A SIMPLICIAL CONTINUATION DIRECT SEARCH METHOD* †

David W. Dreisigmeyer‡
Los Alamos National Laboratory
Los Alamos, NM 87545

April 26, 2007

Abstract

A direct search method for the class of problems considered by Lewis and Torczon [*SIAM J. Optim.*, 12 (2002), pp. 1075-1089] is developed. Instead of using an augmented Lagrangian method, a simplicial approximation method to the feasible set is implicitly employed. This allows the points our algorithm considers to conveniently remain within an *a priori* specified distance of the feasible set. In the limit, a positive spanning set is constructed in the tangent plane to the feasible set of every cluster point. In this way, we can guarantee that every cluster point is a stationary point for the objective function restricted to the feasible set.

Key words: Direct search methods, piecewise-linear approximations, simplicial continuation

AMS subject classifications: 90C56, 53C21

1 Introduction

The general problem we will look at is

$$\min_{\mathbf{x}\in\mathbb{R}^{N+K}} \qquad f(\mathbf{x}) \tag{1.1a}$$

$$\text{subject to :} \qquad \mathbf{g}(\mathbf{x}) = \mathbf{0} \tag{1.1b}$$

where $f : \mathbb{R}^{N+K} \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R}^{N+K} \to \mathbb{R}^N$. Let

$$\mathcal{M} \quad = \quad \{\mathbf{x}|\mathbf{x} \text{ satisfies } (1.1b)\}. \tag{1.2}$$

*LA-UR-07-2755

†Dedicated to Kurt Georg.

‡email: dreisigm@lanl.gov

We'll take $\mathcal{M}$ as a $\mathcal{C}^2$ Riemannian manifold of dimension K, which simply means that $\mathbf{g}(\mathbf{x})$ is $\mathcal{C}^2$ and $\nabla \mathbf{g}$ is full rank on $\mathcal{M}$.

If our problem is given by

$$\min_{\mathbf{y}\in\mathbb{R}^{N+K}} \quad f(\mathbf{y}) \tag{1.3a}$$
$$\text{subject to :} \quad \mathbf{g}(\mathbf{y}) = \mathbf{0} \tag{1.3b}$$
$$\mathbf{h}(\mathbf{y}) \leq \mathbf{0}, \tag{1.3c}$$

where $\mathbf{h} : \mathbb{R}^{N+K} \to \mathbb{R}^L$, we can convert it into the form in (1.1) by the use of slack variables $\mathbf{z} = [z_1, \ldots, z_L]^T$. Letting $\mathbf{x} = [\mathbf{y}^T \ \mathbf{z}^T]^T$ and $\mathbf{z} \odot \mathbf{z} = [z_1^2, \ldots, z_L^2]^T$, (1.3) becomes

$$\min_{\mathbf{x}\in\mathbb{R}^{N+K+L}} \quad f(\mathbf{y}) \tag{1.4a}$$
$$\text{subject to :} \quad \mathbf{g}(\mathbf{y}) = \mathbf{0} \tag{1.4b}$$
$$\mathbf{h}(\mathbf{y}) + \mathbf{z} \odot \mathbf{z} = \mathbf{0}. \tag{1.4c}$$

If the linear independence constraint qualification holds [12], then (1.4) is exactly (1.1).

Now let $\widetilde{\mathcal{M}} = \{\mathbf{y}|\mathbf{y} \text{ satisfies } (1.3b)\}$, $\Omega = \{\mathbf{y}|\mathbf{y} \text{ satisfies } (1.3c)\}$ and $\widetilde{\mathcal{N}} = \widetilde{\mathcal{M}} \cap \Omega$. If $\widetilde{\mathcal{N}}$ is compact, then $\mathcal{M}$ defined by (1.4b) and (1.4c) is also compact because $\mathcal{M}$ consists of multiple 'copies' of $\widetilde{\mathcal{N}}$ that are 'sewn' together. We will assume that $\mathcal{M}$ is compact in the sequel.

Lewis and Torczon developed an augmented Lagrangian pattern search algorithm in [11] that built on the method presented by Conn, Gould and Toint in [4]. The general idea is to use an augmented Lagrangian with a quadratic penalty term to 'guide' our iterates to a solution to (1.1). At each iteration of the algorithm, a fixed augmented Lagrangian was minimized, to a sufficient degree, in the ambient space $\mathbb{R}^{N+K}$ using a pattern search method.

