{x \in \Omega} f(x), \quad \text{where} \quad \Omega \subseteq \mathbb{R}^n \quad (1.1)$$

and where a local minimum is sought. The objective function $f$ maps $\mathbb{R}^n$ into $\mathbb{R} \cup \{+\infty\}$, with the convention that $f$ is assigned the value $+\infty$ in regions where it is undefined. We also focus attention on the cases where $f$ is locally Lipschitz, strictly differentiable [3], or $C^1$. The lack of second derivatives means that stationary points will be accepted as solutions in practice. The case when linear constraints are present is also examined. These constraints are used to define the feasible region $\Omega$. We look at how grids and frames can be chosen to take into account the geometry of $\Omega$.

This paper shows that grid-based methods can be expressed and analyzed in terms of frames, thereby unifying the treatment of grid- and frame-based methods. This unification allows many hybrid methods to be formed, including those which permit arbitrary simple descent steps at every iteration. This is achieved by formulating a frame-based template which can temporarily restrict attention to a subset of $\mathbb{R}^n$ called an admissible set. The concept of an admissible set is a generalization of the idea of a mesh in GPS or of a grid in [6]. A sequence of admissible sets may be used, where these sets eventually become progressively finer. For grid-based methods the grids are the admissible sets. For frame-based methods the admissible sets are equal to $\mathbb{R}^n$. The introduction of admissible sets allows methods which are analogous to grid-based methods but do not use rectangular grids. The template is described in terms of sufficient descent. For appropriate choices of admissible set the phrase
sufficient descent means simple descent; for other choices of admissible set sufficient
descent is stricter than simple descent. The template is strongly connected with GPS
methods when simple descent is always used and only points in the admissible sets
are considered, where the admissible sets are nested grids (or subsets thereof). This
is discussed in detail in [6].

The basic strategy is to generate a sequence of iterates in \( \Omega \) whose cluster points
are solutions of (1.1) under appropriate conditions. The function values at these
iterates form a decreasing sequence. For convenience, one point is said to be lower
(or better) than another if it has a lower function value. At each iteration a search
is conducted for a point which is sufficiently lower than the current iterate. When
the search is unsuccessful, the current iterate is called quasi-minimal. The search
for a sufficiently lower point is required to satisfy a number of conditions. Included
are conditions which ensure it is a finite process and conditions which ensure the
search is not declared unsuccessful until it has adequately explored the region around
the current iterate. This exploration takes into account the local geometry of \( \Omega \) and
evaluates \( f \) at a set of points called a frame.

Frames are defined precisely in section 2, but, loosely speaking, a frame is a group
of points which surround a central point called the frame center. If none of these
surrounding points is significantly lower than the frame center, then the frame and
the frame center are called quasi-minimal. A quasi-minimal frame center (or quasi-
minimal iterate) is, in some sense, a discrete approximation to a local minimum. The
frame center itself is not part of the frame.

The basic approach of a frame-based algorithm is to generate an infinite sequence
of quasi-minimal frames such that the distances between points in these frames shrink
to zero in the limit. The nature of the cluster points of the sequence of quasi-minimal
iterates is examined using Clarke’s nonsmooth analysis. In this part our approach is
similar to Audet and Dennis’s analysis of GPS [1].

We first examine an arbitrary unspecified algorithm that generates a sequence of
iterates \( \{x^{(k)}\} \in \Omega \), where this sequence of iterates contains an infinite subsequence
\( \{z^{(m)}\} \) of quasi-minimal frame centers. The two indices \( k \) and \( m \) count the number
of iterations and quasi-minimal frames, respectively. The function \( k = k(m) \) gives
the number of the iteration in which the \( m \)th quasi-minimal frame occurs. At each
iteration a new iterate \( x^{(k+1)} \) is chosen which satisfies one of two conditions: either
\( x^{(k+1)} \) is an admissible point which is sufficiently lower than \( x^{(k)} \) or the algorithm
finds a quasi-minimal frame centered on \( x^{(k+1)} \), and \( x^{(k+1)} \) is not higher than \( x^{(k)} \).
In the latter case, this new frame center \( x^{(k+1)} \) may be anywhere in \( \Omega \), but the quasi-
minimal frame may be required to consist of points which lie in the current admissible
set. Various strategies are used to ensure a quasi-minimal frame is located in a finite
time. It is shown that this guarantees the sequence \( \{z^{(m)}\} \) is infinite. It is then
shown that the Clarke generalized derivative at each cluster point of \( \{z^{(m)}\} \) in each
limiting direction is nonnegative. In the case when the objective function is \( C^1 \) and
\( \Omega = \mathbb{R}^n \), it is shown that all such cluster points are stationary points of \( f \). These
results are extended to linearly constrained optimization problems by using a barrier
approach [1, 13] and choosing each frame to span the relevant tangent cone.

In section 3 the algorithm template is described, and its behavior is analyzed in
sections 4 and 5. Section 4 develops the main convergence results and applies them to
the unconstrained optimization problem. Section 5 addresses the linearly constrained
optimization problem. It describes how frames can be constructed which take into
account the linear constraints and presents the convergence results for methods using
such frames. Section 6 looks at how the frames’ sizes may be chosen, and concluding remarks are made in section 7.

The template (Template D) described herein is opportunistic, as are framework A in [6] and the framework presented in [4]. This means it can abandon a partially completed frame immediately after discovering a point of sufficient descent. The price paid for this opportunism is that the convergence theory applies only to the subsequence of quasi-minimal iterates \( \{ z^{(m)} \} \). A nonopportunistic approach is presented in framework B of [6] and template C of [14]. These templates require each frame to be completed and to search along the ray from the frame’s center through a point not higher than the lowest frame point. The advantage of this is that the convergence theory applies to the whole sequence of iterates \( \{ x^{(k)} \} \), not merely \( \{ z^{(m)} \} \). The restriction that each frame must be completed is not serious for certain types of algorithms. For instance, methods using finite differences [7] or polytopes [15] must come within one point of completing a frame in order to construct the gradient estimate or polytope.

2. Positive bases and frames. A frame is a finite set of points which strictly contains another point (the frame’s center) in its convex hull. The directions from the frame’s center to each point in the frame form a positive basis [9], which is a set of vectors \( \mathcal{V}_+ = \{ v_i \} \) such that

B1: every vector in \( \mathbb{R}^n \) can be written as a nonnegative combination of the vectors in \( \mathcal{V}_+ \) and

B2: no proper subset of \( \mathcal{V}_+ \) satisfies B1.

The term “nonnegative combination” means a (finite) linear combination without negative coefficients. Sets of vectors which satisfy property B1 only are called positive spanning sets. Any positive spanning set not satisfying B2 must contain a positive basis as a proper subset. It is also shown in [9] that any positive basis for \( \mathbb{R}^n \) must satisfy \( n + 1 \leq |\mathcal{V}_+| \leq 2n \). The members of each positive basis \( \mathcal{V}_+ \) which is constructed are assigned a specific order, and from now on each positive basis is assumed to be ordered unless stated otherwise.

A frame \( \Phi \) is the set of points

\[
\Phi = \Phi(z, h, \mathcal{V}_+) = \{ z + hv : v \in \mathcal{V}_+ \},
\]

where \( z \) is the frame center and the positive scalar \( h \) is the frame size.

A frame \( \Phi \) is called minimal if and only if

\[
f(z) \leq f(x) \ \forall x \in \Phi(z, h, \mathcal{V}_+).
\]

It is useful to work with frames which are only “nearly” minimal. Such frames are called quasi-minimal and are easier to generate than minimal frames. The generation of quasi-minimal (or minimal) frames is important for two reasons: the convergence theory applies to the sequence of centers of quasi-minimal frames, and some algorithm parameters can be altered only after a quasi-minimal frame has been found. A frame \( \Phi \) is called \( \epsilon_z \)-quasi-minimal if and only if

\[
f(z) \leq f(x) + \epsilon_z \ \forall x \in \Phi(z, h, \mathcal{V}_+),
\]

for a preselected nonnegative \( \epsilon_z \). The notation \( \Phi^{(m)} = \Phi(z^{(m)}, h^{(m)}_z, \mathcal{V}^{(m)}_+) \) is used to denote the \( m \)th quasi-minimal frame.

Each quasi-minimal frame may have a different value \( \epsilon^{(m)}_z \) for \( \epsilon_z \). Hence, when a frame \( \Phi^{(m)} \) is called “quasi-minimal” this is understood to mean \( \epsilon^{(m)}_z \)-quasi-minimal.
It is necessary to have values for $\epsilon$ and $h$ at every iteration, and so new sequences $\{\epsilon(k)\}$ and $\{h(k)\}$ are introduced. The sequences $\{h(k)\}$ and $\{h(m)\}$ are linked by the relation $h_z(m) = h(k(m))$. In other words, $h_z(m)$ is the value $h(k)$ takes in the iteration $k(m)$ in which the $m$th frame is located. The quantities $\epsilon_z(m)$ and $\epsilon(k)$ are similarly related.

The sequence of $\epsilon$ values is required to satisfy the following condition:

$$\lim_{k \to \infty} \frac{\epsilon(k)}{h(k)} = 0.$$  \hspace{1cm} (2.3)

A requirement of the convergence theory is that $h(k) \to 0$ as $k \to \infty$, and so one simple choice that satisfies (2.3) is $\epsilon = Nh^\nu$, with $\nu > 1$ and $N \geq 0$. In any case (2.3) requires that $\{\epsilon(k)\}$ goes to zero faster than $\{h(k)\}$.

One could easily define frames using positive spanning sets rather than positive bases. However, there are a number of advantages to the latter (see, e.g., [4, 14]). For convergence purposes a number of restrictions must be imposed on the set $V_+$ used to define a frame, and this is more easily done if $V_+$ is a positive basis rather than a positive spanning set. Second, frame-based templates permit a finite number of arbitrary points (not included in $V_+$) to be examined during each iteration. Including such points in $V_+$ subjects them to unnecessary restrictions. In practice these extra points may be used in a similar way to the members of $V_+$, but for theoretical purposes they are best kept separate.

An upper bound $K$ is imposed on the length of each member of each $V_+^{(m)}$,

$$\|v\| \leq K \quad \forall m \quad \forall v \in V_+^{(m)},$$  \hspace{1cm} (2.4)

where $K$ is independent of $m$ and $k$.

