Optimization and Engineering, 5, 123–144, 2004

Direct Search Methods for Nonlinearly Constrained Optimization Using Filters and Frames

JOHN E. DENNIS JR.
CAAM, Rice University, Houston, Texas, USA
email: dennis@rice.edu

CHRISTOPHER J. PRICE
IAN D. COOPE
Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
email: c.price@math.canterbury.ac.nz
email: i.coope@math.canterbury.ac.nz

Received January 8, 2003; Revised September 17, 2003

Abstract. A direct search method for nonlinear optimization problems with nonlinear inequality constraints is presented. A filter based approach is used, which allows infeasible starting points. The constraints are assumed to be continuously differentiable, and approximations to the constraint gradients are used. For simplicity it is assumed that the active constraint normals are linearly independent at all points of interest on the boundary of the feasible region. An infinite sequence of iterates is generated, some of which are surrounded by sets of points called bent frames. An infinite subsequence of these iterates is identified, and its convergence properties are studied by applying Clarke's non-smooth calculus to the bent frames. It is shown that each cluster point of this subsequence is a Karush-Kuhn-Tucker point of the optimization problem under mild conditions which include strict differentiability of the objective function at each cluster point. This permits the objective function to be non-smooth, infinite, or undefined away from these cluster points. When the objective function is only locally Lipschitz at these cluster points it is shown that certain directions still have interesting properties at these cluster points.

Keywords: derivative free optimization, positive basis methods, non-smooth convergence analysis, frame based methods, filter

1. Introduction

Many approaches have been proposed which address the problem of minimizing a nonlinear objective function subject to nonlinear constraints. The great majority of these approaches make use of the gradients of the objective and constraint functions. Herein we are interested in methods which do not use or approximate the objective function's gradient, and which only require approximations to the constraint functions' gradients. A number of approaches suggest themselves such as simulated annealing, and genetic and evolutionary algorithms (see e.g., Jenkins, 2002; Kim and Myung, 1996; Peng et al., 1998). Other possibilities also exist. For example, one could perform successive minimizations of a quadratic penalty function using a derivative free unconstrained optimization algorithm. More sophisticated approaches which are provably convergent have been examined by several authors (Audet and Dennis, 2003; Lewis and Torczon, 2002; Lucidi et al., 2002; Yu and Li, 1981).

Lewis and Torczon (2002) make use of existing theory by using an augmented Lagrangian approach which does not require objective or constraint gradients. Each augmented Lagrangian is minimized using the derivative free generalized pattern search method for bound constrained optimization described in Lewis and Torczon (1999). Use of the Hestenes–Powell formula (Hestenes, 1969; Powell) (see also Fletcher (1987), p. 291) allows the Lagrange multipliers and penalty parameters to be updated without using or estimating derivatives. Lewis and Torczon show that the method is convergent.

Yu and Li (1981) and Lucidi et al. (2002) propose methods which require constraint gradients, but not the gradient of the objective function. Both consider only inequality constraints, and generate sequences of iterates that are feasible. At each iterate search directions are formed, where the non-negative linear combinations of these search directions span an appropriate cone of feasible directions. Steps along these search directions may give rise to infeasible points. When this occurs feasibility is regained through correction steps (Yu and Li, 1981) or use of search arcs (Lucidi et al., 2002) tangential to the search directions at the current iterate. A step along each search direction in turn is tried until a point of sufficient descent is located. Each method uses a different measure of sufficient descent: Lucidi et al. (2002) requires a decrease proportional to the square of a step length parameter, whereas (Yu and Li, 1981) require a fixed decrease. In Lucidi et al. (2002) each feasible search arc has its own step length parameter, and each unsuccessful step reduces that arc's step length parameter. In contrast, (Yu and Li, 1981) uses the same step length parameter for all search steps, and reduces it only after all search steps are unsuccessful. Both Lucidi et al. (2002) and Yu and Li (1981) demonstrate convergence, with (Yu and Li, 1981) making the simplifying assumption of linearly independent active constraint normals.

Audet and Dennis (2003) describe an extension of generalized pattern search (Torczon, 1987) which uses filters (Fletcher and Leyffer, 2002; Fletcher et al., 2002) to adjudicate between the two aims of minimizing the objective function and minimizing a measure of infeasibility. The use of filters allows (Audet and Dennis, 2003) to consider infeasible points. Loosely speaking, a filter is a finite set of points, none of which is better than any other in both objective function value and constraint violation. An iteration of an algorithm could, for example, select a filter point and generate a new point using a local model of the problem around that filter point. This new point would be rejected by the filter if some existing filter point is better than it. Otherwise it would be accepted, and added to the filter. In contrast a more traditional approach would accept the new point only if it was better than the selected filter point. The filter approach gives an algorithm much more freedom to advance one of the two aims (reducing f and reducing the constraint violation) at the expense of the other. The choice of model would be required to ensure that eventually better points were generated.

This method (Audet and Dennis, 2003) retains the way that generalized pattern search chooses search directions: a finite set $\mathcal{D}$ of search directions is selected before the first iteration, and a subset of these directions is used at each iteration. The set $\mathcal{D}$ is required to possess the property that the number of distinct linear combinations of members of $\mathcal{D}$ with non-negative integer coefficients is finite in any bounded region of R^n. Audet and Dennis

(2003) analyse this method using the non-smooth calculus of Clarke (1990), and show that a subsequence of iterates converges to one or more Karush-Kuhn-Tucker points of a problem *related* to the original problem. This situation is not ideal, and is caused by the fact that at each iteration, the method draws the search directions from a finite set of directions $\mathcal{D}$ which is fixed for all iterations. Hence the method can only generate a finite number of different cones of directions by taking non-negative linear combinations of subsets of the search directions. If the cone of feasible directions at a solution of the original problem is not one of the cones (Audet and Dennis, 2003) can generate using $\mathcal{D}$, then (Audet and Dennis, 2003) must use an approximation, and so no longer solves the original problem.

The approach taken herein is to use filters, but we wish to avoid being restricted to a finite set of search directions. This means the property that non-negative integer sums of the search directions can only reach a finite number of points in any bounded region is lost. It is replaced with a sufficient descent condition (Price and Coope, 2003) in order to retain the desired convergence properties. Hence the concept of an envelope (Fletcher and Leyffer, 2002) around a filter is employed, which serves as a measure of sufficient descent for filter based methods. In order to ensure convergence the algorithm is structured in terms of frames (Coope and Price, 2000; Price and Coope, 2003a, 2003b), and incorporates elements of the feasible correction step ideas from Lucidi et al. (2002) and Yu and Li (1981) to deal with constraint curvature. Hence, like (Lucidi et al., 2002; Yu and Li, 1981), we consider only inequality constraints. For simplicity we follow (Yu and Li, 1981) and assume that the active constraints' normals are linearly independent at all points of interest on the boundary of the feasible region.

The optimization problem may be concisely expressed as

$$\min_{x \in R^n} \quad f(x) \quad \text{subject to} \quad c_i(x) \leq 0 \quad \forall\, i = 1, \ldots, q. \tag{1}$$

The feasible region Ω is the set of points in R^n for which $c(x) \leq 0$, where $c(x)$ is the vector of constraint functions $[c_1(x), \ldots, c_q(x)]^T$. All constraint functions c_i are assumed to be continuously differentiable, but their gradients are not assumed to be available. Here f takes values in $R \cup \{+\infty\}$. The analysis assumes f is locally Lipschitz at some points, and later also assumes strict differentiability at these points. Estimates of the objective function gradient are not formed. The algorithm generates a sequence of iterates $\{x^{(k)}\}_{k=1}^{\infty}$, some of which may be infeasible. Each iteration *may* require estimates of the constraint gradients at one prospective iterate $x_c^{(k)}$. A subsequence of iterates is identified, and the cluster points of this subsequence are shown to be Karush-Kuhn-Tucker (KKT) points of (1) under various conditions. The cases when the objective function is strictly differentiable (Clarke, 1990) at these cluster points, or merely locally Lipschitz are analysed. In the former case it is shown that no feasible direction exists with a negative directional derivative for f when f is strictly differentiable at a cluster point, and that this cluster point is a KKT point under standard conditions. In the latter case we show that the objective function has interesting properties in some directions. These cases are interesting because they permit the objective function to be non-smooth, infinite, or undefined at points away from these cluster points. For later convenience, when f is undefined at a point it is assigned the value infinity at that point.

1.1. Filters

Here we use the standard multiobjective notation for filters (see e.g., Audet and Dennis, 2003), which differs slightly from that of Fletcher and Leyffer (2002). A filter approach compares the merits of various points on the basis of objective function value and size of the constraint violations. The latter is measured by the constraint violation function $h(x)$, where

$$h = \|[c(x)]_+\|_2. \tag{2}$$

If two points x_1 and x_2 are related by $f(x_1) \leq f(x_2)$ and $h(x_1) \leq h(x_2)$ with at least one inequality being strict, then x_1 is clearly superior to x_2. In such cases x_1 is said to dominate x_2 and this is written concisely as $x_1 \prec x_2$. The notation $x_1 \preceq x_2$ is similar, but includes the case when neither inequality is strict. For convenience '$\prec$' and '$\preceq$' have been defined as relations between points in R^n rather that between pairs of values $(f(x), h(x))$.

A filter is a finite set of points $\mathcal{F}$ such that:

F1: no pair of points x_1 and x_2 in the filter satisfy $x_1 \prec x_2$; and
F2: $h(x) \leq h_{\max}$ for all $x \in \mathcal{F}$.