One of the disadvantages of an augmented Lagrangian approach is the need to estimate the Lagrangian multipliers, along with other parameters, as the algorithm proceeds. To overcome this, Fletcher and Leyffer proposed a filter method in [7, 8]. Here one uses a measure of the constraints violations to ensure that we always remain 'close' to the feasible region $\mathcal{M}$.

This paper investigates an alternate way of ensuring we remain nearby $\mathcal{M}$ as the algorithm proceeds. The idea is to use the simplicial approximation algorithm in [3] to control how far away from $\mathcal{M}$ we can possibly be. Here, the maximum distance from $\mathcal{M}$ is measured via the Euclidean distance in $\mathbb{R}^{N+K}$ and is not approximated by the constraints violation, as in the filter method. Additionally, the algorithm in [3] gives us a way to reduce the size of the problem to one dependent on the dimensionality of $\mathcal{M}$, namely K, rather than having to do a direct search in the full ambient space $\mathbb{R}^{N+K}$, as in the Lewis and Torczon method.

We imagine the proposed method to be an alternate to that presented by Lewis and

Torczon in [11]. So the same assumptions used in [4, 11] will hold throughout this paper, slightly restated for our purposes:

AS1: The functions $f(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are $\mathcal{C}^2$, at least in a neighborhood of $\mathcal{M}$.

AS2: The iterates $\{\mathbf{x}^{(k)}\}$ lie on the compact manifold $\mathcal{M}$.

AS3: $\nabla\mathbf{g}(\mathbf{x})$ is full rank in a neighborhood of $\mathcal{M}$.

In the language of differential geometry, we assume that $f(\mathbf{x})$ is a $\mathcal{C}^2$ function defined on the $\mathcal{C}^2$ regular level set (1.1b) that implicitly defines $\mathcal{M}$ [6, 10].

The paper is organized as follows. In section 2 we describe the necessary components of the simplicial continuation algorithm in [3]. Some additional assumptions are spelled out in section 3. Given the assumptions **AS1** - **AS3**, these added requirements are met with probability 1. Our general procedure is describes in section 4. Section 5 derives some convergence results, where we also discuss some practical considerations when implementing the method in section 4. A discussion follows in section 6.

2 The simplicial continuation algorithm

Here we describe the necessary features of the simplicial continuation algorithm in [3] that we will need for our direct search algorithm (see [2] also). Originally this algorithm was designed in order to find a piecewise-linear approximation to implicitly defined manifolds. We are only concerned with using the algorithm in order to implicitly create a mesh such that all points on the mesh lie within some specified distance of the feasible region defined by the constraints in (1.1b).

Let the set $\{v_k\}_{k=0}^{N+K} \subset \mathbb{R}^{N+K}$ of points be affinely independent. Then the $(N+K)$-simplex σ with vertices $\{v_k\}_{k=0}^{N+K}$ is given by

$$\begin{aligned}\sigma &= [v_0, \ldots, v_{N+K}] \\ &= \left\{ v \in \mathbb{R}^{N+K} \,\middle|\, v = \sum_{j=0}^{N+K} \lambda_j v_j,\ \sum_{j=0}^{N+K} \lambda_j = 1 \text{ and } \lambda_j \geq 0 \right\}. \end{aligned} \tag{2.1}$$

The $(N+K-1)$-simplex $\phi_j = [v_0, \ldots, \hat{v}_j, \ldots, v_{N+K}] \subset \sigma$ is called a facet of σ. The hat on $\hat{v}_j$ means that that vertex is omitted from the simplex ϕ_j. A general $(N+K-J)$-face of σ is the convex hull of $(N+K-J+1)$ of the vertices of σ.

Definition 2.1 (Triangulation) *A triangulation $\mathcal{T}$ of $\Omega \subset \mathbb{R}^{N+K}$ is a set of $(N+K)$-simplices such that*

1. $\bigcup_{\sigma\in\mathcal{T}} \sigma = \Omega$*;*

2. *For all $\sigma_1, \sigma_2 \in \mathcal{T}$ either $\sigma_1 \cap \sigma_2$ is empty or a common face, and;*
3. *For every compact subset $C \subset \Omega$, $C \cap \sigma \neq \emptyset$ for only a finite number of $\sigma \in \mathcal{T}$.*

Each of the vertices v_j in a simplex σ is assigned an integer label $l_{\mathbf{g}}(v_j)$ that gives the number of nonnegative entries in $\mathbf{g}(v_j)$. Using this labeling, let us give some definitions.