A set $V_+^{(\infty)} = \{v_1^{(\infty)}, \ldots, v_p^{(\infty)}\}$ is a limit of the sequence of ordered positive bases $\{V_+^{(m)}\}_{m=1}^{\infty}$ if and only if an infinite subsequence of $\{V_+^{(m)}\}$ exists such that each positive basis belonging to this subsequence has cardinality $p$, and

$$\lim_{m \to \infty} v_i^{(m)} = v_i^{(\infty)} \quad \forall i = 1, \ldots, p,$$  \hspace{1cm} (2.5)

where the limit is understood to be taken over this subsequence. Condition (2.4) ensures that such limits exist. The following assumption is needed.

**Assumption 2.1.** All members of the sequence $\{V_+^{(m)}\}$ satisfy (2.4), and each limit $V_+^{(\infty)}$ of the sequence $\{V_+^{(m)}\}$ is an ordered positive basis.

This assumption may be enforced in a variety of ways, some of which are discussed in [4, 14].

3. The algorithm template. The template consists of two nested loops. The outer loop (steps 2–6, indexed by $m$) generates a sequence of quasi-minimal frames with the desired properties. The purpose of the inner loop (steps 3–5, indexed by $k$) is to generate a quasi-minimal frame. Iterations of the inner loop are performed until a quasi-minimal frame is found, where quasi minimality is defined in (2.2) by $\epsilon_z(m)$. Each iteration of the inner loop which does not find a quasi-minimal frame obtains a point of sufficient descent instead. Fixing certain quantities during each iteration of the outer loop (and hence each execution of the inner loop) ensures that a quasi-minimal frame must be located in a finite number of inner loop iterations under standard assumptions. In particular, it is assumed that the sequences of function
values \( \{f^{(k)}\} \) and iterates \( \{x^{(k)}\} \) remain bounded. Here the notation \( f^{(k)} = f(x^{(k)}) \) has been used.

The purpose of the outer loop is to generate a sequence of quasi-minimal frames with the desired properties. In particular, this sequence of quasi-minimal frames must be infinite. In other words, each iteration of the outer loop must be a finite process. Termination of the \( m \)th iteration of the outer loop can be guaranteed either by choosing \( \epsilon \) to be bounded away from zero or by restricting points of sufficient descent to an admissible set \( \mathcal{G}^{(m)} \). In the former case, \( \epsilon \) is given a strictly positive lower bound \( E^{(m)} \), and \( E^{(m)} \) is kept constant between quasi-minimal frames. Sufficient descent means that \( f(x^{(k)}) \) is reduced by more than \( E^{(m)} \) at each iteration of the inner loop. The \( m \)th iteration of the outer loop can fail to terminate only if sufficient descent is always obtained. This means that \( f^{(k)} \to -\infty \) as \( k \) goes to infinity. In the latter case, \( \epsilon = 0 \) is permitted, but \( \mathcal{G}^{(m)} \) must contain only a finite number of points in any bounded subset of \( \mathbb{R}^n \), amongst other things. Hence the inner loop cannot generate a bounded infinite sequence of iterates with strictly decreasing function values. A new \( \mathcal{G}^{(m)} \) can be chosen after each quasi-minimal frame, and so \( \mathcal{G}^{(m)} \) denotes the admissible set used during the search for the \( m \)th quasi-minimal frame. When \( E^{(m)} > 0 \) we define \( \mathcal{G}^{(m)} = \mathbb{R}^n \) for completeness.

At each iteration of the inner loop \( f \) is calculated at a finite number of points. An iteration is completed when either sufficient descent is obtained or a quasi-minimal frame is located. Here sufficient descent means reducing \( f \) by more than \( \epsilon \), where \( \epsilon \) is the same constant used to define quasi minimality. At each iteration the algorithm may calculate \( f \) at a finite number of points. If neither sufficient descent nor a quasi-minimal frame has been obtained, then the algorithm begins forming a frame in \( \mathcal{G}^{(m)} \) about a frame center \( x \), where \( x \) is not higher than the previous iterate \( x^{(k-1)} \). The frame either is quasi-minimal or contains a point in \( \mathcal{G}^{(m)} \) more than \( \epsilon \) lower than \( x^{(k-1)} \). This completes an iteration of the inner loop. If sufficient descent was obtained, then the algorithm increments \( k \) and starts a new iteration of the inner loop. Otherwise, the inner loop terminates.

During each iteration of the outer loop a positive bound \( H^{(m)} \) on \( h^{(k)} \) is imposed. Theoretically this bound is superfluous, but its presence highlights the existence of a lower bound on \( h^{(k)} \) implicit in the choice of \( \mathcal{G}^{(m)} \) (if \( E^{(m)} = 0 \)) or \( E^{(m)} \) (otherwise). Further remarks on this are made later in this section and section 6, respectively.

**Algorithm Template D.**

1. Initialize: set \( k = 1 \), \( m = 1 \), and choose the initial point \( x^{(0)} \in \Omega \).
2. Choose \( H^{(m)} > 0 \), \( E^{(m)} \geq 0 \), and \( \mathcal{G}^{(m)} \).
3. Choose \( h^{(k)} \geq H^{(m)} \) and \( \epsilon^{(k)} \geq E^{(m)} \).
4. Execute any finite process which satisfies one of these conditions:
   (a) generates an iterate \( x^{(k)} \in \Omega \cap \mathcal{G}^{(m)} \) satisfying \( f(x^{(k)}) \leq f^{(k-1)} - \epsilon^{(k)} \);
   or
   (b) generates a quasi-minimal frame \( \Phi^{(m)} = \Phi(z^{(m)}, h^{(m)}_z, \nu^{(m)}_z) \), where \( x^{(k)} \in \Omega \) and \( f^{(k)} \leq f^{(k-1)} \). Here \( z^{(m)} = x^{(k)} \), \( h^{(m)}_z = h^{(k)} \), and \( \epsilon^{(m)} = \epsilon^{(k)} \); or
   (c) case (b) of this step with the added restriction \( \Phi^{(m)} \subset \mathcal{G}^{(m)} \).
5. If \( x^{(k)} \) is not quasi-minimal, increment \( k \) and go to step 3.
6. Increment \( m \) and \( k \). If stopping conditions are not satisfied, go to step 2.

Condition (c) is included in step 4 to highlight the fact that an attempt to satisfy condition (c) by forming a frame in \( \mathcal{G}^{(m)} \) guarantees the satisfaction of either (a) or (c). In contrast, attempts to satisfy either (a) or (b) may end in failure without...
satisfying any of (a)–(c). An example of how condition (c) is used is presented later in Figure 3.1.

The arbitrary process in step 4 allows $f$ to be evaluated at points anywhere in $\Omega$. These points can be used, for example, to include a quasi-Newton step, points chosen by an heuristic, or even randomly selected points. This arbitrary process also permits the lowest point from a previous quasi-minimal frame to be included in the current iteration. This is useful because a quasi-minimal frame which is not minimal contains at least one point which is lower than the frame’s center. Inclusion of such points allows movement away from a strictly concave maximum. For example, if $f = -x^2$ in $R^1$, with $x^{(0)} = 0$, $h^{(0)} = 1$, $V_+ = \{1, -1\}$, and $\epsilon = 2h^2$, then $x^{(0)}$ is a quasi-minimal frame center for all positive $h$. However, if the frame points from the first iteration are included in the arbitrary finite process of the next iteration, then on the second iteration an algorithm will step to a point $x^{(2)}$ satisfying $f^{(2)} \leq -1$ and escape the local maximum. Otherwise, an algorithm might generate an infinite sequence of quasi-minimal frames centered on the origin.

The points examined in step 4’s arbitrary process are useful in the analysis of the template, and so we define $S^{(m)}_+$ as the set containing all nonzero vectors $v$ satisfying (2.4) such that $f(z^{(m)} + h^{(m)}_+ v) \geq f(z^{(m)}) - \epsilon^{(m)}_+$ is established by the arbitrary process in step 4. That is to say,

$$S^{(m)}_+ = \left\{ v : 0 < ||v|| \leq K \text{ and } f \left( z^{(m)} + h^{(m)}_+ v \right) \geq f(z^{(m)}) - \epsilon^{(m)}_+ \text{ is shown in step 4} \right\}.$$  (3.1)

The set $G^{(m)}$ may be a grid [6], $R^n$, or otherwise. For example [6], $G^{(m)}$ may be a grid centered on $z^{(m-1)}$ and containing all points differing from $z^{(m-1)}$ by a sum of integer multiples of the vectors $v^{(m)}_1, \ldots, v^{(m)}_n$, where $v^{(m)}_1, \ldots, v^{(m)}_n$ form a basis for $R^n$. Many other possibilities also exist. The choice of $G^{(m)}$ is subject to a number of restrictions when $E^{(m)}$ is zero. When $E^{(m)} \neq 0$ we use $G^{(m)} = R^n$ without loss of generality.

**Assumption 3.1.** If $E^{(m)} = 0$, then the following two conditions hold:

G1: $G^{(m)}$ contains only a finite number of points in any bounded subset of $\Omega$; and

G2: for all $z \in \Omega$ there exists at least one frame $\Phi(z, h_+, V_+)$ in $G^{(m)}$ for which $h_+$ and $V_+$ satisfy all restrictions required by the template (including Assumption 2.1 or, in the constrained case, Assumption 5.3).

Condition G2 in Assumption 3.1 is used to ensure that an algorithm can always find a frame centered on any point it chooses. Condition G2 excludes such sets as the set of positive integers in $R^1$ because this set does not contain a frame about $x = 0$. Condition G2 conspires with inequality (2.4) to impose a lower limit on $h$. For example, if $G^{(m)}$ is the grid of all integer points in $R^n$, and $K = 5$ in (2.4), then $h$ must be at least 1/5 in order for condition G2 to be satisfied.

When $E^{(m)} = 0$ points of sufficient descent must be chosen from $G^{(m)}$, but quasi-minimal iterates are not required to belong to $G^{(m)}$. This permits an algorithm to consider points not in $G^{(m)}$ at every iteration via the following process. Let step 4 start with an iterate $x^{(k-1)}$ in $G^{(m)}$. The arbitrary finite process in this step selects a point $x \in \Omega$ which is not higher than $x^{(k-1)}$. Note that $x \in G^{(m)}$ is not required. If the algorithm subsequently locates a quasi-minimal frame around $x$, then condition (b) has been achieved, which completes step 4. Otherwise, termination of step 4 is forced by trying to achieve condition (c): that is to say, the algorithm forms a frame in $G^{(m)}$ with $x$ as the frame’s center. Either this frame is quasi-minimal (condition (c)
is satisfied) or a point in $G^{(m)}$ which is sufficiently lower than $x^{(k-1)}$ is located (which satisfies condition (a)). Either way step 4 then terminates. Condition (c) is actually superfluous; if condition (c) holds, then condition (b) is automatically satisfied.