Here we have departed somewhat from Fletcher and Leyffer (2002) in that feasible points have been directly included in the filter. A point x_1 is said to be filtered if either $x_2 \preceq x_1$ for some $x_2 \in \mathcal{F}$, or if $h(x_1) > h_{\max}$. From time to time we shall speak loosely of a pair of values (f_0, h_0) being filtered. This means that there is a point x in the filter for which $f(x) \leq f_0$ and $h(x) \leq h_0$.

We define the most nearly feasible filter point $y^{(k)}$ as the point $y^{(k)} \in \mathcal{F}^{(k)}$ such that $h(y^{(k)}) \leq h(x)$ holds for each $x \in \mathcal{F}^{(k)}$. Clearly $y^{(k)}$ need not be unique, but there is no loss of generality in choosing $y^{(k)}$ arbitrarily when there is more than one most nearly feasible filter point. Note that $y^{(k)}$ may actually be feasible.

Following Fletcher and Leyffer (2002) and Fletcher et al. (2002) we construct an envelope $\mathcal{E}^{(k)}$ around the filter $\mathcal{F}^{(k)}$. This envelope is defined using two positive constants: $\epsilon^{(k)}$ and $\tau^{(k)}$. The set $\mathcal{E}^{(k)}$ contains all infeasible points x_1 such that $f(x_2) \leq f(x_1) + \epsilon^{(k)}$ and $h(x_2) < h(x_1) + \tau^{(k)}$ for some $x_2 \in \mathcal{F}^{(k)}$. The set $\mathcal{E}^{(k)}$ also contains every feasible point x_1 satisfying $f(x_1) \geq f(y) - \epsilon^{(k)}$ for some $y \in \mathcal{F}^{(k)} \cap \Omega$, with the convention that $f(y) = \infty$ if $\mathcal{F}^{(k)} \cap \Omega$ is empty. For convenience any such point which is contained in the envelope $\mathcal{E}^{(k)}$ is said to be envelope filtered, or $\mathcal{E}^{(k)}$-filtered.

2. The algorithm

The idea behind the algorithm is to generate an infinite sequence of structures called frames. Loosely speaking, the points in a frame surround a point called the frame centre. Each frame is defined by its frame centre $x_c^{(k)}$, a frame size $\delta^{(k)}$, and a positive basis $\mathcal{V}_+^{(k)}$. The latter may be defined in terms of positive spanning sets, which are finite sets of vectors whose linear combinations with non-negative coefficients span R^n. If a positive spanning set has the additional property that no proper subset of it is also a positive spanning set, then it is also a positive basis (Davis, 1954). A frame is the set of points

$$\Phi\big(x_c^{(k)}, \delta^{(k)}; \mathcal{V}_+^{(k)}\big) = \big\{x_c^{(k)} + \delta^{(k)} v : v \in \mathcal{V}_+^{(k)}\big\}. \tag{3}$$

Due to the fact that filters treat points somewhat discontinuously, it may be necessary to 'bend' some frames out of shape in order to more closely match the local constraint geometry. This is achieved by 'warping' the directions in $\mathcal{V}_+^{(k)}$, yielding a 'bent' positive basis $\mathcal{B}_+^{(k)}$. The resultant set of points $\Phi(x_c^{(k)}, \delta^{(k)}; \mathcal{B}_+^{(k)})$ is called a bent frame. The notation $\Phi(x_c, \delta; \mathcal{V}_+)$ is used in other papers explicitly for frames. Herein it is extended to include the case when $\mathcal{V}_+$ is any finite set of directions, and such $\Phi(x_c, \delta; \mathcal{V}_+)$ are called sampling sets.

The convergence properties depend on a type of frame called an envelope filtered bent frame. These are bent frames for which every point is $\mathcal{E}^{(k)}$-filtered, and the frame centre is not filtered by $\mathcal{F}^{(k)}$. Their significance is most easily illustrated in the (forbidden) limiting case when $\epsilon = \tau = 0$. In this case each bent frame point $x \in \Phi(x_c^{(k)}, \delta^{(k)}; \mathcal{B}_+^{(k)})$ must satisfy either $f(x) \geq f(x_c^{(k)})$ or $h(x) > h(x_c^{(k)})$. The frames are bent in order to eliminate the latter possibility. Hence the finite difference estimate $f(x) - f(x_c^{(k)})$ of the directional derivative $(x - x_c^{(k)})^T \nabla f(x_c^{(k)})$ is non-negative for each such x.

Guaranteed convergence is obtained by first showing that the subsequence of these envelope filtered bent frames is infinite. Then, taking limits we are able to show certain directional derivatives of f are non-negative at cluster points of these bent frame centres. The bent positive bases are aligned so that they take into account the geometry of any nearby constraints. This alignment, standard assumptions on the constraint gradients, and smoothness of f then establish that these cluster points are KKT points of the problem (1).

The algorithm is now formally stated, and then discussed in detail.

The algorithm

1. Initialize: set $k = 1$, $m = 1$, and choose the initial point $x^{(0)}$. Choose $h_{\max} > 0$ and $\gamma > 0$. Calculate $\mathcal{F}^{(1)}$.
2. Choose $\Delta^{(m)} > 0$, $E^{(m)} > 0$, and $T^{(m)} > 0$.
3. Check stopping conditions and halt if they are satisfied.
4. Choose $\delta^{(k)} \geq \Delta^{(m)}$, $\epsilon^{(k)} \geq E^{(m)}$, and $\tau^{(k)} \geq T^{(m)}$.
5. Execute any finite process which either
   (a) generates an iterate $x^{(k)}$ such that $x^{(k)}$ is not an $\mathcal{E}^{(k)}$-filtered point; or
   (b) generates an $\mathcal{E}^{(k)}$-filtered sampling set $\Phi(x_c^{(k)}, \delta^{(k)}; \mathcal{S}^{(k)})$, where $\mathcal{B}_+^{(k)} \subseteq \mathcal{S}^{(k)}$, and where no point in $\mathcal{F}^{(k)}$ dominates $x_c^{(k)}$. Set $z^{(m)} = x_c^{(k)}$, $\delta_z^{(m)} = \delta^{(k)}$, $\epsilon_z^{(m)} = \epsilon^{(k)}$, $\tau_z^{(m)} = \tau^{(k)}$, and $\mathcal{S}_z^{(m)} = \mathcal{S}^{(k)}$.
6. Update $\mathcal{F}^{(k)}$ to get $\mathcal{F}^{(k+1)}$.
7. If $x^{(k)}$ is not $\mathcal{E}^{(k)}$-filtered, increment k and go to Step 3, otherwise increment m and k, and go to Step 2.

The algorithm consists of two nested loops. Each iteration of the outer loop (Steps 2–7) generates an envelope filtered bent frame. Each iteration of the inner loop (Steps 3–7) either generates a point which is not envelope filtered, or generates an envelope filtered bent frame. Iterations of the inner loop are performed until this latter event occurs.

Each iteration of the outer loop selects positive minimum values $\Delta^{(m)}$, $E^{(m)}$, and $T^{(m)}$ for the frame size $\delta^{(k)}$, and envelope parameters $\epsilon^{(k)}$ and $\tau^{(k)}$. These minimum values ensure that each iteration of the outer loop (and hence each execution of the inner loop) is a finite process provided the sequence of function values $\{f^{(k)}\}$ is bounded below.

At iteration k of the inner loop, values for $\delta^{(k)}$, $\epsilon^{(k)}$, and $\tau^{(k)}$ are chosen which meet or exceed their preselected minimum values. The algorithm then tries to find one or more points which are not $\mathcal{E}^{(k)}$-filtered by calculating f and h at a finite number of arbitrary points. If successful, $x^{(k)}$ is chosen as a non envelope filtered point, the filter is updated, and a new inner loop iteration is begun. Otherwise a working set of constraints indexed by $W^{(k)}$ is identified and a bent frame centre $x_c^{(k)}$ is chosen. The centre $x_c^{(k)}$ must not be dominated by any point in the filter $\mathcal{F}^{(k)}$. A bent frame around $x_c^{(k)}$ is then constructed, and the values of f and h are calculated at each bent frame point. The orientation and shape of this bent frame are chosen to reflect the local geometry of constraints indexed by $W^{(k)}$. This bent frame either contains a point which is not envelope filtered, or the entire frame is envelope filtered. In the former case $x^{(k)}$ is chosen as a non envelope filtered point, the filter is updated, and a new iteration of the inner loop is begun. In the latter case this execution of the inner loop terminates, having achieved its purpose. The bent frame centre $x_c^{(k)}$ becomes both $x^{(k)}$ and the next member of the sequence $\{z^{(m)}\}$ of envelope filtered bent frame centres (or, more simply, envelope filtered centres). The sequence $\{z^{(m)}\}$ is vital because the convergence theory shows that this sequence is infinite, and the cluster points of $\{z^{(m)}\}$ are solutions of (1). The convergence theory is silent on the subsequence of non envelope filtered iterates.

Each iteration the filter is updated. This is done by first augmenting $\mathcal{F}^{(k)}$ with some or all of the points at which f and h were calculated, including $x^{(k)}$. Then, all points in the filter which are dominated by another filter point are deleted. Finally, points for which h exceeds $h_{\max}$ are also deleted from the filter. This yields $\mathcal{F}^{(k+1)}$.