Definition 2.2 (Completely Labeled) *An N-simplex is called completely labeled if, up to permutations of the vertices,*

$$\{l_{\mathbf{g}}(v_0), \ldots, l_{\mathbf{g}}(v_N)\} = \{0, \ldots, N\}.$$

Definition 2.3 (Transverse) *An $(N+J)$-simplex, $J = 1, \ldots, K$, ν is transverse if there is a completely labeled N-face contained in ν.*

We have the following result:

Theorem 2.4 ([3]) *Let the $(N+K)$-simplex σ contain a completely labeled N-face τ. Then the number of completely labeled N-faces of σ is between $(K+1)$ and 2^K, and the number of completely labeled facets of σ is between $(K+1)$ and $(K+\min(K, N+1))$.*

The way we will move from one simplex to a neighboring simplex is by pivoting. In particular, we will pivot the simplex by a reflection of one of its vertices through a neighboring edge. For a vertex v_j define the left and right neighbors as

$$v_{j-} \quad = \quad \begin{cases} v_{j-1} & \text{if } 0 < j \leq N+K, \\ v_{N+K} & \text{if } j = 0, \end{cases} \quad \text{and,} \tag{2.2a}$$

$$v_{j+} \quad = \quad \begin{cases} v_{j+1} & \text{if } 0 \leq j < N+K, \\ v_0 & \text{if } j = N+K, \end{cases} \tag{2.2b}$$

respectively. Then the reflection of v_j across the neighboring edge $[v_{j-}, v_{j+}]$ is given by

$$r(v_j) = v_{j-} - v_j + v_{j+}. \tag{2.3}$$

The new reflected simplex is given by

$$\sigma R_j = [v_0, \ldots, r(v_j), \ldots, v_{N+K}], \tag{2.4}$$

where R_j is the matrix corresponding to the operation in (2.3).

This pivoting by reflection generates a Freuenthal-Kuhn triangulation $\mathcal{F}(\sigma_0)$ of $\mathbf{R}^{N+K}$ given any initial simplex σ_0 [1]. The Kuhn triangulation $\mathcal{K}(\mathbf{v}^0, \delta)$ is given by the reflection triangulation of the initial simplex

$$\sigma_0 \quad = \quad \left[\mathbf{v}^0, \mathbf{v}^0 + \delta \mathbf{e}_1, \ldots, \mathbf{v}^0 + \delta \sum_{i=1}^{N+K} \mathbf{e}_i\right], \tag{2.5}$$

where the $\mathbf{e}_i$ are the standard basis elements for $\mathbb{R}^{N+K}$. Note that the diameter of σ_0 is $\mathrm{diam}(\sigma_0) = (N+K)\delta$. Any triangulation $\mathcal{F}(\sigma_0)$ can be obtained from a triangulation $\mathcal{K}(\mathbf{v}^0, 1)$ by an affine transformation of all the simplices in $\mathcal{K}(\mathbf{v}^0, 1)$. That is, there exists an $M \in \mathcal{GL}(N+K)$ and a $\mathbf{c} \in \mathbb{R}^{N+K}$ such that under the invertible affine map

$$\mathcal{A}(\mathbf{x}) \quad = \quad M\mathbf{x} + \mathbf{c}, \tag{2.6}$$

if $\sigma \in \mathcal{F}(\sigma_0)$ then $\mathcal{A}(\sigma) \in \mathcal{K}(\mathbf{0}, 1)$ [3].

3 Further assumptions

Now we'll examine the algorithms in section 2 in order to find out what we will need to assume in order for our direct search method to work correctly. These assumptions will place restrictions on the allowed level sets $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ and, on the allowed triangulations given a level set. The results here will be stated for $\mathcal{K}(\mathbf{v}^0, \delta)$. Completely analogous results hold for $\mathcal{F}(\sigma_0)$ via the relationship (2.6).