An illustration of step 4 is given in Figure 3.1. Here $E^{(m)} = \epsilon^{(k)} = 0$, and $G^{(m)}$ is the set of all points which are intersections of either two lines or of a line and a circle. The lines form two irregularly spaced parallel sets. The circles are centered on integer points in $R^2$, and all have the same radius. Clearly $G^{(m)}$ satisfies Assumption 3.1, provided $h^{(k)}$ is not too small. Points considered by step 4 are marked with dots, and also the legend of the form $f(x_i) = F$. Here points are used in the order given by the index $i$, and $F$ is the function value of the $i$th such point. The index $i$ is not an iteration number; inside step 4 both $k$ and $m$ are fixed. The notations $x_i$ and $f_i = f(x_i)$ are restricted to this paragraph and Figure 3.1. Arrows point from each frame center to the points in the corresponding frame. Step 4 begins with the current iterate $x_1$. It calculates $f_2$. Now $x_2$ is an admissible point, so if $x_2$ were also lower than $x_1$, then step 4 would terminate under condition (a) and return $x_2$ as a point of sufficient descent. However, $x_2$ is higher than $x_1$ and is thus rejected. Step 4 then calculates $f$ at $x_3$, which is lower than $x_1$. Now $x_3$ is not admissible, so step 4 cannot return $x_3$ as a point of sufficient descent. Instead step 4 forms a frame around $x_3$, consisting of $x_1$, $x_4$, and $x_5$. If this frame were quasi-minimal (which is the same as minimal since $\epsilon^{(k)} = 0$), then step 4 would terminate under condition (b) and return $x_3$ as a quasi-minimal iterate. However, the frame is not quasi-minimal because $x_5$ is lower than $x_3$. Unfortunately, $x_5$ is not admissible, and so it cannot be returned as a point of sufficient descent. Step 4 then forms a frame around $x_5$ consisting only of admissible points: $x_6$, $x_7$, and $x_8$. This forces the termination of step 4: either at least one of $x_6$, $x_7$, and $x_8$ is lower than $x_5$ (and hence lower than $x_1$) or all three are at least as high as $x_5$. In the former case step 4 would terminate under condition (a) and return the lowest of $x_6$, $x_7$, and $x_8$. In the latter case step 4 would terminate under condition (c) and return $x_5$ as a quasi-minimal center (which is what happens).
Hence $x_5$ becomes both the new $x^{(k)}$ and the new $z^{(m)}$. The set $S^{(m)}_+$ consists of the vectors $(x_i - x_5)/h_x^{(m)}$, $i \neq 5$, which satisfy the inequalities in (3.1).

Template D uses the same set of admissible points $G^{(m)}$ for each iteration between quasi-minimal frames. At each iteration attention could be restricted to a subset of $G^{(m)}$ which satisfies conditions G1 and G2, and this would not affect the convergence results. This has not been done for two reasons. First, the current form of Template D and the resulting analysis is clearer. Second, by judicious choice of $h_x^{(m)}$ and $V^{(m)}_+$ restricting attention to a subset of $G^{(m)}$ can be achieved implicitly.

The convergence analysis examines the asymptotic properties of the sequences of iterates when the stopping conditions are never invoked. Practical considerations make stopping conditions essential, which is why they are featured in Template D. The current placement of stopping conditions ensures that the algorithm always terminates with a quasi-minimal frame. Stopping conditions could also be checked in the inner loop, for example at step 5.

4. The main convergence results. First it is shown that the subsequence of quasi-minimal frames is infinite under appropriate conditions.

**Theorem 4.1.** Assume that for each $m$ either $E^{(m)} > 0$ or $G^{(m)}$ satisfies conditions G1 and G2 in Assumption 3.1. Then at least one of the three following possibilities holds:

(i) the subsequence of quasi-minimal iterates is infinite; or
(ii) the sequence of iterates is unbounded; or
(iii) $f^{(k)} \to -\infty$ as $k \to \infty$.

**Proof.** We assume case (i) does not occur and that $J$ is the final value of $m$. In the case when $E^{(J)} = 0$ it is then shown that (ii) must occur. Similarly, when $E^{(J)}$ is strictly positive it is shown that (iii) must occur.

If $E^{(J)} = 0$, then step 4 generates a sequence of points in $G^{(J)} \cap \Omega$ with strictly decreasing function values. This sequence must contain an infinite number of distinct points in $G^{(J)} \cap \Omega$. However, $G^{(J)} \cap \Omega$ can contain only a finite number of points inside any bounded subset of $\Omega$, by condition G1. Hence the sequence of iterates must be unbounded.

Let $E^{(J)}$ be strictly positive. Once $m = J$ occurs, step 4 is executed endlessly, and it reduces the best known function value by more than $E^{(J)}$ each time it is executed. Hence $f^{(k)} \to -\infty$ in the limit $k \to \infty$, as required. \(\square\)

In addition to conditions on the sequences of ordered positive bases and admissible sets, the following assumption is needed to establish convergence.

**Assumption 4.2.** The following conditions hold:

(a) the points at which $f$ is calculated lie in a compact subset of $\mathbb{R}^n$;
(b) the sequence of function values $\{f^{(k)}\}$ is bounded below;
(c) $h^{(k)} \to 0$ as $k \to \infty$; and
(d) $\epsilon^{(k)}/h^{(k)} \to 0$ as $k \to \infty$.

The first two parts of this assumption eliminate possibilities (ii) and (iii) of Theorem 4.1, which guarantees that the sequence $\{z^{(m)}\}$ has cluster points. Parts (c) and (d) ensure that these cluster points have interesting properties. Satisfaction of these latter two parts can be ensured by an appropriate implementation of the template. Collectively parts (c) and (d) ensure $\epsilon^{(k)} \to 0$ as $k \to \infty$.

The next theorem establishes the basic convergence result using Clarke’s generalized derivative [3], which is
Provided \( f \) is locally Lipschitz at \( x \) it can be shown [3] that \( f^\circ(x;v) \) is subadditive and positively homogeneous in \( v \). Moreover, if \( M \) is a Lipschitz constant for \( f \) at \( x \), then \( |f^\circ(x;v)| \leq M|v| \).

**Theorem 4.3.** Let \( f \) be locally Lipschitz at \( z^{(\infty)} \). Let \( v \) be any vector such that there exists a sequence \( \{(z^{(m)}, v^{(m)})\} \) with \( v^{(m)} \in S_+^{(m)} \) for all \( m \) and such that \( (z^{(\infty)}, v) \) is a cluster point of this sequence, where \( S_+^{(m)} \) is defined in (3.1). Then

\[
f^\circ(z^{(\infty)};v) \geq 0.
\]

**Proof.** We restrict our attention to a subsequence of \( \{z^{(m)}\} \) for which the corresponding subsequence \( \{(z^{(m)}, v^{(m)})\} \) converges uniquely to \( (z^{(\infty)}, v) \). The definition (3.1) implies that

\[
f\left(z^{(m)} + h^z z^{(m)} v^{(m)} \right) - f\left(z^{(m)} \right) + \epsilon z^{(m)} \geq 0.
\]

Hence

\[
limit_{m \to \infty} \sup_{m \to \infty} \frac{f\left(z^{(m)} + h^z z^{(m)} \left(w^{(m)} + v \right) \right) - f\left(z^{(m)} + h^z z^{(m)} w^{(m)} \right) + f\left(z^{(m)} + h^z z^{(m)} w^{(m)} \right) - f\left(z^{(m)} \right)}{h^z z^{(m)}} \geq 0,
\]

where \( w^{(m)} = v^{(m)} - v \). Now \( w^{(m)} \to 0 \) as \( m \to \infty \), and so the first two terms provide a lower bound on \( f^\circ \), which yields

\[
f^\circ(z^{(\infty)};v) + \lim_{m \to \infty} \sup_{m \to \infty} \frac{f\left(z^{(m)} + h^z z^{(m)} w^{(m)} \right) - f(z^{(m)})}{h^z z^{(m)}} \geq 0.
\]

The last term vanishes because \( f \) is locally Lipschitz and because \( w^{(m)} \to 0 \), which yields the required result. \( \Box \)

An alternative way of looking at Theorem 4.3 is as follows.

**Corollary 4.4.** There does not exist an open halfspace on which \( f^\circ(z^{(\infty)};v) \) is negative for all \( v \) in this halfspace.

**Proof.** The template guarantees \( \mathcal{V}_+^{(m)} \subseteq S_+^{(m)} \) for all \( m \). Assumption 2.1 and Theorem 4.3 imply there exists a positive basis \( \mathcal{V}_+^{(\infty)} \) such that

\[
f^\circ(z^{(\infty)};v) \geq 0 \quad \forall v \in \mathcal{V}_+^{(\infty)}.
\]

Now every open halfspace contains a member of \( \mathcal{V}_+^{(\infty)} \). Hence no open halfspace exists on which the generalized derivative of \( f \) at \( z^{(\infty)} \) is negative. \( \Box \)

**4.1. The differentiable case.** Corollary 4.4 is useful because all \( C^1 \) functions have open halfspaces of descent directions at all nonstationary points. We now look at the case when \( f \) is strictly differentiable [3] at \( z^{(\infty)} \), i.e.,

\[
\exists v \in \mathbb{R}^n \text{ such that } f^\circ(z^{(\infty)};v) = w^T v \quad \forall v \in \mathbb{R}^n.
\]

This yields the following important corollary.
COROLLARY 4.5. If $f$ is strictly differentiable at $z^{(\infty)}$, then $z^{(\infty)}$ is a stationary point of $f$.

Proof. Strict differentiability implies

$$
\exists w \in \mathbb{R}^n \text{ such that } f^\circ(z^{(\infty)}; v) = w^T v \quad \forall v \in \mathbb{R}^n.
$$

If $w$ is nonzero, then $f^\circ(z^{(\infty)}; v)$ is negative on the open halfspace $\{v : w^T v < 0\}$, contradicting Corollary 4.4. The only remaining possibility is that $w = 0$. Hence $z^{(\infty)}$ must be a stationary point of $f$. \qed

The difference between these two corollaries is that Corollary 4.4 can eliminate many points of nondifferentiability from the set of possible cluster points of $\{z^{(m)}\}$, whereas Corollary 4.5 cannot. For example, let $f = \min\{|x|, |x|^2\}$, where the 2-norm has been used. Corollary 4.5 has $x = 0$ and $\{x : \|x\| = 1\}$ as possible cluster points, whereas Corollary 4.4 shows that $x = 0$ is the only possible cluster point.