Step 5 permits a finite number of arbitrary points to be examined at each iteration. This would allow the algorithm, for example, to try a quasi-Newton step, points selected by a heuristic, or even randomly selected points. In Step 5(b) various quantities are relabelled for use in the convergence analysis. The 'z' subscript indicates that these quantities correspond directly to the members of the subsequence $\{z^{(m)}\}$. These relabelled quantities are not needed to implement the algorithm.

The $z^{(m)}$ have their own separate counter m. This counter is linked to the iteration counter k via the function $k = k(m)$, which gives the iteration number in which the mth envelope filtered centre was found. Certain quantities (such as the $W^{(k)}$) need only be defined when a bent frame centre $x_c^{(k)}$ is selected. Hence the members of the sequences $\{x_c^{(k)}\}$, $\{W^{(k)}\}$, $\{\mathcal{V}_+^{(k)}\}$, and $\{\mathcal{B}_+^{(k)}\}$, amongst others, are not defined for all values of k. Any reference to these sequences or their members is understood to implicitly restrict k to the range of values for which $x_c^{(k)}$ have been selected. In particular, $x_c^{(k)}$ are defined for all iteration numbers $k = k(m)$ in which an envelope filtered centre $z^{(m)}$ is found.

When no feasible point is known, one would normally make gaining feasibility the primary aim by choosing $x_c^{(k)}$ to be the most nearly feasible point $y^{(k)}$ in $\mathcal{F}^{(k)}$ at each iteration. Once feasibility has been achieved, selecting each $x_c^{(k)}$ so that $h(x_c^{(k)})$ is either small or zero would ensure that feasibility is retained. The constant γ is used to determine which points are near the boundary of Ω. Specifically, if the maximum of the constraint values at a point x is at least $-\gamma$ and $h(x) \leq \gamma$, then x is regarded as near the boundary of Ω. In practice the constants $h_{\max}$ and γ can be adjusted a finite number of times as long as they always remain positive.

The filter approach requires that, at each point x near the boundary $\partial\Omega$ of the feasible region, a direction $w_d(x)$ along which h is decreasing can be constructed. This direction is used to warp the members of $\mathcal{V}_+^{(k)}$ to yield $\mathcal{B}_+^{(k)}$. This warping is needed because h is not smooth on $\partial\Omega$. Accordingly, at points far from $\partial\Omega$ there is no need to warp $\mathcal{V}_+^{(k)}$, and w_d need not be formed. There are two possibilities when far from $\partial\Omega$: the algorithm is either looking at points deep in the interior of Ω, or at points far away from Ω. In the first case the algorithm is acting as an unconstrained algorithm. In the second the algorithm is primarily attempting to reduce h in an area where it is smooth, with reducing f as a secondary aim. In the latter case convergence to an infeasible point is clearly possible. Guaranteeing an algorithm does not converge to such a point is a global optimization problem, and, as such, is outside the scope of this paper.

The use of the 2-norm in (2) is preferred because it is smooth everywhere except at the boundary $\partial\Omega$ of Ω. In contrast norms such as the 1-norm and ∞-norm are also non-smooth at infeasible points. Following Yu and Li (1981) we assume that the active constraint gradients are linearly independent at every point of interest on the boundary $\partial\Omega$ of Ω. This assumption guarantees the construction of a descent direction w_d for h at points near $\partial\Omega$.

The next section examines cones of feasible directions at limit points z^* of the sequence $\{z^{(m)}\}$. First the case when z^* is an interior point of Ω is discussed, partially in the context of unconstrained optimization. This allows many of the ideas central to frame-based algorithms to be explored. Cones of feasible directions at boundary points are then looked at, and a method of construction for sets of generators of these cones is given. For simplicity, much of Section 3 (only) restricts attention to C^1 objective functions. Section 4 describes how these sets of generators can be used to construct bent frames which adequately conform to the local constraint geometry. Section 5 analyses the algorithm's convergence properties. Concluding remarks are made in Section 6.

3. Tangent cones and generators

We wish to generate an infinite sequence of envelope filtered centres whose cluster point(s) have non-negative directional derivatives for f along all feasible directions. This means we must examine all feasible directions at each cluster point z^* of this sequence. For z^* in the interior of Ω the set of feasible directions is all directions, whereas for z^* on the boundary of Ω it is all directions in the tangent cone $T(z^*)$ of Ω at z^* (Clarke, 1990). The set of feasible directions is examined using a finite set of directions whose non-negative linear combinations span the set of feasible directions. In the case when z^* is an interior point any positive basis will suffice as this finite set of directions. When z^* is a boundary point this finite set of directions is partially defined by the active constraints at z^*. In either case the directions in this finite set are the limits of those in $\{\mathcal{S}_z^{(m)}\}$. Hence bent positive bases for frames near $\partial\Omega$ must be aligned with nearby constraints. This section discusses how this alignment is achieved. For convenience we focus attention on a single cluster point of the sequence of envelope filtered centres. If necessary this is done by replacing the sequence of envelope filtered centres with an appropriate infinite subsequence of itself.

3.1. Interior cluster points

The case when z^* is an interior point of Ω is looked at in this subsection. The discussion is partly in terms of an unconstrained direct search algorithm (Coope and Price, 2000), as this illustrates many of the key ideas of frame-based methods. We look at f along a finite set of directions $\{v_1, \dots, v_p\}$ which form a positive basis. Positive bases are useful because if $v_j^T \nabla f(z^*) \geq 0$ for all $j = 1, \dots, p$ and f is C^1, then $v^T \nabla f(z^*) \geq 0$ for all $v \in R^n$, and so $\nabla f(z^*)$ must be zero. In practice we estimate each $v_j^T \nabla f$ at an approximation $x_c^{(k)}$ to z^* by using the finite differences

$$f\left(x_c^{(k)} + \delta^{(k)} v_j\right) - f_c^{(k)} \quad j = 1, \dots, p,$$

where $f_c^{(k)} = f(x_c^{(k)})$ and $\delta^{(k)} > 0$.

An unconstrained algorithm (Coope and Price, 2000) would be required to generate an infinite sequence of 'quasi-minimal' frames. The notion of quasi-minimality is the unconstrained equivalent of envelope filtered. A frame is quasi-minimal if and only if

$$f\left(x_c^{(k)} + \delta^{(k)} v\right) + \epsilon^{(k)} \geq f_c^{(k)} \quad \forall\, v \in \mathcal{V}_+^{(k)} \tag{4}$$

where $\epsilon^{(k)}$ is the positive constant used to define the envelope $\mathcal{E}^{(k)}$, and $\epsilon^{(k)}$ is chosen at the start of the kth iteration. The sequences $\{\delta^{(k)}\}$ and $\{\epsilon^{(k)}\}$ are both required to converge to zero as $k \to \infty$, with $\{\epsilon^{(k)}\}$ converging faster than $\{\delta^{(k)}\}$ (i.e. $\epsilon^{(k)}/\delta^{(k)} \to 0$ as $k \to \infty$).

The importance of a sequence of quasi-minimal frames can most easily be seen by assuming that all $\mathcal{V}_+^{(k)}$ are identical. Equation (4) then yields

$$\frac{f\left(x_c^{(k)} + \delta^{(k)} v\right) - f_c^{(k)}}{\delta^{(k)}} + \frac{\epsilon^{(k)}}{\delta^{(k)}} \geq 0 \quad \forall v \in \mathcal{V}_+^{(k)}. \tag{5}$$

In the limit $k \to \infty$ this gives $v^T \nabla f(z^*) \geq 0$ for all v in $\mathcal{V}_+^{(k)}$. Thus, if f is C^1 then z^* must be a stationary point of f.

In practice using the same positive basis for all $\mathcal{V}_+^{(k)}$ leads to an unacceptable loss of flexibility. Instead, the sequence of positive bases $\{\mathcal{V}_+^{(k)}\}$ is forbidden to collapse in the limit $k \to \infty$. That is to say, the limits of this sequence of positive bases must also be positive bases. More precisely, each $\mathcal{V}_+^{(k)}$ is regarded as an ordered set, and limits of the sequence $\{\mathcal{V}_+^{(k)}\}$ are defined as follows.

Definition 3.1. A set of vectors $\mathcal{V}_+^* = \{v_1^*, \dots, v_p^*\}$ is a limit of the sequence of sets of vectors $\{\mathcal{V}_+^{(k)}\}_{k=1}^{\infty}$ if and only if there exists an infinite subset $\mathcal{K}$ of positive integers such that $\mathcal{V}_+^{(k)}$ has cardinality p for all $k \in \mathcal{K}$; and

$$\lim_{k \to \infty} v_j^{(k)} = v_j^* \quad \forall j = 1, \dots, p$$

where the limit is over values of k in $\mathcal{K}$.

The existence of limits for the sequence $\{\mathcal{V}_+^{(k)}\}$ is guaranteed by imposing an upper bound K on the length of every member of every $\mathcal{V}_+^{(k)}$. The cardinality of each positive basis is at least $n + 1$ and at most $2n$ (Davis, 1954). Hence infinite subsequences of positive bases of the same cardinality must exist. The upper bound K ensures these subsequences have limits.

An appropriate choice (Coope and Price, 2000; Price and Coope, 2003) of the method by which the $\mathcal{V}_+^{(k)}$ are generated ensures that the limits of $\{\mathcal{V}_+^{(k)}\}$ are positive bases for R^n. By considering subsequences if necessary, if $\mathcal{V}_+^{(k)} \to \mathcal{V}_+^*$ and $x_c^{(k)} \to z^*$ as $k \to \infty$ then Eq. (5) ensures that the directional derivatives of a C^1 function f at z^* along each member of $\mathcal{V}_+^*$ are non-negative. Since $\mathcal{V}_+^*$ is a positive basis, the only possibility is that $\nabla f(z^*) = 0$.