We can summarize what we need fairly quickly. Let $\mathcal{M}$ be the surface implicitly defined by $\mathbf{g}(\mathbf{x}) = \mathbf{0}$. There's no loss in generality in also assuming that $\mathcal{M}$ is connected, so we'll let $\mathcal{M}$ have only one component. Now define $\widetilde{\mathcal{K}}^m(\mathbf{v}^0, \delta)$, $m = 0, 1, 2, \ldots$, to be the set of all of the transverse simplices σ in $\mathcal{K}(\mathbf{v}^0, \delta/2^m)$ such that $\sigma \cap \mathcal{M} \neq \emptyset$. Let $\widetilde{\chi}^m = \{\tau^m_{k,i} \cap \mathcal{M}\}$, where $\tau^m_{k,i} \subset \sigma^m_k \in \widetilde{\mathcal{K}}^m(\mathbf{v}^0, \delta)$ is an N-face. Here k is the index on the simplices and i is the index on the N-faces of a given simplex. What we need is that $\widetilde{\chi}^\infty$ is dense in $\mathcal{M}$. This can be achieved under the following assumptions:

Complete labeling assumption: Every N-face τ that is sufficiently small and intersects $\mathcal{M}$ at one point can be completely labeled.

Finite intersection assumption: All of the (not necessarily transverse) N-faces of a given triangulation intersect $\mathcal{M}$ at a finite number of points.

Dimension of intersection assumption: Let ν be an $(N+K-J)$-face of σ, $J = 0, \ldots, K$, and $\mathrm{relint}(\nu)$ be the relative interior of ν. If $\nu \cap \mathcal{M} \neq \emptyset$ then $\dim(\mathrm{relint}(\nu) \cap \mathcal{M}) = K - J$.

The complete labeling assumption is a condition on our function $\mathbf{g}(\mathbf{x})$. **AS3** guarantees this will hold. The finite intersection assumption is a condition on our triangulation given a level set $\mathbf{g}(\mathbf{x}) = \mathbf{0}$. It is designed so that we 'pick up' all of the pieces of $\mathcal{M}$ as our algorithm proceeds. Under this assumption, the $\widetilde{\chi}^m$ have the nesting property

$$\widetilde{\chi}^0 \subset \quad \widetilde{\chi}^1 \subset \quad \widetilde{\chi}^2 \cdots . \tag{3.1}$$

The dimension of intersection assumption occurs with probability 1 for an arbitrary triangulation. If it holds and $\mathcal{M}$ is compact (which is **AS2**), then the finite intersection assumption also holds.

Similar to above, let $\widehat{\mathcal{K}}^m(\mathbf{v}^0, \delta)$, $m = 0, 1, 2, \ldots$, be the set of all of the (not necessarily transverse) simplices σ in $\mathcal{K}(\mathbf{v}^0, \delta/2^m)$ such that $\sigma \cap \mathcal{M} \neq \emptyset$. Additionally, let $\widehat{\chi}^m = \{\tau^m_{k,i} \cap \mathcal{M}\}$, where $\tau^m_{k,i} \subset \sigma^m_k \in \widehat{\mathcal{K}}^m(\mathbf{v}^0, \delta)$ is an N-face. The $\widehat{\chi}^m$ also have the nesting property

$$\widehat{\chi}^0 \subset \quad \widehat{\chi}^1 \subset \quad \widehat{\chi}^2 \cdots . \tag{3.2}$$

Since $\mathcal{M} \subset \widehat{\mathcal{K}}^m(\mathbf{v}^0, \delta)$, we have the following:

Lemma 3.1 *$\widehat{\chi}^\infty$ is dense in $\mathcal{M}$.*

PROOF. *Choose any point $\mathbf{x} \in \mathcal{M}$, and any $\epsilon > 0$. Then there exists an M such that there is a $\sigma \in \widehat{\mathcal{K}}^M(\mathbf{v}^0, \delta)$ where $\sigma \subset B_\epsilon(\mathbf{x})$, the open ball of radius ϵ centered at $\mathbf{x}$. Now pass to the subspace basis for $\mathcal{M}$ and the resulting topology.*

□

We can use the complete labeling and finite intersection assumptions to give us our desired result:

Theorem 3.2 *Let the triangulation $\widehat{\mathcal{K}}^m(\mathbf{v}^0, \delta)$ satisfy the finite intersection assumption for every $m < \infty$, and $\mathcal{M}$ satisfy the complete labeling assumption. Then $\widetilde{\chi}^\infty$ is dense in $\mathcal{M}$.*