Clearly if $f$ is continuously differentiable at $z$, then it is also strictly differentiable there, and so Corollary 4.5 establishes the convergence results of [4] and framework A of [6]. It also establishes convergence for methods which do not conform to either [4] or [6]. An example of such a method is any algorithm which uses $E = \epsilon = 0$ and also uses frame centers which are not necessarily members of the current admissible set $\mathcal{G}^{(m)}$. Further examples include any method using $E = \epsilon = 0$ and a grid with hexagonal, triangular, or circular symmetries in some dimensions. An example of an admissible set with both circular and rectangular symmetries is the set of all points in $R^2$ which have either integer Cartesian coordinates $(x_1$ and $x_2$) or have integer values for $r$ and $r\theta/\pi$, where $r$ and $\theta$ are the standard polar coordinates. An admissible set like this could be used with functions that may have both straight grooves and circular grooves centered on the origin. Many other possibilities for the admissible set exist, including those which incorporate random elements. For example, $\mathcal{G}^{(m)}$ could be the set of all points $x + v(x) \in R^n$, where all components of $x$ are integer and where $v(x)$ is a random vector function of $x$ over the set of vectors satisfying $\|v(x)\| \leq 1$.

4.2. The Lipschitz condition. In this subsection the case when $f$ is locally Lipschitz but not differentiable is discussed. Let $f$ be locally Lipschitz with Lipschitz constant $M$ at $z^{(\infty)}$, and let $v$ be a direction satisfying $f^\circ(z^{(\infty)}; v) < 0$ at $z^{(\infty)}$. Let $u$ be a unit vector, and let $\eta \in \mathbb{R}$ be positive. Then

$$
f^\circ\left(z^{(\infty)}; v + \eta u\right) \leq f^\circ\left(z^{(\infty)}; v\right) + \eta M.
$$

This shows that $f^\circ(z^{(\infty)}; \cdot)$ is negative for all directions in a cone containing $v$ in its interior. Hence an algorithm conforming to the template will eventually find a descent direction if it looks along a sequence of directions converging to $v$ as $z$ goes to $z^{(\infty)}$.

It should be noted that the existence of descent directions at a point does not guarantee that $f^\circ$ is negative along these directions. A very simple example is the function $f = -|x|$ in one dimension at the origin. A more interesting example in two dimensions is

$$
f = \begin{cases} 
  r, & |\theta| \geq \theta_0, \\
  r(2|\theta| - \theta_0) / \theta_0, & |\theta| < \theta_0,
\end{cases}
$$

again at the origin. For clarity, this example is described using polar coordinates $r$ and $\theta$, with $r \geq 0$ and $-\pi < \theta \leq \pi$. The function is well defined for all $\theta_0$ values, but we are primarily interested in $0 < \theta_0 < \pi$. For these values $f$ looks like an
upward pointing cone with a notch slanting downwards along $\theta = 0$. The example would be presented to an algorithm as an unconstrained problem in the rectangular coordinates $x_1 = r \cos(\theta)$ and $x_2 = r \sin(\theta)$). Simple calculations show that $f^2$ is positive for every direction whenever $\theta_0 < \pi/2$. However, directions with $2|\theta| < \theta_0$ are descent directions at the origin.

The necessity of the Lipschitz condition can be seen by considering, for example, the function $f = -x_2 + 5\sqrt{|x_1|}$. Elementary calculations show the directional derivative $f'(0; e_2) = -1$, where $e_2$ is the $i$th unit vector. However, if the direction $e_2$ is replaced by the parabolic arc $te(t), where $v(t) = e_2 + te_1$, then

$$f'_{\text{arc}}(0; v(\cdot)) = \lim_{t \to 0} \frac{f(0 + tv(t)) - f(0)}{t} = 4.$$ 

Here the fact that the direction of $v(t)$ alters as $t$ goes to zero means that a descent step is not located even though $v(t)$ becomes parallel to the descent direction $e_2$ as $t$ tends to zero. There is nothing special about keeping the direction constant. Similar calculations with the function $f = -x_2 + 5\sqrt{|x_1| - x_2^2}$ give $f'_{\text{arc}}(0; v(\cdot)) = -1$ and $f''(0; e_2) = 4$. This time the fixed direction fails, and by curving $v(t)$ into the limiting direction $e_2$, a descent step is found. So if $f$ is not locally Lipschitz and lacks any other special properties, then little can be said.

There is one computationally expensive way to attack such problems using the arbitrary finite process in step 4 of the template. The idea is eventually to look everywhere in some neighborhood of each cluster point of the sequence of iterates. Let $x^{(k)} + y, y \in Y^{(k)}$, be the set of points at which $f$ is calculated in the arbitrary finite process in step 4 during iteration $k$.

**Theorem 4.6.** If $f$ is continuous and the sequence of sets $\{Y^{(k)}\}$ satisfies the following two properties,

- **Y1:** the sequence is eventually nested, i.e., $Y^{(k)} \subseteq Y^{(k+1)}$ for all $k$ sufficiently large; and
- **Y2:** there exists a positive constant $\mu$ such that $\bigcup_{k=1}^{\infty} Y^{(k)}$ is dense in the open ball of radius $\mu$ centered on the origin,

then all cluster points of the sequence of iterates are local minimizers of $f$.

**Proof.** The proof is by contradiction. Let $x^{(\infty)}$ be a cluster point of the sequence of iterates which is not a local minimizer. Replace the sequence of iterates $\{x^{(k)}\}$ with an infinite subsequence $\{x^{(k)}\}$ such that all members of this subsequence are within $\mu/3$ of $x^{(\infty)}$, and replace $\{Y^{(k)}\}$ with the corresponding subsequence of itself. We note that this subsequence of $\{Y^{(k)}\}$ satisfies both Y1 and Y2. Property Y1 ensures that property Y2 is not lost when moving to this subsequence. Now there exists a point $x_\mu$ within $\mu/3$ of $x^{(\infty)}$ which is strictly lower than $x^{(\infty)}$. Continuity of $f$ means that there is a ball of strictly positive radius $\xi < \mu/3$ around $x_\mu$ on which $f$ is strictly less than $f(x^{(\infty)})$. Property Y2 means that for some finite $k$ step 4 will evaluate $f$ at a point in the ball of radius $\xi$ about $x_\mu$. This contradicts the fact that the sequence of function values $\{f(x^{(k)})\}$ is monotonically decreasing. \(\square\)

This is not a particularly practical way of ensuring convergence except on very small problems. However, it is one way of gaining some confidence in a solution when $f$ is not smooth.

**4.3. Generalizing sufficient descent.** The sequential algorithms of García-Palomares and Rodríguez [10] conform to Template D except on one point: the choice of sufficient descent condition. Herein the same measure of sufficient descent (i.e., $\epsilon$) is used for all search steps, whereas the prototype sequential algorithms in [10] use
a different value for each search direction. Template D is easily adapted to include these prototype algorithms. This is done by replacing the sequence of constants \( \{ \epsilon(k) \} \) with a sequence of functions \( \{ \epsilon_k(v) \} \). The sufficient descent condition for a step \( h^{(k)}v \) from the iterate \( x^{(k)} \) becomes

\[
\begin{align*}
  f(x^{(k)} + h^{(k)}v) &< f(x^{(k)}) + \epsilon^{(k)}(v), \\
  \lim_{k \to \infty} \left( \sup_{v \in R^n} \epsilon^{(k)}(v) \right)/h^{(k)} &= 0
\end{align*}
\]

A frame \( \Phi \) which contains no point of sufficient descent is quasi-minimal. The sequence \( \{ \epsilon^{(k)}(v) \} \) is required to have the following properties:

\[
\lim_{k \to \infty} \left( \sup_{v \in R^n} \epsilon^{(k)}(v) \right)/h^{(k)} = 0
\] (4.1)

and

\[
\epsilon^{(k)}(v) \geq E^{(m(k))} \quad \forall v \in R^n \quad \forall k.
\]

Here \( m(k) \) is the value of \( m \) at step 3 of iteration \( k \), which is the index of the quasi-minimal frame the template is searching for at iteration \( k \). A corresponding sequence of functions \( \{ \epsilon_x^{(m)}(v) \} \) is also defined, with \( \epsilon_x^{(m)}(v) = \epsilon^{(k(m))}(v) \) for all \( v \), as before.

Equation (4.1) ensures that the proof of Theorem 4.3 is still valid, and the rest of the convergence theory depends only on the lower bounds \( E^{(m)} \), not on \( \epsilon \) itself.

5. The linearly constrained case. Following [1, 13, 18] we develop a theory for the linearly constrained optimization problem (LCOP)

\[
\begin{align*}
  \min_{x \in \Omega} f(x), \quad &\text{where } \Omega = \{ x : a_i^T x + b_i = 0 \quad \forall i = 1, \ldots, q \text{ and } a_i^T x + b_i \leq 0 \quad \forall i = q+1, \ldots, L \}. \\
\end{align*}
\]

(5.1)

We regard any point \( x \in \Omega \) as a solution of (5.1) if and only if no feasible direction exists at \( x \) along which the directional derivative of \( f \) is negative. The constraints defining \( \Omega \) are imposed via a barrier function. A new objective function \( f_c(x) = f(x) + \psi(x) \) is defined, where \( \psi(x) \) is the indicator function for the set \( \Omega \). Hence \( f_c(x) = f(x) \) if \( x \in \Omega \), and \( f_c(x) = \infty \) otherwise. Algorithms conforming to Template D may be applied to \( f_c \); however, the discontinuous nature of \( f_c \) means that Theorem 4.3 does not guarantee convergence to one or more solutions of (5.1). To ensure that an algorithm locates solution(s) of the LCOP we consider a specialization of Template D which requires that each positive basis \( \mathcal{V}_+(m) \) conforms to the shape of the feasible region \( \Omega \) near \( z^{(m)} \). This specialization is presented later as Template E.

For the unconstrained case the crucial feature of a positive basis is that at any point \( x \) it positively spans the set of feasible directions at \( x \) and also at any point near \( x \). For constrained problems we need finite sets of directions with the same property, although, in general, the set of feasible directions is now a closed polyhedral cone rather than \( R^n \). The set of feasible directions can also vary from point to point, in contrast to the unconstrained case.