3.2. *Boundary and infeasible cluster points*

We take an approach similar to that in the previous subsection when z^* lies on the boundary of Ω. In this case, the tangent cone $T(z^*)$ is no longer equal to R^n. A first order constraint qualification is required if the tangent cone is to be written in terms of the constraint gradients. Herein the following assumption is used.

Assumption 3.1. At each point x on the boundary of Ω, the gradients of the constraints active at x are linearly independent.

Clearly this can be relaxed at points far away from all points considered by the algorithm. Using this first order constraint qualification, the tangent cone at z^* may be expressed as

$$T(z^*) = \{v : v^T \nabla c_i(z^*) \le 0 \quad \forall i \text{ such that } c_i(z^*) = 0\}.$$

This definition can not be used directly for several reasons. First z^* is unknown, only estimates of it are available. Second, the active set at z^* is not always obvious. Last, only estimates $a_i^{(k)}$ of the gradients $\nabla c_i(x_c^{(k)})$ may be available. Hence, whether or not exact constraint gradient information is available, the tangent cones at boundary points of interest must be estimated. We construct an estimate $C^{(k)}(W)$ of the tangent cone $T(z^*)$ at each iterate as follows:

$$C^{(k)}(W) = \left\{v : v^T a_i^{(k)} \le 0 \quad \forall i \in W\right\}. \tag{6}$$

The construction used for $C^{(k)}(W)$ has the property that it also implicitly constructs each cone $C^{(k)}(W_1)$, where $W_1 \subseteq W$. This means that one does not have to *exactly* determine the set of indices W^* of the active constraints at z^*; a superset of W^* will do provided the estimated constraint normals of this superset are linearly independent.

We require that the estimated tangent cone $C^{(k)}(W^*)$ converge to $T(z^*)$ in the limit $k \to \infty$. This requires that the errors in the estimated constraint gradients go to zero as $k \to \infty$. Accordingly, the maximum error L_k between the actual and estimated gradients at $x_c^{(k)}$ is defined as

$$L_k = \max\left\{\left\|a_i^{(k)} - \nabla c_i\left(x_c^{(k)}\right)\right\| : i \in W^{(k)}\right\},$$

and the following assumption is made.

Assumption 3.2. The errors in the gradient estimates go to zero in the limit $k \to \infty$, that is to say $\lim_{k\to\infty} L_k = 0$.

When $x_c^{(k)}$ is near the boundary of Ω, the set $W^{(k)}$ must index all constraints which are maximal at $x_c^{(k)}$, or nearly so. For a preselected constant $\gamma > 0$, if $h(x) \leq \gamma$ then

$$\forall\, i = 1, \ldots, q, \quad c_i\left(x_c^{(k)}\right) \geq -\gamma \quad \Rightarrow \quad i \in W^{(k)}. \tag{7}$$

If $h > \gamma$ then the empty set can always be used for $W^{(k)}$. The reasoning behind (7) is as follows: if h is large (i.e. greater than γ) then $h(x)$ is smooth near $x_c^{(k)}$ and there are no kinks along which one *must* align the positive bases. If h is not large, then the positive bases are aligned with all constraints which are within γ of being active.

Assumption 3.1 allows γ to be defined implicitly. To do this first list the constraints in increasing order of $|c_i(x_c^{(k)})|$, and then add constraints to $W^{(k)}$ in that order, starting with $W^{(k)}$ as the empty set. One stops adding constraints when adding one more constraint to $W^{(k)}$ would make the set $\{a_i^{(k)} : i \in W^{(k)}\}$ linearly dependent. In practice, one would choose a value for γ, and use the above construction only when $h \leq \gamma$. In this case one would only need to apply the above construction to the subset of constraints for which $|c_i(x_c^{(k)})| \leq \gamma$. Constraints not satisfying (7) *may* also be included in $W^{(k)}$, provided the corresponding set of constraint normals $A^{(k)} = \{a_i^{(k)} : i \in W^{(k)}\}$ remains linearly independent.

The C^1 nature of the constraints means that if some subsequence of $\{x_c^{(k)}\}$ converges to z^*, then, for k large, each corresponding $W^{(k)}$ contains the active set at z^*. This follows from Assumption 3.1 and from (7) on noting $c_i(z^*) = 0$ for all i in W^*.

The tangent cone at any feasible point near $x_c^{(k)}$ is defined by some subset of the constraints in the working set $W^{(k)}$. Each cone of the form $C^{(k)}(W)$ is a polyhedral cone and hence (Theorem 4.18 of van Tiel (1984)) can be written as the non-negative combinations of a finite number of vectors:

$$\exists v_1, \ldots, v_p \quad \text{such that} \quad C^{(k)}(W) = \left\{\sum_{j=1}^{p} \eta_j v_j : \eta_j \geq 0 \quad \forall j\right\}. \tag{8}$$

The vectors $v_1, \ldots, v_p$ are often referred to as a *set of generators* of the cone $C^{(k)}(W)$.

Sets of generators are extremely useful. If f is C^1 and if $\nabla f^T v_j \geq 0$ for each generator v_j, then $\nabla f^T v \geq 0$ for all v in the cone $C^{(k)}(W)$. This follows from the non-negativity of each η_j in (8) and the linearity of $v^T \nabla f$ in v. Thus we only need to consider sets of generators, not the cones themselves.

Sets of generators are needed for every cone $C^{(k)}(W)$ for which $W \subseteq W^{(k)}$. Their construction is the subject of the following subsection. For judicious choices of these sets of generators it turns out that they have many members in common, and their union $\mathcal{V}_+^{(k)}$ can be a fairly small positive spanning set for R^n. When the working set $A^{(k)} = \{a_i^{(k)} : i \in W^{(k)}\}$ of estimated constraint normals is linearly independent, this union $\mathcal{V}_+^{(k)}$ is a positive basis for R^n, and hence (Davis, 1954) has at most $2n$ members. Any positive basis (or positive spanning set) which contains a set of generators for each cone $C^{(k)}(W)$, $W \subseteq W^{(k)}$, is said to be *aligned* with the working set $W^{(k)}$ at iteration k.

3.3. *Constructing sets of generators*

An aligned positive basis for the working set $W^{(k)}$ at $x_c^{(k)}$ is constructed in two parts: one each for the subspace containing the members of $A^{(k)} = \{a_i^{(k)} : i \in W^{(k)}\}$ and for the subspace orthogonal to the members of $A^{(k)}$. For notational simplicity assume that $W^{(k)} = \{1, \ldots, \ell\}$. Let $S^{(k)} = [s_1^{(k)} \ldots s_n^{(k)}]$ be any invertible matrix satisfying $(S^{(k)})^T a_i^{(k)} = e_i$, where $s_1^{(k)}, \ldots, s_n^{(k)}$ are the columns of $S^{(k)}$, and e_i is the ith column of the identity matrix. Let $\mathcal{U}_+^{(k)}$ be a positive basis for the subspace spanned by $e_{\ell+1}, \ldots, e_n$. Clearly $\{\pm e_1, \ldots, \pm e_\ell\} \cup \mathcal{U}_+^{(k)}$ is a positive basis for R^n. Pre-multiplying by $S^{(k)}$ gives

$$\mathcal{V}_+^{(k)} = \left\{\pm s_i^{(k)} : i = 1, \ldots, \ell\right\} \cup \left\{S^{(k)} u : u \in \mathcal{U}_+^{(k)}\right\} \tag{9}$$

which is also a positive basis for R^n because $S^{(k)}$ is invertible. The next theorem shows that $\mathcal{V}_+^{(k)}$ is aligned with the working set $W^{(k)}$ at $x_c^{(k)}$.

Theorem 3.1. *For each $W \subseteq W^{(k)}$, the set of generators*

$$\mathcal{G}^{(k)}(W) = \left\{-s_i^{(k)} : i \in W\right\} \cup \left\{\pm s_i^{(k)} : i \in W^{(k)} - W\right\} \cup \left\{S^{(k)} u : u \in \mathcal{U}_+^{(k)}\right\} \tag{10}$$

for the cone $C^{(k)}(W)$ is contained in the positive basis $\mathcal{V}_+^{(k)}$. Hence $\mathcal{V}_+^{(k)}$ is aligned with the working set $W^{(k)}$ at $x_c^{(k)}$.