PROOF. *Take the σ from Lemma 3.1. If $\sigma \in \widetilde{\mathcal{K}}^M(\mathbf{v}^0, \delta)$ we're done, so assume $\sigma \notin \widetilde{\mathcal{K}}^M(\mathbf{v}^0, \delta)$. Pick any N-face $\tau \subset \sigma$ such that $\tau \cap \mathcal{M} \neq \emptyset$. $\tau \cap \mathcal{M}$ has a finite number of points. Choose one of these points and call it $\mathbf{y}$. We need to show that $\mathbf{y} \in \tau' \cap \mathcal{M}$ for some $M' > M$ and $\tau' \subset \sigma' \in \widetilde{\mathcal{K}}^{M'}(\mathbf{v}^0, \delta)$.*

We can refine our triangulation, which also refines τ, until $\mathbf{y}$ is the only point in $\tau' \cap \mathcal{M}$, where $\tau' \subset \tau$. By the complete labeling assumption, doing further refinements if necessary, τ' is a completely labeled N-face and hence $\tau' \subset \sigma' \in \widetilde{\mathcal{K}}^{M'}(\mathbf{v}^0, \delta)$, $M' > M$.

□

Notice what this theorem tells us. First, $\mathcal{M} \not\subset \widetilde{\mathcal{K}}^m(\mathbf{v}^0, \delta)$ in general. However, as we refine the triangulation, we will contain more of $\mathcal{M}$ in $\widetilde{\mathcal{K}}^m(\mathbf{v}^0, \delta)$. In the limit, we can get arbitrarily close to any point in $\mathcal{M}$.

Let the point $\widetilde{\chi}^m_{k,i} \in \mathcal{M}$ satisfy $\widetilde{\chi}^m_{k,i} \subset \tau^m_{k,i}$ and $\widetilde{\chi}^m_{k,i} \subset \widetilde{\chi}^m$. Then we can find $\widetilde{\chi}^m_{k,i}$ by solving the following problem:

$$\min_{\mathbf{x} \in \tau^m_{k,i}} \qquad G(\mathbf{x}) = \mathbf{g}^T(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}). \tag{3.3}$$

Equation (3.3) will need to be solved repeatedly in the sequel. This can be done, e.g., by evaluating $\mathbf{g}(\mathbf{x})$ at the barycenter $b(\tau^m_{k,i})$ of $\tau^m_{k,i}$. This point will replace exactly one of the vertices of $\tau^m_{k,i}$ to give us a new completely labeled N-face. This procedure can be repeated until we converge to $\widetilde{\chi}^m_{k,i}$. Provided m is large enough, the unique solution $\widetilde{\chi}^m_{k,i}$ satisfies $G(\widetilde{\chi}^m_{k,i}) = 0$.

4 The general procedure

Now we can state our general procedure for solving (1.1). Start with an initial simplex $\sigma_0 \subset \mathbb{R}^{N+K}$ that generates a triangulation $\mathcal{F}(\sigma_0)$ with a completely labeled N-face $\tau_{0,0}$ that contains the point $\widetilde{\chi}_{0,0}$. Now, there are K other vertices of σ_0 that are not contained in $\tau_{0,0}$. Pick any one of these vertices, call it $v_\imath$, which will have the label $l_{\mathbf{g}}(v_\imath)$. Replace the unique vertex $v_{\hat{\imath}} \subset \tau_{0,0}$ that has the label $l_{\mathbf{g}}(v_{\hat{\imath}}) = l_{\mathbf{g}}(v_\imath)$ with $v_\imath$. This gives us another completely labeled N-face $\tau_{0,1}$ with an associated point $\widetilde{\chi}_{0,1}$. We can repeat this process to have $(K+1)$ completely labeled N-faces $\tau_{0,i}$ with their associated points $\widetilde{\chi}_{0,i}$, $i = 0, \ldots, K$. Notice that the set of vectors $\{\widetilde{\chi}_{0,i} - \widetilde{\chi}_{0,0}\}_{i=1}^K$ are linearly independent.