Template E generates a sequence of feasible iterates which contains an infinite subsequence \( \{ z^{(m)} \} \) of quasi-minimal frame centers. At each frame center the constraints which could be active (i.e., hold with equality) at or near this frame center are identified. The directions in that frame’s positive basis are aligned with the identified set of constraints. More precisely, for any cone of feasible directions defined by a subset of those constraints, there is a subset of the frame’s positive basis which
positively spans that cone of feasible directions. These *aligned* positive bases can be used to extend the convergence theory in section 4 to the linearly constrained case.

For each frame a subset of the constraints is selected which includes those constraints which are active at or near the quasi-minimal center \( z^{(m)} \). This is done by choosing a positive constant \( \delta \) and selecting all constraints with residuals not more than \( \delta \). These constraints are indexed by the working set \( W^{(m)} \) which must satisfy

\[
\left| a_i^T z^{(m)} + b_i \right| \leq \delta \implies i \in W^{(m)},
\]

where \( \delta > 0 \) is independent of \( m \). The feasibility of each \( z^{(m)} \) means that every constraint which is active (which includes all equality constraints) at some point near \( z^{(m)} \) appears in \( W^{(m)} \). Hence, for any \( x \in \Omega \) near \( z^{(m)} \), the set of active constraints at \( x \) is contained in \( W^{(m)} \). The positive basis \( \mathcal{V}_{x}^{(m)} \) is then constructed so that some subset of it positively spans the cone of feasible directions at \( x \). In practice \( W^{(m)} \) would often contain constraints with residuals much greater than \( \delta \). This would assist an algorithm in traversing the boundary of the feasible region more quickly.

The constraints in \( W^{(m)} \) define a polyhedral cone

\[
\mathcal{K}^{(m)} = \left\{ v : a_i^T v = 0 \quad \forall i = 1, \ldots, q \quad \text{and} \quad a_i^T v \leq 0 \quad \forall i = q + 1, \ldots, \ell \right\}
\]

which is the cone of feasible directions at any point in \( \Omega \) for which \( W^{(m)} \) is the active set of constraints. A positive basis for the null space of the equality constraints is constructed which contains a positive basis for any cone (see section 5.1) defined by any subset of \( W^{(m)} \) containing all equality constraints. A positive basis which satisfies these conditions is said to be *aligned* with the set of constraints \( W^{(m)} \) at \( z^{(m)} \) or, more simply, aligned. Occasionally the phrase “aligned with a cone” is used; it means aligned with the set of constraints defining that cone. A frame is constructed by the same process used in (2.1). Any such frame is also called aligned. In the next section the formation of aligned frames is discussed, followed by the barrier approach to linearly constrained problems.

5.1. Generating aligned positive bases and frames. A polyhedral cone \( \mathcal{K} \) may be defined as the intersection of a finite number of halfspaces and hyperplanes:

\[
\mathcal{K} = \left\{ v : a_i^T v = 0, \quad i = 1, \ldots, q \quad \text{and} \quad a_i^T v \leq 0, \quad i = q + 1, \ldots, \ell \right\}.
\]

For convenience we have omitted the \((m)\) superscripts and have assumed that the first \( \ell - q \) inequality constraints are those in the current working set. *In this subsection only,* the constraints under discussion are those defining the cones of feasible directions. These constraints are of the form

\[
a_i^T v = 0, \quad i = 1, \ldots, q \quad \text{and} \quad a_i^T v \leq 0, \quad i = q + 1, \ldots, \ell.
\]

That is to say, the constants \( b_i \) have been omitted from the constraints which define \( \Omega \). Any such cone can be rewritten as a finitely generated cone

\[
\exists v_1, \ldots, v_p \quad \text{such that} \quad \mathcal{K} = \left\{ \sum_{i=1}^{p} \eta_i v_i : \eta_i \geq 0 \quad \forall i = 1, \ldots, p \right\}
\]

as is shown by Theorem 4.18 of [16]. The vectors \( v_1, \ldots, v_p \) are often referred to as a set of *generators* of the cone \( \mathcal{K} \). A minimal set of generators \( \mathcal{V}_+ \) for a closed polyhedral cone \( \mathcal{K} \) is a set of vectors \( \{v_1, \ldots, v_p\} \) such that
K1: \{v_1, \ldots, v_p\} satisfies (5.5) and
K2: no proper subset of \(\mathcal{V}_+\) satisfies (5.5).

Initially we consider the special case where the \(a_i, i \leq \ell\), are linearly independent. A positive basis aligned with \(K\) is constructed in two parts: one each for the subspace containing these \(a_i\) and for the subspace orthogonal to these \(a_i\). For illustrative purposes, choose any basis for \(R^n\) which satisfies \(a_i = e_i\) for \(i = 1, \ldots, \ell\) but is otherwise arbitrary. Here \(e_i\) is the \(i\)th unit vector. If \(\mathcal{U}_+\) is any positive basis for the subspace spanned by \(e_{\ell+1}, \ldots, e_n\), then
\[
\{-e_i : i = q + 1, \ldots, \ell\} \cup \mathcal{U}_+
\]
is a set of generators for \(K\). Interestingly, this is a subset of the following positive basis for the null space of the equality constraints
\[
\mathcal{V}_+ = \{\pm e_i : i = q + 1, \ldots, \ell\} \cup \mathcal{U}_+.
\]
This positive basis for the null space of the equality constraints contains a set of generators for every polyhedral cone defined by the equality constraints and any subset of the constraints \(v^T e_i \geq 0, v^T e_i \leq 0, \text{ and } v^T e_i = 0\) for \(i = q+1, \ldots, \ell\). This property is crucial: it means that \(\mathcal{V}_+\) contains a set of generators for every possible cone of feasible directions at \(z^{(m)}\) and at all points near \(z^{(m)}\).

We now revert back to the original basis for \(R^n\) and work with \(a_i\). The assumption that the set \(\{a_i : i \in W^{(m)}\}\) is linearly independent is retained. For notational simplicity we continue to assume that \(W^{(m)} = \{1, \ldots, \ell\}\). Let \(A = [a_1, \ldots, a_\ell]\) and select an invertible matrix \(S = [s_1, \ldots, s_n]\) satisfying \(S^T A = [e_1, \ldots, e_\ell]\). This allows the following Theorem to be stated.

**Theorem 5.1.** If \(W\) and \(S\) are as defined above, then the set
\[
\{-s_i : i = q + 1, \ldots, \ell\} \cup \{Su : u \in \mathcal{U}_+\}
\]
is an ordered minimal set of generators for the cone \(K\) defined in (5.4). Here \(\mathcal{U}_+\) is an ordered positive basis for the subspace spanned by \(e_{\ell+1}, \ldots, e_n\).

**Proof.** First, it is clear that all members of (5.6) lie in \(K\). We now must show that an arbitrary \(w_1 \in K\) can be expressed as a nonnegative linear combination of the members of (5.6). Since \(w_1 \in K\), it follows that \(A^T w_1 \leq 0\). Moreover, the first \(q\) elements of \(A^T w_1\) must be zero. For convenience let \(A^T w_1 = y\). Now, for appropriate nonnegative choices of \(\eta_i, i = q + 1, \ldots, \ell\), the vector
\[
w_2 = \sum_{i=q+1}^\ell \eta_i (-s_i) \text{ solves } A^T w_2 = \sum_{i=q+1}^\ell -\eta_i e_i = y \leq 0.
\]
Hence \(w_2 - w_1\) is a member of the null space of \(A^T\) (hereafter \(N(A^T)\)). Clearly \(w_2\) is a nonnegative linear combination of the members of (5.6). Moreover, \(\{Su : u \in \mathcal{U}_+\}\) is an ordered positive basis for the null space \(N(A^T)\), and so \(w_1\) can be written as a nonnegative linear combination of the members of (5.6).

Minimality can be seen as follows. For a specific \(j \in q+1, \ldots, \ell\) one has \(A^T (-s_j) = -e_j\), whereas \(e_j^T A^T v = 0\) for all other \(v\) in (5.6). Hence \(-s_j\) cannot be expressed as a nonnegative linear combination of the remaining members of (5.6). Finally, assume some \(Su_j, u_j \in \mathcal{U}_+\) is redundant, i.e.,
\[
(5.7) \quad Su_j = \sum_{i=q+1}^\ell \sigma_i (-s_i) + \sum_{i \neq j} \theta_i Su_i
\]
for some \( \theta_i, \sigma_i \geq 0 \). Now \( A^T S u = 0 \) for all \( u \in \mathcal{U}_+ \), which implies \( \sigma_i = 0 \) for all \( i \). Multiplying (5.7) by \( S^{-1} \) yields a contradiction with the fact that \( \mathcal{U}_+ \) is a positive basis. \( \square \)

**Corollary 5.2.** The set

\[
V_+ = \{ s_i : i = q + 1, \ldots, \ell \} \cup \{ -s_i : i = q + 1, \ldots, \ell \} \cup \{ Su : u \in \mathcal{U}_+ \}
\]

contains a set of generators for any cone defined by the equality constraints and any subset of the inequality constraints in (5.4).

**Proof.** Without loss of generality let the selected subset of inequality constraints be indexed by \( i = q + 1, \ldots, r \), where \( r \leq \ell \). Using \( \mathcal{U}_+ \) as a positive basis for the subspace spanned by \( e_{q+1}, \ldots, e_n \) as above, the set

\[
\{ \pm s_i : i = r + 1, \ldots, \ell \} \cup \{ Su : u \in \mathcal{U}_+ \}
\]

is a positive basis for the null space \( N([a_1 \ldots a_\ell]^T) \). The corollary then follows from Theorem 5.1. \( \square \)

As an illustration, consider the constraints \( x_1 \leq 0 \) and \( x_2 \leq 0 \) in \( \mathbb{R}^2 \). Equation (5.8) gives \( V_+ = \{ \pm e_1, \pm e_2 \} \). There are four possible sets of active constraints: none; \( x_1 \leq 0 \) only; \( x_2 \leq 0 \) only; and both. The sets of generators for the corresponding tangent cones are \( V_+, \{ -e_1, \pm e_2 \}, \{ \pm e_1, -e_2 \}, \) and \( \{ -e_1, -e_2 \} \). Equation (5.8) is used to define each \( V_+^{(m)} \). Corollary 5.2 means that every \( V_+^{(m)} \) is a positive basis for the null space of the set of equality constraints.