Proof: Clearly $\mathcal{G}^{(k)}(W) \subseteq \mathcal{V}_+^{(k)}$ for all $W \subseteq W^{(k)}$. Thus we need to show that each $\mathcal{G}^{(k)}(W)$ is a set of generators for the cone $C^{(k)}(W)$, $W \subseteq W^{(k)}$. Consider an arbitrary $W \subseteq W^{(k)}$. Without loss of generality let $W^{(k)} = \{1, \ldots, \ell\}$ and re-order the constraints in $W^{(k)}$ so that $W = \{1, \ldots, r\}$. To show that $\mathcal{G}^{(k)}(W)$ is a set of generators for the cone $C^{(k)}(W)$, we first note that

$$v \in C^{(k)}(W) \Leftrightarrow v^T a_i^{(k)} \leq 0 \quad \forall\, i \in W.$$

This is clearly true for all members of $\mathcal{G}^{(k)}(W)$, and so it remains to show that the members of $\mathcal{G}^{(k)}(W)$ positively span $C^{(k)}(W)$. Let v be an arbitrary member of $C^{(k)}(W)$. Then

$$v^T a_i^{(k)} = \theta_i \leq 0 \quad \forall\, i \in W.$$

Using the fact that $(s_j^{(k)})^T a_i^{(k)} = 1$ when $i = j$ and zero otherwise, it follows that

$$v - \sum_{j \in W} \theta_j s_j^{(k)}$$

must lie in the null space of $a_i^{(k)}$, $i \in W$. However the second and third parts of $\mathcal{G}^{(k)}(W)$ positively span this null space, so v can be expressed as a non-negative linear combination of the members of $\mathcal{G}^{(k)}(W)$. Since W was an arbitrary subset of $W^{(k)}$, the positive basis $\mathcal{V}_+^{(k)}$ is aligned with $W^{(k)}$ at iteration k. □

We need to make sure that the limits of the sequence of positive bases $\{\mathcal{V}_+^{(k)}\}$ are also positive bases. This is done in two parts. For the first part, an upper bound K is imposed on the length of every $u \in \mathcal{U}_+^{(k)}$ for all values of k, where the positive constant K is independent of k. Consider an arbitrary increasing sequence of positive integers $\mathcal{K}$ for which $W^{(k)}$ exists, and is the same for all $k \in \mathcal{K}$. Then the positive bases $\mathcal{U}_+^{(k)}$, $k \in \mathcal{K}$ all positively span the same subspace of R^n. Limits for this subsequence of $\mathcal{U}_+^{(k)}$ of positive bases are defined using Definition 3.1. The set of all such limits for all acceptable choices of $\mathcal{K}$ is the set of limits for the full sequence $\{\mathcal{U}_+^{(k)}\}_{k=1}^{\infty}$. The following assumption ensures the unconstrained parts of $\mathcal{V}_+$ do not asymptotically collapse.

Assumption 3.3. All limits of the sequence of ordered positive bases $\{\mathcal{U}_+^{(k)}\}$ are ordered positive bases.

This assumption can be enforced using the same techniques described in Coope and Price (2000) and Price and Coope (2003) in an unconstrained setting.

For the second part, select an arbitrary increasing sequence of positive integers $\mathcal{K}$ for which $W^{(k)}$ is the same for all $k \in \mathcal{K}$, and such that the subsequences $\{\mathcal{U}_+^{(k)}\}_{k\in\mathcal{K}}$ and $\{x_c^{(k)}\}_{k\in\mathcal{K}}$ have limits $\mathcal{U}_+^*$ and z^* respectively. Then $a_i^{(k)} \to \nabla c_i(z^*)$ for all i in $W^{(k)}$ in the limit $k \to \infty$, $k \in \mathcal{K}$, by Assumption 3.2. Define limits of the subsequence of matrices $\{S^{(k)}\}_{k\in\mathcal{K}}$ in the usual way using the matrix 2-norm. The set of all such limits for all acceptable choices of $\mathcal{K}$ is the set of limits of the full sequence $\{S^{(k)}\}$.

Assumption 3.4. All limits of the sequence $\{S^{(k)}\}$ are non-singular, and share a common upper bound on their 2-norm.

The linear independence of the set of active constraint normals at z^* ensures that this assumption can be enforced.

Assumptions 3.3 and 3.4 ensure each limit $\mathcal{V}_+^*$ of $\{\mathcal{V}_+^{(k)}\}$ is a positive basis. The construction for $\mathcal{V}_+^{(k)}$ in (9) is continuous with respect to $S^{(k)}$ and $\mathcal{U}_+^{(k)}$, and hence is also valid for $\mathcal{V}_+^*$, as is Theorem 3.1. Specifically, choose $\mathcal{K}$ so that the subsequences $\{\mathcal{U}_+^{(k)}\}_{k\in\mathcal{K}}$ and $\{S^{(k)}\}_{k\in\mathcal{K}}$ have limits $\mathcal{U}_+^*$ and S^*, and all $W^{(k)}$ are equal for $k \in \mathcal{K}$. Then (9) gives the limit $\mathcal{V}_+^*$ of the corresponding subsequence $\{\mathcal{V}_+^{(k)}\}_{k\in\mathcal{K}}$ when $\mathcal{U}_+^{(k)}$ and $S^{(k)}$ are replaced with $\mathcal{U}_+^*$ and S^* respectively. Theorem 3.1 shows $\mathcal{V}_+^*$ is aligned with the superset $W^{(k)}$ of W^* at z^*, where $k \in \mathcal{K}$.

In addition to conditions on the sequence of ordered positive bases, the following assumption is needed to establish convergence.

Assumption 3.5. The following conditions hold:

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

It is shown in Section 5 that part (b) of this assumption guarantees that the subsequence $\{z^{(m)}\}$ is infinite. Part (a) then implies that this subsequence also 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 algorithm.

4. Sampling sets and bent frames

At each iteration f and h are calculated at members of a sampling set until either a point which is not envelope filtered is found, or the sampling set is exhausted. The members of the sampling set are expressed in the form $x_c^{(k)} + \delta^{(k)}v$, where v ranges over a finite set $\mathcal{S}^{(k)}$ which contains the bent positive basis $\mathcal{B}_+^{(k)}$. At first sight one would expect $\mathcal{S}^{(k)}$ would include all directions in the positive basis $\mathcal{V}_+^{(k)}$, but this leads to theoretical difficulties. The crux of these difficulties is that each envelope filtered point $x_c^{(k)} + \delta^{(k)}v$ provides information which indicates that *either* f *or* h is increasing along v. The problem is most easily seen when there is only one constraint which is strictly convex, and $a_1^{(k)}$ is exact. For directions tangential to the constraint h is increasing, and hence the fact that a sampling point is envelope filtered conveys no information about f. Unfortunately information about f is what we seek—not information about h.

For example let $c_1(x) = x_1^2 + x_2^2 - 9$ in R^2. At the point $x^{(k)} = (3, 0)$ we have $\mathcal{V}_+^{(k)} = \{\pm 2e_1, \pm 2e_2\}$, with $\{\pm 2e_2\}$ being the '$\mathcal{U}_+$' part of the positive basis. The three vectors $\{-2e_1, \pm 2e_2\}$ generate the cone of feasible directions at (3, 0). Clearly $h(x^{(k)} \pm 2\delta^{(k)}e_2) > h(x^{(k)})$, for any positive $\delta^{(k)}$, and hence if the sampling set is envelope filtered it provides information on f only in the single direction $-2e_1$, for all sufficiently small positive $\delta^{(k)}$. This is clearly not sufficient to determine whether or not a feasible direction exists at (3, 0) along which f has a negative slope.

This problem is circumvented by generating a direction $w_d^{(k)}$ which is a descent direction for each constraint in $W^{(k)}$ at $x_c^{(k)}$. This direction is scaled to give a non-ascent step $w_s^{(k)}$ for h at $x_c^{(k)}$. Members of $\mathcal{V}_+^{(k)}$ are bent towards $w_s^{(k)}$, yielding $\mathcal{B}_+^{(k)}$.

In the two dimensional example above, if $w_s^{(k)} = -e_1$ and $\delta^{(k)} = 1$, then one choice for $\mathcal{B}_+^{(k)}$ is $\{-3e_1, 2e_2 - e_1, -2e_2 - e_1, e_1\}$. An envelope filtered sample set based on this $\mathcal{B}_+^{(k)}$ would provide information about f along the first three directions in $\mathcal{B}_+^{(k)}$. Whilst these three directions still do not generate the tangent cone at $(3, 0)$, they do generate a cone much bigger than that generated by $\{-2e_1\}$ on its own. It is shown later that, as $k \to \infty$, the amount of 'bending' required for directions tangential or interior to $\partial\Omega$ goes to zero. It is then shown that, in the limit, these directions positively span the relevant tangent cones.

Assumptions 3.1 and 3.2 guarantee that the $a_i^{(k)}$ are linearly independent for sufficiently large k. If $W^{(k)}$ is non-empty then $w_d^{(k)}$ is chosen as

$$w_d^{(k)} = -\sum_{i \in W^{(k)}} s_i^{(k)}, \tag{11}$$

otherwise $w_d^{(k)} = 0$ is used. Assumption 3.4 implies that the sequence $\{w_d^{(k)}\}$ is bounded, and (11) gives $(a_i^T w_d)^{(k)} = -1$ for all $i \in W^{(k)}$. When $w_d^{(k)}$ is non-zero, it is scaled by a factor $\lambda^{(k)}$ to yield a non-ascent step $w_s^{(k)}$ for h at $x_c^{(k)}$. The factor $\lambda^{(k)}$ is chosen using a back-tracking search. For a preselected $\rho \in (0, 1)$, members of the sequence $\lambda = 1, \rho, \rho^2, \dots$ are tried in that order, and the first one satisfying

$$h\big(x_c^{(k)} + \lambda\delta^{(k)} w_d^{(k)}\big) \le h\big(x_c^{(k)}\big) \tag{12}$$

is accepted as $\lambda^{(k)}$, giving $w_s^{(k)} = \lambda^{(k)} w_d^{(k)}$. If no acceptable λ value is found before λ reaches machine precision, then more accurate constraint gradient estimates are formed, and the process is repeated. If $w_d^{(k)} = 0$ then $w_s^{(k)} = 0$ is used.