We'll also want K additional points. To generate these consider the opposing simplex ς_0 that is obtained from σ_0 as follows. All of the vertices in $\tau_{0,0}$ will also be vertices of ς_0. To find the additional K vertices of ς_0, let $b(\tau_{0,0})$ be the barycenter (or centroid) of $\tau_{0,0}$. Take the vertex $v_\imath \notin \tau_{0,0}$, $\imath = 1, \ldots, K$, and reflect it about $b(\tau_{0,0})$ along the line $[v_\imath, b(\tau_{0,0})]$ to get $v_\imath^*$. The set $\{v_\imath^*\}_{\imath=1}^K$ will also be vertices of ς_0. Note that generally $\varsigma_0 \notin \mathcal{F}(\sigma_0)$. Now we will repeat the above procedure to find K additional points $\widetilde{\chi}_{0,i} \subset \varsigma_0$, $i = K+1, \ldots, 2K$. As above, the set of vectors $\{\widetilde{\chi}_{0,i} - \widetilde{\chi}_{0,0}\}_{i=K+1}^{2K}$ are linearly independent.

Definition 4.1 (Quasi-Minimal Frame and Quasi-Minimal Point) *The set of points* $\mathsf{V} = \{\widetilde{\chi}_{0,i}\}_{i=1}^{2K}$ *is called a frame centered around* $\widetilde{\chi}_{0,0}$*. The frame is said to be quasi-minimal if*

$$f(\widetilde{\chi}_{0,i}) + \epsilon \quad \geq \quad f(\widetilde{\chi}_{0,0}), \;\; i = 1, \ldots, 2K, \tag{4.1}$$

for some $\epsilon > 0$*. The point* $\widetilde{\chi}_{0,0}$ *is then said to be quasi-minimal.*

If $\widetilde{\chi}_{0,0}$ is quasi-minimal, σ_0 is contracted as follows. Find the vectors from $\widetilde{\chi}_{0,0}$ to all of the vertices of σ_0. Halving the length of these vectors defines a new set of vertices that in turn define the contracted simplex σ_0^*. Then we can repeat the above procedure using $\sigma_0^* \in \mathcal{F}(\sigma_0^*)$.

If $\widetilde{\chi}_{0,0}$ is not quasi-minimal, then there is a $\widetilde{\chi}_{0,i}$, $i \neq 0$, that satisfies the sufficient descent criterion specified by (4.1). If $\widetilde{\chi}_{0,i} \in \sigma_0$, we can repeat the above procedure using $\mathcal{F}(\sigma_0)$. If $\widetilde{\chi}_{0,i} \in \varsigma_0$, we repeat the above procedure using $\mathcal{F}(\varsigma_0)$. In either case, we only need to find a new opposing simplex and the required points on it.

One thing we will need to do is align our simplex σ_0 with $\mathcal{M}$ to find the new simplex σ_a, called the aligned simplex. We do this with an initial translation: Let

$\sigma_t = \sigma_0 + \widetilde{\chi}_{0,0} - b(\tau_{0,0})$. This may require us to contract σ_t so that the N-face $\tau^*_{0,0}$ of σ_t associated with $\widetilde{\chi}_{0,0}$ is still completely labeled. Now find the $\widetilde{\chi}_{0,i}$, $i = 1, \ldots, K$, associated with σ_t. Form the unit length vectors $w_i = (\widetilde{\chi}_{0,i} - \widetilde{\chi}_{0,0})/\|(\widetilde{\chi}_{0,i} - \widetilde{\chi}_{0,0})\|$. If

$$|\det([w_1, \ldots, w_K])| \quad \geq \quad d^* \tag{4.2}$$

for some user specified $d^* > 0$, we say that σ_t is aligned. Then $\sigma_a = \sigma_t$. Otherwise rotate σ_t, holding $\widetilde{\chi}_{0,0}$ fixed, until (4.2) is satisfied. Since the affine plane M defined by the set of points $\{\widetilde{\chi}_{0,i}\}_{i=0}^{K}$ forms a local model for $\mathcal{M}$, we will likely only need to rotate once in order to satisfy (4.2). First translate everything by $\widetilde{\chi}_{0,0}$ to the origin. Let the columns of the $(N+K)$-by-K matrix M be a basis for the translation of M. Then choose any $(N+K)$-by-$(N+K)$ unitary matrix U such that