When degeneracy is present in a set of active constraints the above approach must be modified. (Readers not interested in the degenerate case may wish to proceed directly to Assumption 5.3.) The existence of an aligned positive spanning set \( V_+^{ld} \) is guaranteed by Theorem 4.18 of [16], but its construction can be computationally expensive [13]. The superscript “\( ld \)” is used to highlight the fact that \( V_+^{ld} \) is not necessarily a positive basis and is no longer defined by (5.8). The set \( V_+^{ld} \) must contain a set of generators for the cone \( \mathcal{K} \) in (5.4) and also for every cone defined by any subset of the constraints in \( W \) which includes all equality constraints. If the constraints are linearly dependent, then \( V_+^{ld} \) is a positive spanning set for the subspace defined by the equality constraints, but it is no longer a positive basis. For convenience, in the following discussion we assume any linear dependence in the subset of equality constraints has been removed by deleting redundant equality constraints.

The construction of \( V_+^{ld} \) is in two parts. The first part is a positive basis for the null space of the normals of the constraints indexed by \( W \). The second part is for the subspace \( T \) spanned by \( a_{q+1}, \ldots, a_\ell \), where \( T \) is of dimension \( r - q \).

For the first part of \( V_+^{ld} \), order the constraints so that \( a_1, \ldots, a_{q+r} \) are linearly independent. Let the invertible matrix \( S = [s_1, \ldots, s_n] \) satisfy \( S^T[a_1, \ldots, a_r] = [e_1, \ldots, e_r] \). The first part of \( V_+^{ld} \) is

\[
Su_+ = \{ Su : u \in \mathcal{U}_+ \},
\]

where \( \mathcal{U}_+ \) is a positive basis for the subspace spanned by \( e_{r+1}, \ldots, e_n \). Clearly \( Su_+ \) positively spans the null space \( N([a_1 \ldots a_\ell]^T) \).

The second part of \( V_+^{ld} \) contains a positive scalar multiple of each vector \( \pm v \) which lies in \( T \) and satisfies with equality any \( r - 1 \) linearly independent constraints indexed by the set \( W \), including all equality constraints. Note that \( \pm v \) may violate
any constraint it is not required to satisfy with equality. Clearly \( \mathcal{V}_{d}^d \) must contain the \( \mathcal{V}_{+} \) defined by (5.8) for each subset of \( W \) which includes all equality constraints and has \( r \) linearly independent constraint normals. The set of all such vectors for all such subsets of \( W \) includes many pairs of vectors which are positive scalar multiples of one another. Eliminating such pairs gives \( \mathcal{V}_{d}^d \). In the particular case when all constraint normals indexed by \( W \) are linearly independent, these vectors \( \pm v \) are positive scalar multiples of \( \pm s_{q+1}, \ldots, s_{\ell} \) in (5.8), and the definition of \( \mathcal{V}_{d}^d \) reverts back to that in (5.8).

It is now shown that \( \mathcal{V}_{d}^d \) contains a set of generators for every cone defined by any subset of the constraints in \( W \) which includes all equality constraints. Consider a cone \( \mathcal{K}_s \) defined by a subset \( W_s \) of \( W \) which contains all equality constraints but is otherwise arbitrary. Let the dimension of \( \mathcal{K}_s \) be \( \rho \). For convenience reorder the inequality constraints so that \( a_1, \ldots, a_\rho \) are linearly independent, and \( 1, \ldots, \rho \in W_s \). Define \( T_s \) to be the subspace spanned by \( a_{q+1}, \ldots, a_\rho \). We now add further constraint normals \( a_i, i \in W \), to \( a_1, \ldots, a_\rho \) to obtain a maximal linearly independent set \( a_1, \ldots, a_r \). The construction of \( \mathcal{V}_{d}^d \) ensures that it contains an ordered positive basis (given by (5.8)) defined by the working set \( \{1, \ldots, r\} \). Hence \( \mathcal{V}_{d}^d \) must contain a positive basis for the null space of \( a_1, \ldots, a_\rho \) by Corollary 5.2.

It remains to show that \( \mathcal{V}_{d}^d \) contains a set of generators for the cone \( \mathcal{K}_s \cap T_s \). Define the hyperplane \( \mathcal{H} = \{v \in T_s : (a_{q+1} + \cdots + a_\rho)^Tv = -1\} \). Clearly \( \mathcal{K}_s \cap T_s \) is contained in the cone \( \{v \in T_s : a_i^Tv \leq 0 \text{ for all } i = q+1, \ldots, \rho\} \). Also, because \( a_{q+1}, \ldots, a_\rho \) is a basis for \( T_s \), it is clear that \( \mathcal{K}_s \cap T_s \cap \mathcal{H} \) is bounded, and hence is a polytope \( P \). It can be shown [16] that a set of generators for \( \mathcal{K}_s \cap T_s \) is precisely the set of vectors from the origin to the vertices of \( P \). Each of these vectors \( v \) satisfies \( a_i^Tv = 0 \) for all but one of \( i = 1, \ldots, \rho \). By adding an appropriate member of \( N([a_1, \ldots, a_\rho]^T) \) to \( v \) one can obtain a vector \( v_+ \) which satisfies \( a_i^Tv_+ = 0 \) for all but one \( i \in 1, \ldots, r \). Hence each such \( v_+ \) is a positive scalar multiple of a member of \( \mathcal{V}_{d}^d \). Thus \( \mathcal{V}_{d}^d \) contains an ordered set of generators for every cone \( \mathcal{K}_s \) defined by any subset of constraints in \( W \) which includes all equality constraints.

The following assumption is needed to ensure the limits of the sequence of positive bases have the required properties.

**Assumption 5.3.**

(a) All limits of the sequence of ordered positive bases \( \{\mathcal{U}_+^{(m)}\} \) are ordered positive bases.

(b) The methods used to generate \( S \) and \( \mathcal{V}_+ - S\mathcal{U}_+ \) are repeatable. That is to say, they will always return the same \( S \) and \( \mathcal{V}_+ - S\mathcal{U}_+ \) when given the same working set \( W \).

(c) Each \( \mathcal{V}_+^{(m)} \) satisfies (2.4).

This assumption is the equivalent of Assumption 2.1 for the LCOP (5.1). In the case when constraints are absent, \( \mathcal{U}_+^{(m)} = \mathcal{V}_+^{(m)} \) for all \( m \), and Assumption 5.3 reduces to Assumption 2.1. For the case when the normals of the constraints in \( W \) are linearly independent Assumption 5.3(b) amounts to returning the same \( S \) when given the same \( W \). It is possible that some members of \( \mathcal{V}_+^{(m)} \) violate the bound in (2.4). This bound is imposed retrospectively by replacing any \( v \in \mathcal{V}_+^{(m)} \) violating (2.4) with \( Kv/\|v\| \).

### 5.2. The template for linearly constrained problems.

The following template lists the specialized form of Template D required for linearly constrained problems.
Algorithm Template E.
1. Initialize: set \( k = 1, m = 1 \), and choose the initial point \( x^{(0)} \in \Omega \). Choose \( \delta \geq 0 \).
2. Choose \( H^{(m)} > 0 \), \( E^{(m)} \geq 0 \), and \( \mathcal{G}^{(m)} \).
3. Choose \( h(k) \geq H^{(m)} \) and \( \varepsilon(k) \geq E^{(m)} \).
4. Execute any finite process which satisfies one of these conditions:
   (a) generates an iterate \( x^{(k)} \in \Omega \cap \mathcal{G}^{(m)} \) satisfying \( f(x^{(k)}) < f^{(k-1)} - \varepsilon(k) \); or
   (b) generates a quasi-minimal frame \( \Phi^{(m)} = \{z^{(m)}, h^{(m)}_z, \gamma^{(m)}_z\} \), where \( x^{(k)} \in \Omega \) and \( f^{(k)} \leq f^{(k-1)} \). Here \( z^{(m)} = x^{(k)}, h^{(m)} = h^{(k)} \), and \( \epsilon^{(m)} = \epsilon(k) \). The frame \( \Phi^{(m)} \) must be aligned with an identified working set \( W^{(m)} \) satisfying (5.2); or
   (c) case (b) of this step with the added restriction \( \Phi^{(m)} \in \mathcal{G}^{(m)} \).
5. If \( x^{(k)} \) is not quasi-minimal, increment \( k \) and go to step 3.
6. Increment \( m \) and \( k \). If stopping conditions are not satisfied, go to step 2.

If the active constraint normals are linearly independent at every point on the boundary of \( \Omega \), then the constant \( \delta \) in Template E can be defined implicitly. The case when all constraint normals are linearly independent is trivial. For the remaining case, the equality constraints are indexed by \( i = 1, \ldots, q \) as above, and the inequality constraints are ordered so that \( |a^T z^{(m)} + b| \) is an increasing function of \( i \). The working set \( W \) is chosen as the largest set \{1, \ldots, r\} for which the corresponding constraint normals are linearly independent. The residual of the \( r + 1 \)st constraint in this list must have a uniform positive lower bound for all feasible \( z \) in an arbitrary compact set \( \Xi \), and \( \delta \) can be chosen as this bound. If this were not the case there would be a linearly dependent set of constraints indexed by \( W \), say, and also a sequence of points \( \{z_j\} \subset \Omega \) for which

\[
\lim_{j \to \infty} \left( \max_{i \in W} |a^T z_j + b_i| \right) = 0.
\]

Continuity of the constraint functions then implies a degenerate point exists on the boundary of \( \Omega \), contradicting the initial assumption. Assumption 4.2(a) ensures a compact set \( \Xi \) exists which contains all points of interest.

5.3. Convergence results for the linearly constrained case. Algorithms conforming to Template D may be applied to the LCOP by applying such methods to the barrier function \( f_c \). The non-Lipschitz nature of \( f_c \) means that Theorem 4.3 is not directly applicable, and convergence to solution points of the LCOP (5.1) must be established some other way. In order to guarantee convergence to solution(s) of the LCOP under standard conditions a further restriction must be imposed. Specifically, each frame \( \Phi^{(m)} \) generated by such an algorithm must be aligned with the working set of constraints \( W^{(m)} \). This working set includes all constraints with small residuals \( (\leq \delta) \) at \( z^{(m)} \). With this restriction, Template D becomes Template E.