Theorem 4.1.

$$\liminf_{k\to\infty} \big\|w_s^{(k)}\big\| > 0$$

where k ranges over the positive integers for which $W^{(k)}$ is defined and non-empty.

Proof: Assume that there exists an increasing sequence of positive integers $\mathcal{K}$ such that

$$\lim_{k \in \mathcal{K}} \big\|w_s^{(k)}\big\| = 0.$$

By replacing $\mathcal{K}$ with a subsequence of itself if necessary, assume that all $W^{(k)}$ are the same for $k \in \mathcal{K}$. Assumption 3.4 and Eq. (11) imply that the sequence of $\|w_d^{(k)}\|$ values is bounded away from zero. Therefore $\lambda^{(k)} \to 0$ as $k \to \infty$ over $\mathcal{K}$. Hence, for all sufficiently large $k \in \mathcal{K}$

$$h\big(x_c^{(k)} + \delta^{(k)}\lambda^{(k)} w_d^{(k)}/\rho\big) > h\big(x_c^{(k)}\big),$$

and the sequence of pairs $\{(w_d^{(k)}, x_c^{(k)})\}$ has a limit point (w_d^*, x_c^*) for which $(w_d^*)^T \nabla c_i(x_c^*) \geq 0$ must hold for some $i \in W^{(k)}$.

On the other hand $(w_d^T a_i)^{(k)} = -1$ for all $i \in W^{(k)}$ and for all k. Now $w_d^{(k)} \to w_d^*$ and $a_i^{(k)} \to \nabla c_i(x_c^*)$ for all i, so $\nabla c_i(x_c^*)^T w_d^* = -1$ for all $i \in W^{(k)}$—a contradiction. □

The members of $\mathcal{V}_+^{(k)}$ are 'bent' towards $w_s^{(k)}$ so that the steps along these bent directions are non-increasing steps for h. Each $v_j^{(k)}$ is bent to give the direction $b_j^{(k)}$. Collectively the $b_j^{(k)}$ form the bent positive basis $\mathcal{B}_+^{(k)}$. Each $b_j^{(k)}$ is a function of a bending parameter α as follows:

$$b_j^{(k)}(\alpha) = (1-\alpha)v_j^{(k)} + \alpha w_s^{(k)} \quad \text{where } \alpha \in [0, 1]. \tag{13}$$

The quantity $\alpha_j^{(k)}$ denotes the value of α used by the algorithm to bend $v_j^{(k)}$. The choice $\alpha_j^{(k)} = 0$ corresponds to no bending, and $\alpha_j^{(k)} = 1$ to replacing $v_j^{(k)}$ with the non-ascent step $w_s^{(k)}$ for h at $x_c^{(k)}$. The value $\alpha_j^{(k)}$ is chosen so that

$$h\big(x_c^{(k)} + \delta^{(k)} b_j^{(k)}\big(\alpha_j^{(k)}\big)\big) \leq \max\big\{h\big(x_c^{(k)}\big) - \tau^{(k)}, h\big(x_c^{(k)} + \delta^{(k)} w_s^{(k)}\big)\big\}. \tag{14}$$

For convenience, the notation $b_j^{(k)}$ is used as a shorthand for $b_j^{(k)}(\alpha_j^{(k)})$. Clearly (14) always holds for $\alpha = 1$. When $x_c^{(k)}$ is infeasible (14) can be stronger than non-ascent. The following is *an example* of a sub-algorithm that could be used to implement Step 5(b) of the main algorithm.

An example sub-algorithm for Step 5(b)

(i) Select $W^{(k)}$ and calculate $a_i^{(k)}$, $i \in W^{(k)}$.
(ii) Calculate $s_i^{(k)}$ and form $\mathcal{V}_+^{(k)}$ using Eq. (9).
(iii) Calculate $w_s^{(k)}$ via (11) and (12).
(iv) Form $\mathcal{B}_+^{(k)}$. For $j = 1, \ldots, |\mathcal{V}_+^{(k)}|$ choose $\alpha_j^{(k)}$ as the least element in, for example, $\{0, 2^{-k}, 2.2^{-k}, 3.2^{-k}, \ldots, 1\}$ which satisfies (14), where $b_j^{(k)}$ is given by (13).
(v) Select $x_c^{(k)}$ as $y^{(k)}$ and $\mathcal{S}^{(k)}$ as $\mathcal{B}_+^{(k)}$. Calculate f and h at points in the sampling set $\Phi(x_c^{(k)}, \delta^{(k)}; \mathcal{S}^{(k)})$ until either a non $\mathcal{E}^{(k)}$ filtered point is found or the sampling set is exhausted.

The purpose behind how each α is chosen in Step (iv) is to, asymptotically, select the smallest value acceptable to (14). The minimum acceptable value $\beta_j^{(k)}$ of the bending coefficient $\alpha_j^{(k)}$ for $v_j^{(k)}$ at $x_c^{(k)}$ is

$$\begin{aligned}\beta_j^{(k)} = \min \big\{\beta \in [0,1] : h\big(x_c^{(k)} + \delta^{(k)} b_j^{(k)}(\beta)\big) \\ \leq \max\big[h\big(x_c^{(k)}\big) - \tau^{(k)}, h\big(x_c^{(k)} + \delta^{(k)} w_s^{(k)}\big)\big]\big\}.\end{aligned}$$

Each $\beta_j^{(k)}$ is the smallest value in [0, 1] for which $b_j^{(k)}(\beta_j^{(k)})$ satisfies (14) at $x_c^{(k)}$. The following assumption requires that the actual amount of bending asymptotically approach the minimal bending required by (14).

Assumption 4.1. The method used to choose each $\alpha_j^{(k)}$ ensures that

$$\lim_{k\to\infty} \max\left\{\left|\beta_j^{(k)} - \alpha_j^{(k)}\right| : \forall j = 1, \ldots, \left|\mathcal{V}_+^{(k)}\right|\right\} = 0.$$

The next two theorems show that these minimum acceptable values tend to zero for limiting directions which are tangential or descent directions for all active constraints. The first establishes a useful intermediate result.

Theorem 4.2. *For $c \in R^q$ and $\tau > 0$*

$$\|[c-\tau]_+\|_2 \leq [\|[c]_+\|_2 - \tau]_+. \tag{15}$$

Proof: The result is obvious if $[c-\tau]_+ = 0$. Otherwise, let u be the unit vector parallel to $[c-\tau]_+$. Then $[c-\tau]_+ + \tau u \leq [c]_+$ because each element u_i of u satisfies $u_i \leq 1$, and $u_i > 0$ if and only if $c_i > \tau$. Also $\|[c-\tau]_+ + \tau u\|_2 \leq \|[c]_+\|_2$ because both vectors are non negative. Finally, $\|[c-\tau]_+ + \tau u\|_2 = \|[c-\tau]_+\|_2 + \tau$ because u and $[c-\tau]_+$ are parallel, which gives the result. $\square$

Theorem 4.3. *Let $\mathcal{K}$ be an increasing sequence of positive integers such that the sequences $\{x_c^{(k)}\}_{k\in\mathcal{K}}$ and $\{\mathcal{B}_+^{(k)}\}_{k\in\mathcal{K}}$ have limits x_c^* and $\mathcal{B}_+^*$ respectively. Here x_c^* is feasible, and the limits of $\{\mathcal{B}_+^{(k)}\}_{k\in\mathcal{K}}$ are given by Definition* 3.1. *Then $\mathcal{B}_+^*$ contains a set of generators for the tangent cone $T(x_c^*)$.*

Proof: Let W^* index the set of constraints which are active at x_c^*. Assume k is large enough so that no point of the form $x_c^{(k)} + \delta^{(k)} b_j^{(k)}$ violates any constraint not indexed by W^*. Clearly $W^* \subseteq W^{(k)}$ for all sufficiently large $k \in \mathcal{K}$. If $W^{(k)}$ is empty, then $T(x_c^*) = R^n$, and it is positively spanned by the positive basis $\mathcal{V}_+^{(k)} = \mathcal{B}_+^{(k)}$, as required. Otherwise, the set of generators in $\mathcal{V}_+^{(k)}$ for the tangent cone $C^{(k)}(W^*)$ is

$$\mathcal{G}^{(k)}(W^*) = \left\{-s_i^{(k)} : i \in W^*\right\} \cup \left\{\pm s_i^{(k)} : i \in W^{(k)} - W^*\right\} \cup \left\{S^{(k)}u : u \in \mathcal{U}_+^{(k)}\right\}.$$

All members v_j of $\mathcal{G}^{(k)}(W^*)$ satisfy $v_j^T a_i^{(k)} \leq 0$ for all $i \in W^*$ by definition of a set of generators. Now the change $\Delta c_i(b_j^{(k)}(\alpha))$ in constraint i as a result of taking a step $\delta^{(k)} b_j^{(k)}(\alpha)$ from $x_c^{(k)}$ is

$$\begin{aligned}\Delta c_i\big(b_j^{(k)}(\alpha)\big) &= c_i\big(x_c^{(k)} + \delta^{(k)}\big((1-\alpha)v_j^{(k)} + \alpha w_s^{(k)}\big)\big) - c_i\big(x_c^{(k)}\big) \\ &= \delta^{(k)}\big(a_i^{(k)}\big)^T\big((1-\alpha)v_j^{(k)} + \alpha w_s^{(k)}\big) + o\big(\delta^{(k)}\big) + L_k\, O\big(\delta^{(k)}\big).\end{aligned} \tag{16}$$