$$M^T U \hat{\tau}^*_{0,0} \quad \approx \quad 0, \tag{4.3}$$

where $\hat{\tau}^*_{0,0}$ is the translation of $\tau^*_{0,0} \subset \sigma_t$ by $\widetilde{\chi}_{0,0}$. Now translate everything back to $\widetilde{\chi}_{0,0}$ to have the new simplex $\bar{\sigma}_t$ with the new N-face $\bar{\tau}^*_{0,0}$ associated with $\widetilde{\chi}_{0,0}$. What we are trying to do is make $\bar{\tau}^*_{0,0}$ approximately normal to $\mathcal{M}$. After each rotation we need to check that the new $\bar{\tau}^*_{0,0}$ is still completely labeled, contracting $\bar{\sigma}_t$ if necessary. Now check if $\bar{\sigma}_t$ satisfies (4.2). The exact solution to (4.3) is given by the following: Let $M = W_M S_M V_M^T$ and $\hat{\tau}^*_{0,0} = WSV^T$ be the reduced SVDs [9]. Let $Z_M = W_M^\perp$ and $Z = W^\perp$. Then

$$U \quad = \quad [Z_M \; W_M] \, [W \; Z]^T \, . \tag{4.4}$$

Combining all of the above gives us our algorithm for solving (1.1):

Algorithm 4.2 ***(Basic Method)***

Step 0: *Let $m = k = 0$. Choose an initial simplex $\sigma^{(0)}$, an initial point $\widetilde{\chi}^{(0)} \subset \sigma^{(0)}$ and the constants $1 > \beta > 0$ and $\epsilon > 0$.*

Step 1: *Align $\sigma^{(k)}$ to find $\sigma_a^{(k)}$. Find the points $\{\widetilde{\chi}_i^{(k)}\}_{i=1}^{K} \subset \sigma_a^{(k)}$ around $\widetilde{\chi}^{(k)}$. Finish constructing the frame $\mathsf{V}^{(k)} = \{\widetilde{\chi}_i^{(k)}\}_{i=1}^{2K}$ using the aligned opposing simplex $\varsigma_a^{(k)}$ and calculate the function values on this frame. Let $\delta^{(k)} = \min_i(\|\widetilde{\chi}_i^{(k)} - \widetilde{\chi}^{(k)}\|)$.*

Step 2: *Choose the lowest point $\widetilde{\chi}^{(k+1)}$ in $\mathsf{V}^{(k)} \cup \{\widetilde{\chi}^{(k)}\}$. Let $\sigma^{(k+1)}$ be the simplex associated with $\widetilde{\chi}^{(k+1)}$.*

Step 3: *If $f(\widetilde{\chi}^{(k+1)}) < f(\widetilde{\chi}^{(k)}) - \epsilon(\delta^{(k)})^{1+\beta}$, let $k \leftarrow k+1$ and go to Step 1.*

Step 4: *Contract $\sigma^{(k)}$ to find the simplex $\sigma^{(k+1)}$. Let $\zeta^{(m)} = \widetilde{\chi}^{(k)}$ and $\mathsf{Z}^{(m)} = \mathsf{V}^{(k)}$. Set $\widetilde{\chi}^{(k+1)} = \widetilde{\chi}^{(k)}$. Let $m \leftarrow m+1$ and $k \leftarrow k+1$.*

Step 5: *If the stopping criteria are satisfied, return $\widetilde{\chi}^{(k)}$. Otherwise go to Step 1.*

Algorithm 4.2 bears a strong resemblance to the frame based method in [5]. We will use this resemblance in section 5 to prove our convergence results using the quasi-minimal iterates $\{\zeta^{(m)}\}_{m=0}^{\infty}$.

We can also employ an optional `SEARCH` step to have an additional finite set of points $\mathsf{S}^{(k)} \subset \mathcal{M}$. Then $\widetilde{\chi}^{(k+1)}$ in Step 2 would be chosen from $\mathsf{S}^{(k)} \cup \mathsf{V}^{(k)} \cup \{\widetilde{\chi}^{(k)}\}$. The user is free to specify this `SEARCH` step by any finite method.