First we note that Theorem 4.1 is directly applicable, and parts (a) and (b) of Assumption 4.2 ensure the sequence of quasi-minimal iterates \( \{z^{(m)}\} \) is infinite and has cluster points. Next we establish a constrained version of Theorem 4.3.

Definition 5.4. Let \( z^{(\infty)} \) be a cluster point of the sequence of quasi-minimal iterates. Define \( S(\infty)(W) \) as the set of all vectors \( v \) for which there exists an infinite subsequence \( \{(z^{(m)}, v^{(m)})\}_{m \in M} \) with the following properties:

(i) this subsequence converges uniquely to \( (z^{(\infty)}, v) \);
(ii) \( z^{(m)} + h_z^{(m)}v^{(m)} \in \Omega \) for all \( m \in \mathcal{M} \);
(iii) \( v^{(m)} \in \mathcal{S}_{+}^{(m)} \) for all \( m \in \mathcal{M} \), where \( \mathcal{S}_{+}^{(m)} \) is as defined in (3.1);
(iv) \( W^{(m)} = W \) for all \( m \in \mathcal{M} \);
(v) \( \|z^{(m)} - z^{(\infty)}\| < \gamma \) for all \( m \in \mathcal{M} \), where \( \gamma > 0 \);
(vi) no point in the closed ball of radius \( 2\gamma \) about \( z^{(\infty)} \) violates any constraint not in the active set for the point \( z^{(\infty)} \); and
(vii) \( h_z^{(m)} < \gamma/K \), where \( K \) is the constant used in the upper bound (2.4) on each \( \|v^{(m)}\| \).

The notation \( \mathcal{S}_{+}^{(\infty)} \) is used to denote the union of all \( \mathcal{S}_{+}^{(\infty)}(W) \) when \( W \) ranges over all possible subsets of the set of constraint indices \( \{1, \ldots, L\} \).

Condition (ii) in Definition 5.4 requires that \( v \) be a feasible direction at \( z^{(\infty)} \). Note that the finiteness of the number of different working sets \( W \) means that any \( \mathcal{M} \) satisfying all conditions except (iv) will have an infinite subset which satisfies all seven conditions. Hence condition (iv) does not exclude any \( v \) from \( \mathcal{S}_{+}^{(\infty)} \). For any \( z^{(\infty)}, v \), conditions (iv)–(vii) can always be satisfied for an appropriate choice of \( \mathcal{M} \), provided the first three conditions can. This follows directly from condition (i) and the facts that the number of constraints is finite and \( h_z^{(m)} \to 0 \) as \( m \to \infty \). Conditions (v)–(vii) mean that constraints not in \( W \) are automatically satisfied by all \( z^{(m)} + h_z^{(m)}v \) when \( v \in \mathcal{S}_{+}^{(m)} \) and \( m \in \mathcal{M} \). In particular, this includes all points in the frames \( \Phi^{(m)}, m \in \mathcal{M} \). Conditions (i)–(iii) are needed for the proof of Theorem 5.5. The last four conditions are superfluous to the proof of Theorem 5.5 but are needed in the proof of Theorem 5.6.

**Theorem 5.5.** Let \( z^{(\infty)} \) be a cluster point of the sequence of quasi-minimal iterates. Then \( f^c(z^{(\infty)}, v) \geq 0 \) for all \( v \in \mathcal{S}_{+}^{(\infty)} \).

**Proof.** Since every \( z^{(m)} \) is feasible, condition (ii) of Definition 5.4 allows us to use the fact that \( f \equiv f_c \) on \( \Omega \). Conditions (i) and (iii) of Definition 5.4 allow Theorem 4.3 to be invoked, yielding the required result. \( \square \)

Next it is shown that Theorem 5.5 applies to a set of directions rich enough to include a set of generators for the cone of feasible directions at \( z^{(\infty)} \).

**Theorem 5.6.** Let \( z^{(\infty)} \) be a limit point of the sequence of quasi-minimal iterates. The set \( \mathcal{S}_{+}^{(\infty)} \), as defined in Definition 5.4, contains a set of generators for the cone of feasible directions \( \mathcal{K}^{(\infty)} \) at the limit point \( z^{(\infty)} \).

**Proof.** Let \( W^{(\infty)} \) be the set of active constraints at \( z^{(\infty)} \). Consider an infinite increasing sequence of positive integers \( \mathcal{M} \) with the following properties:

(a) \( z^{(m)} \to z^{(\infty)} \) as \( m \to \infty \), \( m \in \mathcal{M} \);
(b) \( \mathcal{V}_{+}^{(m)} \to \mathcal{V}_{+}^{(\infty)} \) as \( m \to \infty \), \( m \in \mathcal{M} \);
(c) \( W^{(m)} \) is the same for all \( m \in \mathcal{M} \);
(d) \( \|z^{(m)} - z^{(\infty)}\| < \gamma \) for all \( m \in \mathcal{M} \), where \( \gamma \) is a positive constant;
(e) no point in the closed ball of radius \( 2\gamma \) centered on \( z^{(\infty)} \) violates any constraint not in \( W^{(\infty)} \); and
(f) \( h_z^{(m)} < \gamma/K \) for all \( m \in \mathcal{M} \), where \( K \) is the constant in (2.4).

The existence of \( \mathcal{M} \) is guaranteed by the following facts: (a) holds because \( z^{(\infty)} \) is a limit point of \( \{z^{(m)}\} \); (b) holds by Assumption 5.3; (c) and (e) hold because the number of different possible working sets is finite; (d) follows from (a); and (f) follows from Assumption 4.2(c).

If attention is restricted to the sequence \( \{v_i^{(m)}\}_{m \in \mathcal{M}} \) for a fixed value of \( i \), then (a) and (b) together, and (c), (d), (e), and (f), respectively, yield items (i), (iv), (v),
(vi), and (vii) of Definition 5.4. The fact that $\mathcal{V}'^{(m)}$ is contained in $\mathcal{S}'^{(m)}$ for all $m$ yields item (iii) of Definition 5.4, which leaves just condition (ii).

Now (5.2) implies $W^{(\infty)} \subseteq W^{(m)}$ for all $m$ such that $z^{(m)}$ is sufficiently near $z^{(\infty)}$. Hence (a) and (c) imply $W^{(\infty)} \subseteq W^{(m)}$ for all $m \in \mathcal{M}$. Therefore each $\mathcal{V}'^{(m)}$, $m \in \mathcal{M}$, contains a set of generators for the cone $\mathcal{K}^{(\infty)}$ of feasible directions at $z^{(\infty)}$. Each such set of generators for $\mathcal{K}^{(\infty)}$ consists of two parts. The first part (hereafter $\mathcal{W}_+$) is the part contained in the span of $\{u_i : i \in W^{(m)}\}$ and the second part is $SU^{(m)}_+ = \{ Su : u \in U^{(m)}_+ \}$. The construction of the first part depends only on $W^{(m)}$, and so $\mathcal{W}_+$ is the same for all $m \in \mathcal{M}$ by Assumption 5.3(b). The matrix $S$ is also independent of $m$, again by Assumption 5.3(b). Item (b) means that $\{U^{(m)}_+\}_{m \in \mathcal{M}}$ has a unique limit $U^{(\infty)}_+$, which is an ordered positive basis, by Assumption 5.3(a). Clearly $\mathcal{W}_+$ and $SU^{(\infty)}_+ = \{ Su : u \in U^{(\infty)}_+ \}$ are both subsets of $\mathcal{S}^{(\infty)}(W^{(m)})$. The set $\mathcal{W}_+ \cup SU^{(\infty)}_+$ is also a set of generators for $\mathcal{K}^{(\infty)}$.

Each $z^{(m)}$ lies in $\Omega$, and each member of $\mathcal{W}_+ \cup SU^{(m)}_+$ lies in $\mathcal{K}^{(\infty)}$. Hence $z^{(m)} + h_k^{(m)}v$ does not violate any constraint indexed by the set $W^{(\infty)}$ for all $m \in \mathcal{M}$ and for all $v \in \mathcal{W}_+ \cup SU^{(m)}_+$. Items (d)–(f) and the bound on $\|v\|$ in (2.4) imply no member of $\mathcal{W}_+ \cup SU^{(m)}_+$, $m \in \mathcal{M}$, can violate any constraint not in $W^{(\infty)}$. Thus all conditions of Definition 5.4 hold for all $v^{(m)} \in \mathcal{W}_+ \cup SU^{(m)}_+$ for every $m \in \mathcal{M}$. Their limits $\mathcal{W}_+ \cup SU^{(\infty)}_+$ are a set of generators for $\mathcal{K}^{(\infty)}$ contained in $\mathcal{S}^{(\infty)}$, as required.

Theorems 5.5 and 5.6 show that a set of generators for the cone $\mathcal{K}^{(\infty)}$ exists such that $f^\circ$ is nonnegative at $z^{(\infty)}$ along each of these generators, where $\mathcal{K}^{(\infty)}$ is the cone of feasible directions at $z^{(\infty)}$. The following result extends this to all feasible directions in the case when $f$ is strictly differentiable at $z^{(\infty)}$.

**Theorem 5.7.** If $f$ is strictly differentiable at $z^{(\infty)}$, then no feasible direction exists at $z^{(\infty)}$ along which $f$ has a negative directional derivative.

**Proof.** Now, for a general $v \in \mathcal{K}^{(\infty)}$, we can write

$$v = \sum_{i=1}^{P} \eta_i v_i^{(\infty)}, \quad \text{where} \quad \eta_i \geq 0 \quad \forall i$$

and where $\{v_1^{(\infty)}, \ldots, v_p^{(\infty)}\} \subseteq \mathcal{S}^{(\infty)}_+$ is a set of generators for $\mathcal{K}^{(\infty)}$. The strict differentiability of $f$ at $z^{(\infty)}$ yields

$$\nabla f^T \nabla \mathcal{V} = \sum_{i=1}^{P} \eta_i \nabla f^T v_i^{(\infty)} = \sum_{i=1}^{P} \eta_i f^\circ \left(z^{(\infty)}; v_i^{(\infty)}\right) \geq 0,$$

as required. Hence no feasible direction exists at $z^{(\infty)}$ along which $f$ has a negative directional derivative, and $z^{(\infty)}$ is a solution of the LCOP (5.1).