Here $o(\delta)$ and $O(\delta)$ denote quantities that respectively go to zero faster than δ, and at least as fast as δ, in the limit $k \to \infty$. Noting that $v_j^T a_i^{(k)}$ is always non-positive for all v_j in $\mathcal{G}^{(k)}(W^*)$,

$$\frac{1}{\delta^{(k)}}\left[\Delta c_i\left(b_j^{(k)}(\alpha)\right) + \tau^{(k)}\right] \le \alpha\left(a_i^{(k)}\right)^T w_s^{(k)} + o(1) + L_k\, O(1) + \frac{\tau^{(k)}}{\delta^{(k)}},$$

where Assumption 3.5(d) means that the right hand τ/δ term can be incorporated into the $o(1)$ term. This right hand side is non-positive for values of α satisfying

$$\alpha \ge \left[(o(1) + L_k O(1)) \big/ \left(a_i^{(k)}\right)^T w_s^{(k)}\right] \tag{17}$$

for each $i \in W^*$. So, for these α values,

$$c_i\left(x_c^{(k)} + \delta^{(k)} b_j^{(k)}(\alpha)\right) \le c_i\left(x_c^{(k)}\right) - \tau^{(k)}.$$

Now $c_i \le 0$ at all points of interest for all $i \notin W^*$. Therefore if the same α satisfies (17) for all i in W^*, then Theorem 4.2 implies

$$h\left(x_c^{(k)} + \delta^{(k)} b_j^{(k)}(\alpha)\right) \le \left[h\left(x_c^{(k)}\right) - \tau^{(k)}\right]_+ \forall j : v_j^{(k)} \in \mathcal{G}^{(k)}(W^*). \tag{18}$$

The right hand side of (18) is a lower bound on the right hand side of (14). Hence

$$\beta_j^{(k)} \le \max_{i \in W^*}\left[(o(1) + L_k O(1)) \big/ \left(a_i^{(k)}\right)^T w_s^{(k)}\right]_+ \quad \forall j : v_j^{(k)} \in \mathcal{G}_j^{(k)}(W^*).$$

Now $(a_i^T w_s)^{(k)} = -\|w_s^{(k)}\| / \|w_d^{(k)}\|$ for all $i \in W^*$ and $k \in \mathcal{K}$, so Theorem 4.1 implies the denominator in (17) is bounded away from zero. Hence $\beta_j^{(k)} \to 0$ and also $\alpha_j^{(k)} \to 0$ as $k \to \infty$ for all j such that $v_j^{(k)} \in \mathcal{G}^{(k)}(W^*)$, by Assumptions 3.2 and 4.1. Therefore $b_j^{(k)} - v_j^{(k)} \to 0$ as $k \to \infty$ for all of these values of j. However the limiting values of $\mathcal{G}^{(k)}(W^*)$ span the tangent cone $T(x_c^*)$ by Theorem 3.1, and Assumptions 3.3 and 3.4, which yields the result. □

Theorem 4.3 also holds at each infeasible limit x_c^* because the tangent cone $T(x_c^*)$ is empty at any infeasible x_c^*.

Asymptotically (14) either reduces h by at least τ, or to zero, as is shown next.

Theorem 4.4. *Let $\{x_c^{(k)}\}$ converge to $x^* \in \Omega$, if necessary by replacing $\{x_c^{(k)}\}$ with a subsequence of itself. Then for all k sufficiently large*

$$x \in \Phi\left(x_c^{(k)}, \delta^{(k)}; \mathcal{B}_+^{(k)}\right) \Rightarrow h(x) \le \left[h\left(x_c^{(k)}\right) - \tau^{(k)}\right]_+. \tag{19}$$

Proof: There are two cases. First let x^* be an interior point of Ω. Assumption 3.5(c) implies that all frame points are feasible for k large, which yields the result.

Second, let $x^* \in \partial\Omega$. We show that $h(x_c^{(k)} + \delta^{(k)} w_s^{(k)}) \leq [h(x_c^{(k)}) - \tau^{(k)}]_+$, by showing $w_d^{(k)} = w_s^{(k)}$ for k large. Together with (14), this yields the required result. By considering a subsequence of $\{x_c^{(k)}\}$ if necessary, let $W^{(k)}$ be independent of k and let $\mathcal{V}_+^{(k)} \to \mathcal{V}_+^*$. Let k be large enough so that $c_i < 0$, $i \notin W^{(k)}$ at all points of interest. Now, for all $i \in W^{(k)}$,

$$\begin{aligned} c_i\left(x_c^{(k)} + \delta^{(k)} w_d^{(k)}\right) &= c_i\left(x_c^{(k)}\right) + \delta^{(k)}\left(w_d^{(k)}\right)^T a_i^{(k)} + L_k O\left(\delta^{(k)}\right) + o\left(\delta^{(k)}\right) \\ &= c_i\left(x_c^{(k)}\right) - \delta^{(k)} + L_k O\left(\delta^{(k)}\right) + o\left(\delta^{(k)}\right) \end{aligned}$$

after noting that $(w_d^T a_i)^{(k)} = -1$ for all i in $W^{(k)}$. Then for all i in $W^{(k)}$,

$$c_i\left(x_c^{(k)} + \delta^{(k)} w_d^{(k)}\right) = c_i\left(x_c^{(k)}\right) - \tau^{(k)} - \delta^{(k)} + L_k O\left(\delta^{(k)}\right) + o\left(\delta^{(k)}\right) \leq c_i\left(x_c^{(k)}\right) - \tau^{(k)}$$

for sufficiently large k by Assumptions 3.5(d) and 3.5(c). Theorem 4.2 implies $w_d^{(k)}$ is a non increasing step for h when k is large, and hence $w_d^{(k)} = w_s^{(k)}$. Theorem 4.2 and (14) then yield the required result. □

5. Convergence results

The convergence properties of the sequence of envelope filtered centres $\{z^{(m)}\}$ are analysed in this section using the non-smooth calculus of Clarke (1990). The Clarke generalized derivative is defined for locally Lipschitz functions as

$$f^\circ(x; v) = \limsup_{t \downarrow 0\, y \to x} \frac{f(y + tv) - f(y)}{t}.$$

If f is not locally Lipschitz at x then this limit superior may be infinite. Provided f is locally Lipschitz at x it can be shown (Clarke, 1990) that $f^\circ(x; v)$ is subadditive and positively homogeneous in v. Moreover, if M is a Lipschitz constant for f near x, then $|f^\circ(x; v)| \leq M\|v\|$.

Theorem 5.1. *At least one of the following possibilities holds*:
(i) *the subsequence of envelope filtered centres is infinite*; *or*
(ii) *the sequence of function values is unbounded below.*

Proof: We assume that case (i) does not occur, and that there are $J - 1$ envelope filtered bent frames in total. Hence the final value of m is J. Consider the subsequence of iterates generated after $m = J$ occurs.

Both $E^{(J)}$ and $T^{(J)}$ are positive, and so each iteration of Step 5 must locate a point x which is not envelope filtered. There are three ways this can happen. First, $h(x) + \tau^{(k)} \leq h(y^{(k)})$, where $y^{(k)}$ is the least infeasible point in $\mathcal{F}^{(k)}$. In this case a new least infeasible point $y^{(k)}$ is found, and it satisfies $h(y^{(k+1)}) \leq h(y^{(k)}) - T^{(J)}$. Clearly this can only occur a finite number of times, and we label the final such new y as y_{last}.

Second, x is feasible, which means that either x is the first feasible point located (which can happen exactly once) or $y^{(k)}$ is also feasible and $f(x) < f(y^{(k)}) - E^{(J)}$. This can only occur a finite number of times or (ii) occurs.

Third, x is infeasible, and $(f(x) + E^{(J)}, h(x) + T^{(J)})$ is not dominated by any point in the filter. We define the area of a filter $\mathcal{F}^{(k)}$ as the area of the region in the f, h plane consisting of filtered pairs of values (f, h) satisfying $0 \leq h \leq h_{\max}$ and $f \leq f(y_{\text{last}})$. Clearly each new point accepted by the filter increases the area of the filter by at least $E^{(J)}T^{(J)}$. Hence if the sequence of envelope filtered frames is finite the area of $\mathcal{F}^{(k)}$ must tend to ∞ as k goes to infinity. This can only happen if $f^{(k)} \to -\infty$ as k goes to infinity, which is case (ii). $\square$

Condition (ii) of the above theorem is weaker than for the unconstrained case in that f is found to be unbounded below, but possibly for a sequence of infeasible points which are bounded away from the feasible region.

Theorem 5.2. *If h and f are locally Lipschitz near z^*, then*

$$h^\circ(z^*; v) \geq 0 \quad \text{or} \quad f^\circ(z^*; v) \geq 0,$$

where v is any vector such that there exists a sequence $\{(z^{(m)}, v^{(m)})\}$ with $v^{(m)} \in \mathcal{S}_z^{(m)}$ for all m, and such that (z^*, v) is a cluster point of this sequence.