For the moment we continue to consider a subsequence of quasi-minimal iterates as defined in the proof of Theorems 5.5–5.7. It is shown in the proof of Theorem 5.6 that $z^{(m)} + h_k^{(m)}v \in \Omega$ for every $v \in \mathcal{W}_+ \cup SU^{(m)}_+$ when $m \in \mathcal{M}$. In contrast, in early iterations $z + hv$ can easily violate constraints not in $W^{(\infty)}$. When this occurs $f_c = +\infty$ at $z + hv$, and the direction $v$ is effectively ignored. Rather than do this, one could evaluate $f$ at $z + \alpha v$, where $\alpha \in R$ is the largest value such that $z + \alpha v \in \Omega$. Any such function evaluations can be included in the arbitrary finite process of step 4.
of the template, which means that Theorem 5.6 still applies. The advantage of such function evaluations is that an algorithm can look along the direction \( v \) immediately, rather than having to wait until \( h \) is small enough to make \( z + hv \) feasible.

6. Selecting the frame size. Template D imposes a number of restrictions on \( h \). In addition to the explicit requirement that \( h \) tend to zero, there is also a sequence of lower bounds \( \{H^{(m)}\} \). These lower bounds are not required for convergence purposes, but other lower bounds on \( h \) are implicit in Assumption 4.2(d) and condition G2. The presence of the explicit lower bounds on \( h \) in Template D is to reinforce the fact that the implicit lower bounds exist. These implicit lower bounds are discussed first, and the cases \( E = 0 \) and \( E > 0 \) are treated separately.

If \( E = 0 \), then condition G1 and the bound (2.4) mean that condition G2 can not be satisfied if \( h \) is too small. In practice one could define \( G^{(m)} \) using a length \( H^{(m)} \) which would become a lower bound for \( h \) until the next quasi-minimal frame is located. For example, in [6]

\[
G^{(m)} = \left\{ x_0 + h \sum_{i=1}^{n} \eta_i v_i : \eta_i \text{ is integer } \forall i = 1, \ldots, n \right\}
\]

is used, where \( x_0 \) is the origin of the grid and where \( v_1, \ldots, v_n \) are a basis for \( R^n \). In [6] \( h \) is used to define both the admissible set \( G^{(m)} \) and also the quasi-minimal frame \( \Phi \) contained in that set. Therefore in [6], \( h \) is kept constant between quasi-minimal frames. Under Template D the \( h \) value used to define the grid would become the lower bound \( H^{(m)} \), and \( h \) values in excess of this would be permitted.

If \( E > 0 \), then the requirement that \( \epsilon/h \to 0 \) means that \( h \) must approach zero more slowly than \( E \). The simplest method of ensuring this is to connect \( \epsilon \) and \( h \) via a relation like \( \epsilon = NH^\nu \), where \( N > 0 \) and \( \nu > 1 \). The bound \( \epsilon \geq E \) is then equivalent to a positive lower bound on \( h \). In fact, in [4] the lower bound on \( \epsilon \) is imposed indirectly via this relation and a specific positive lower bound \( H \) on \( h \).

The convergence theory requires that \( h \to 0 \) as \( k \to \infty \) but does not state how this is to be done. Simple approaches such as using \( h^{(k)} = 2^{-k} \) have obvious drawbacks. Indeed, \( h \) permanently falls below machine precision after a fixed number of iterations. Such an approach takes no account of how quickly or slowly the sequence of iterates is converging. When a solution is located quickly \( h \) should become small quickly in order to verify that it is indeed a solution. In contrast, if good reductions in \( f \) occur with \( h \) large, then \( h \) should remain large until such reductions cease. Similarly, if the sequence of iterates moves from a region where small steps are necessary into a region where large steps are better, then \( h \) should increase. This suggests that \( h \) should vary in sympathy with the lengths of recent steps and also with the recent reductions in \( f \). One possibility is to impose an upper bound on \( h \) of the form

\[
(6.1) \quad h^{(k)} \leq \max \left\{ \gamma H^{(m(k))}, \frac{\gamma}{\alpha} \right\}.
\]

Here \( \gamma_f \) and \( \gamma_x \) are moving averages of the past decreases in function values and step lengths, respectively, and \( \gamma \) is a constant satisfying \( \gamma \geq 1 \). The value \( m(k) \) is the value of \( m \) at step 3 of iteration \( k \). The two moving averages are defined in terms of two sequences \( \{\omega_i\}_{i=1}^{\infty} \) and \( \{\beta_i\}_{i=1}^{\infty} \) of nonnegative weights as follows:

\[
\gamma_f^{(k)} = \sum_{i=1}^{k-1} \omega_{k-i} \left| f^{(i+1)} - f^{(i)} \right| \quad \text{and} \quad \gamma_x^{(k)} = \sum_{i=1}^{k-1} \beta_{k-i} \left\| x^{(i)} - x^{(i-1)} \right\|,
\]
with the convention that \( \Upsilon_f^{(1)} = \Upsilon_x^{(1)} = 0 \). The two sequences of weights chosen are so that
\[
\sum_{i=1}^{\infty} \omega_i \quad \text{and} \quad \sum_{i=1}^{\infty} \beta_i
\]
are both finite. If the sequence of iterates is bounded, then the sequence of \( \Upsilon_f^{(k)} \) values is bounded. If \( f \) is bounded below on any bounded set, then the sequence of \( \Upsilon_f^{(k)} \) values must converge to zero. Given the quadratic nature of smooth functions near local minima, one could replace \( \Upsilon_f^{(k)} \) with its square root in (6.1).

An alternative for the case when \( E \) is always positive is presented in [4]. There \( \epsilon = Nh^\nu \) is used, with \( \nu > 1 \), and \( N > 0 \). The bound \( \epsilon \geq E \) is imposed indirectly by imposing a strictly positive lower bound \( H^{(m)} \) on \( h \). When sufficient descent is obtained \( h \) may be increased by up to a fixed multiple of itself. If a quasi-minimal frame is located, then \( h \) is reduced in such a way that if \( h \) is reduced repeatedly, then \( h \rightarrow 0 \). In essence, if the sequence \( \{h\} \) has a strictly positive lower bound (\( h_{\text{min}} \), say), then \( f \) is reduced by at least \( Nh_{\text{min}}^\nu \) an infinite number of times. Hence either \( h \rightarrow 0 \) or \( f \rightarrow -\infty \). More details are presented in [4].

Template D and related algorithm frameworks gain much flexibility by not making \( h \rightarrow 0 \) a direct consequence of conforming to the template. In contrast, GPS [17] guarantees \( h \rightarrow 0 \) when \( f \) is \( C^1 \) and the sequence of iterates is bounded. If Template D satisfies (6.1) or uses the approach in [4], then it also guarantees \( h \rightarrow 0 \) under the same conditions.

7. Concluding remarks. Template D contains algorithms which bear a striking resemblance to the implicit filtering algorithms in [2, 11]. These implicit filtering algorithms use the positive basis \( \{\pm e_1, \ldots, \pm e_n\} \) to form frames (or, in the language of [2, 11], stencils) about the current iterate \( x \). Using the frame an estimate \( g \) of the gradient at \( x \) is formed, and a search direction is generated. A finite line search is conducted along this direction, where satisfaction of a sufficient descent condition is sought. If sufficient descent is not obtained, if the gradient at \( x \) is small, or (in [2]) if the frame is minimal, then \( h \) is reduced. Given that convergence occurs, [11] shows that convergence rate is linear for bound constrained problems when the line search direction is \( -g \), except that any infeasible point in the line search is replaced with the closest feasible point to it. In the absence of bounds, [2] shows that a superlinear rate can be obtained when a quasi-Newton search direction is used.

Implicit filtering differs from Template D on a number of points, including the type of sufficient descent condition in the line search and that frame points are not considered as candidates for the next iterate. This last difference has enormous theoretical implications. The possibility of a frame point becoming the next iterate rather than a line search point means that the rate theorems of implicit filtering are not guaranteed to apply to any algorithm conforming to Template D. The absence of steps to frame points in implicit filtering means that the convergence theory behind Template D is inapplicable to implicit filtering. Steps to frame points are crucial to the convergence theory, and so implicit filtering in its current form [2, 11] falls outside the scope of Template D.

Nevertheless, from a practical perspective, implicit filtering is very similar to some algorithms conforming to Template D. Minor modifications to implicit filtering would make it conform to Template D, and hence provably convergent. In the case when steps to frame points occur only in early iterations (numerical experiments in [8]
suggest that this is common) the rate theorems of implicit filtering would apply. The numerical results for implicit filtering [2, 11] and the work on the global aspects of finite differences [19] show that the use of frames can enable algorithms to “step over” many local minima to find a much lower minimum.

A frame-based template for unconstrained and linearly constrained optimization has been developed. Applicability to linearly constrained problems is achieved by aligning frames with active and nearly active constraints. The use of frames means that clearly inactive constraints can be ignored, and the linear constraints can involve irrational numbers, in contrast to [1, 13]. It has been shown that algorithms conforming to the template generate sequences of quasi-minimal iterates whose cluster points are stationary points of the optimization problem under mild conditions. The cluster points of the sequence of quasi-minimal iterates retain interesting properties even when the objective function is not differentiable.

The approach taken unifies methods using sufficient descent and simple descent. The former use the sufficient descent condition to ensure quasi-minimal frames are generated. The latter do so when necessary by restricting the frame points (but not the frame centers) to admissible sets. The frame centers are not restricted, which allows these simple descent methods to select quasi-minimal iterates which lie outside of the admissible set every iteration. The facts that the admissible sets can be unrelated to one another, can incorporate random elements, and can sometimes yield quasi-minimal frames outside the admissible set means that for some algorithms conforming to Template D there is no “pattern” restricting the locations of iterates. All that can be said is that the admissible sets get finer as h approaches zero. This is a level of flexibility not present in previous simple descent methods such as GPS [13, 17] or [6]. These earlier simple descent methods also specifically use rectangular grids or subsets of them. The greater choice of admissible sets allows these sets to possess, for example, circular or spherical symmetries in some dimensions. This could be very useful when, for example, minimizing a quadratic penalty function involving nonlinear constraints with known symmetries.

Template D encompasses a wide class of algorithms including existing frame-based and grid-based methods. Numerical results for existing methods in this class [5, 8, 15] show that there are effective methods conforming to Template D. There is much scope for future work in developing algorithms which exploit the great flexibility afforded by the template.

Acknowledgments. The authors wish to thank Professor John Dennis for some very useful suggestions, including those concerning the use of Clarke’s nonsmooth analysis. The authors also wish to thank Professor Tim Kelley for drawing implicit filtering to their attention and three anonymous referees for their many helpful comments and suggestions.

REFERENCES