Proof: First we restrict m to an increasing sequence of positive integers $\mathcal{M}$ such that $\{(z^{(m)}, v^{(m)})\}_{m \in \mathcal{M}}$ converges to (z^*, v). Each sampling set $\Phi(z^{(m)}, \delta_z^{(m)}; \mathcal{S}_z^{(m)})$ is envelope filtered, so for each $v^{(m)} \in \mathcal{S}_z^{(m)}$, either

$$f\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right) + \epsilon_z^{(m)} \geq f\left(z^{(m)}\right) \tag{20}$$

or both

$$h\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right) + \tau_z^{(m)} > h\left(z^{(m)}\right) \quad \text{and} \quad h\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right) > 0. \tag{21}$$

If this were not so, a point x in the filter would satisfy

$$f(x) \leq f\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right) + \epsilon_z^{(m)} < f\left(z^{(m)}\right)$$

and also either

$$0 = h(x) = h\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right) \leq h\left(z^{(m)}\right)$$

or

$$h(x) < h\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right) + \tau_z^{(m)} \leq h\left(z^{(m)}\right).$$

The left hand inequalities are because x envelope filters $z^{(m)} + \delta_z^{(m)} v^{(m)}$ and the right hand inequalities stem from the negation of (20, 21). Together, they imply that $z^{(m)}$ is filtered by x, which is excluded by the algorithm.

At least one of the two possibilities in (20, 21) must occur infinitely often. If the former occurs infinitely often, then

$$f\left(z^{(m)} + \delta_z^{(m)} v\right) - f\left(z^{(m)}\right) \geq \left[f\left(z^{(m)} + \delta_z^{(m)} v\right) - f\left(z^{(m)} + \delta_z^{(m)} v^{(m)}\right)\right] - \epsilon_z^{(m)}. \quad (22)$$

The convergence of $v^{(m)}$ to v and Assumption 3.5(d) imply that the first pair of terms and the final term on the right hand side of (22) go to zero faster than $\delta_z^{(m)}$. Hence, after dividing by $\delta_z^{(m)}$ and taking the limit $m \to \infty$, (22) shows $f^\circ(z^*; v)$ is bounded below by zero. A similar argument for h yields either

$$f^\circ(z^*; v) \geq 0 \quad \text{or} \quad h^\circ(z^*; v) \geq 0$$

as required. □

Corollary 5.1. *If either $h(z^{(m)} + \delta_z^{(m)} v^{(m)}) + \tau_z^{(m)} \leq h(z^{(m)})$ or $h(z^{(m)} + \delta_z^{(m)} v^{(m)}) = 0$ holds for each member of a subsequence $\{(z^{(m)}, v^{(m)})\}$ converging uniquely to (z^*, v), then $f^\circ(z^*, v) \geq 0$.*

Proof: When either $h(z^{(m)} + \delta_z^{(m)} v^{(m)}) + \tau_z^{(m)} \leq h(z^{(m)})$ or $h(z^{(m)} + \delta_z^{(m)} v^{(m)}) = 0$ then the point $z^{(m)} + \delta_z^{(m)} v^{(m)}$ must be envelope filtered because $f(z^{(m)} + \delta_z^{(m)} v^{(m)}) + \epsilon_z^{(m)} \geq f(z^{(m)})$. Hence $f^\circ(z^*; v) \geq 0$ by Theorem 5.2. □

Corollary 5.2. *The inequality $f^\circ(z^*; v) \geq 0$ holds for each (z^*, v) which is the limit of a subsequence of $\{(v^{(m)}, z^{(m)})\}$, for which $z^* \in \Omega$ and $v^{(m)} \in \mathcal{B}_+^{(k(m))}$ for all m.*

Proof: Theorem 4.4 shows that the bending condition (14) implies either $h(z^{(m)} + \delta_z^{(m)} v^{(m)}) + \tau_z^{(m)} \leq h(z^{(m)})$ or $h(z^{(m)} + \delta_z^{(m)} v^{(m)}) = 0$, for m sufficiently large. The required result then follows from Corollary 5.1. □

This second corollary is crucial. The limits of $\{\mathcal{B}_+^{(k(m))}\}$ contain sets of generators for the tangent cones at the limits of $\{z^{(m)}\}$, by Theorem 4.3. This provides information about f° along each member of these sets of generators. When f is strictly differentiable at these limit points, it allows convergence to KKT point(s) to be established under standard conditions.

Theorem 5.3. *Let z^* be a feasible limit point of the the sequence of envelope filtered centres. Then*

(i) *there is a set of generators for the tangent cone $T(z^*)$ at z^* such that $f^\circ(z^*; v) \geq 0$ for each of these generators*; *and*

(ii) *if f strictly differentiable at z^* then $v^T \nabla f(z^*) \geq 0$ for all feasible directions at z^*.*

Proof: Let $\mathcal{M}$ be an infinite increasing sequence of positive integers. Theorem 4.3 shows that if $\{z^{(m)}\}_{m\in\mathcal{M}}$ converges uniquely to z^* then the limits of $\{\mathcal{B}_+^{(k(m))}\}_{m\in\mathcal{M}}$ contain a set of generators for $T(z^*)$. Hence Corollary 5.2 shows that $f^\circ(z^*; v) \geq 0$ for all of these generators, which proves part (i). Strict differentiability means that $f^\circ(z^*; v) = v^T \nabla f^*$ for all v. Linearity implies that $v^T \nabla f^*$ is non-negative for all directions in the tangent cone, which yields part (ii). □

If both parts of Theorem 5.3 hold, then Assumption 3.1 means that Farka's lemma can be applied, and z^* is a KKT point of (1).

6. Summary

An algorithm for inequality constrained nonlinear optimization has been presented. The method uses filters, which allows it to handle infeasible starting points, and generate infeasible iterates. The method generates an infinite sequence of envelope filtered bent frames. The shape of these bent frames is somewhat restricted at frame centres near the boundary of the feasible region, but parts of the bent frames which lie in the null space of the estimated active constraint gradients may be chosen freely. This allows second order information to be incorporated in the shape of these frames.

In order to ensure convergence, a number of conditions are imposed upon the algorithm. In particular, linear independence of the active constraint gradients at all points on the boundary of the feasible region is assumed. It is shown that an identifiable subsequence of iterates has cluster points, and the nature of these cluster points is examined using Clarke's non-smooth calculus. This allows the analysis to be applied to objective functions which are locally Lipschitz, but not necessarily continuously differentiable. It is shown that cluster points at which the objective function is strictly differentiable are also KKT points. It is also shown that the objective function has interesting properties in certain directions at other cluster points. The analysis is purely local, and so the objective function may be non-smooth, or infinite at points remote to these cluster points.

Acknowledgments

The authors wish to thanks two anonymous referees for their many helpful comments which led to a much improved form of the paper.

References

C. Audet and J. E. Dennis Jr, "A pattern search filter method for nonlinear programming without derivatives," *SIAM Journal on Optimization* vol. 13, pp. 889–903, 2003.

F. H. Clarke, "*Optimization and Non-Smooth Analysis*, *SIAM Classics in Applied Mathematics*, 1990.

I. D. Coope and C. J. Price, "Frame based methods for unconstrained optimization," *Journal of Optimization Theory and Applications* vol. 107, pp. 261–274, 2000.

C. Davis, "Theory of positive linear dependence," *American Journal of Mathematics* vol. 76, pp. 733–746, 1954.

R. Fletcher, *Practical Methods of Optimization*," J. Wiley and Sons, 1987.

R. Fletcher and S. Leyffer, "Nonlinear programming without a penalty function," *Math. Prog. Series A* vol. 91, pp. 239–269, 2002.

R. Fletcher, S. Leyffer, and P. L. Toint, "On the global convergence of a filter-SQP algorithm," *SIAM Journal on Optimization* vol. 13, pp. 44–59, 2002.

M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications* vol. 4, pp. 303–320, 1969.

W. M. Jenkins, "A decimal coded evolutionary algorithm for constrained optimization," *Comput. and Structures* vol. 80, pp. 471–480, 2002.

J. H. Kim and H. Myung, "A two-phase evolutionary programming for general constrained optimization problem," in *Evolutionary Programming V*, MIT Press, 1996, pp. 295–304.

R. M. Lewis and V. Torczon, "Pattern search algorithms for bound constrained optimization," *SIAM Journal on Optimization* vol. 9, pp. 1082–1099, 1999.

R. M. Lewis and V. Torczon, "A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds," *SIAM Journal on Optimization* vol. 12, pp. 1075–1089, 2002.

S. Lucidi, M. Sciandrone, and P. Tseng, "Objective-derivative-free methods for constrained optimization," *Math. Prog. Series A* vol. 92, pp. 37–59, 2002.

M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, ed., Academic Press: London.

H. Peng, Z. Z. Feng, and L. H. Yan, "A comparison of evolutionary strategies and hybrid evolutionary strategies for constrained optimization problems," *Chinese Journal of Numerical Mathematics and Applications* vol. 20, pp. 55–60, 1998.

C. J. Price and I. D. Coope, "Frame based ray search algorithms in unconstrained optimization," *Journal of Optimization Theory and Applications* vol. 116, pp. 359–377, 2003.

C. J. Price and I. D. Coope, "Frames and grids in unconstrained and linearly constrained optimization: A non-smooth approach," *SIAM Journal on Optimization* vol. 14, pp. 415–438, 2003.

J. van Tiel, *Convex Analysis*, J. Wiley and Sons, 1984.

V. Torczon, "On the convergence of pattern search algorithms," *SIAM Journal on Optimization* vol. 7, pp. 1–25, 1997.

W.-C. Yu and Y.-X. Li, "A direct search method by the local positive basis for the nonlinearly constrained optimization," *Chinese Annals of Mathematics* vol. 2, pp. 269–280, 1981.