5 Convergence analysis and practical considerations

Now we will prove our convergence results for Algorithm 4.2. The results here will be completely analogous to the convergence results presented in [5]. First, since $\mathcal{M}$ is compact and $f(\mathbf{x})$ is $\mathcal{C}^2$ on $\mathcal{M}$, we have the following result:

Theorem 5.1 ([5]) *The sequence of quasi-minimal iterates $\{\zeta^{(m)}\}$ is infinite.*

Now pick some cluster point $z^{(\infty)}$ of $\{\zeta^{(m)}\}$ and a subsequence $\{\widetilde{\zeta}^{(m)}\}$ of $\{\zeta^{(m)}\}$ that converges to $z^{(\infty)}$. An element $\widetilde{\zeta}^{(i)} \in \{\widetilde{\zeta}^{(m)}\}$ will be associated with the set $\widetilde{\mathsf{Z}}^{(i)}$. What we want to show is that $z^{(\infty)}$ is a stationary point of $f(\mathbf{x})$ on $\mathcal{M}$. Pick an $\varepsilon > 0$ sufficiently small and let $\mathcal{V}_\varepsilon(z^{(\infty)}) = B_\varepsilon(z^{(\infty)}) \cap \mathcal{M}$ be a neighborhood on $\mathcal{M}$ around $z^{(\infty)}$, where $B_\varepsilon(z^{(\infty)}) \subset \mathbb{R}^{N+K}$ is the open ball of radius ε around $z^{(\infty)}$. We'll assume that $\{\widetilde{\zeta}^{(m)}\} \subset \mathcal{V}_\varepsilon(z^{(\infty)})$. This gives us the following:

Theorem 5.2 *Each cluster point $z^{(\infty)}$ of the sequences of quasi-minimal iterates is a stationary point of $f(\mathbf{x})$ on $\mathcal{M}$.*

PROOF. Consider the tangent plane to $\mathcal{M}$ at some cluster point $z^{(\infty)}$, denoted by $\mathcal{T}_{z^{(\infty)}}\mathcal{M}$. For ε sufficiently small, we can model $\mathcal{V}_\varepsilon(z^{(\infty)})$ by an appropriate neighborhood of the origin of $\mathcal{T}_{z^{(\infty)}}\mathcal{M}$. From the sufficient descent condition

$$f(\widetilde{\chi}_k^{(m)}) - f(\widetilde{\zeta}^{(m)}) \quad \geq \quad -\epsilon(\delta^{(m)})^{1+\beta}, \tag{5.1}$$

for $k \in \{1, \ldots, 2K\}$. Additionally

$$f(\widetilde{\chi}_i^{(m)}) - f(\widetilde{\zeta}^{(m)}) \quad = \quad \nabla f(\widetilde{\zeta}^{(m)}) \cdot (\widetilde{\chi}_i^{(m)} - \widetilde{\zeta}^{(m)}) + O((\delta^{(m)})^2), \tag{5.2}$$

where $i \in \{1, \ldots, K\}$. From the construction of the opposing simplex, for every $j \in \{K+1, \ldots, 2K\}$ and for some $i(j) \in \{1, \ldots, K\}$,

$$f(\widetilde{\chi}_j^{(m)}) - f(\widetilde{\zeta}^{(m)}) \quad = \quad -C_{i,j}^{(m)} \nabla f(\widetilde{\zeta}^{(m)}) \cdot (\widetilde{\chi}_{i(j)}^{(m)} - \widetilde{\zeta}^{(m)}) + O((\delta^{(m)})^2), \tag{5.3}$$

where $C_{i,j}^{(m)} > 0$ is some constant.

Now, because our simplices are aligned, $(N+K)^2\delta^{(m)} \geq \|\widetilde{\chi}_i^{(m)} - \widetilde{\zeta}^{(m)}\| \geq \delta^{(m)}$. Letting $\hat{x}_i^{(m)} = (\widetilde{\chi}_i^{(m)} - \widetilde{\zeta}^{(m)})/\|\widetilde{\chi}_i^{(m)} - \widetilde{\zeta}^{(m)}\|$, we then have that

$$|\nabla f(\widetilde{\zeta}^{(m)}) \cdot \hat{x}_i^{(m)}| \quad \leq \quad O((\delta^{(m)})^\beta). \tag{5.4}$$

Additionally, the set of vectors $\{\widetilde{\chi}_i^{(m)} - \widetilde{\zeta}^{(m)}\}_{i=1}^K$ are linearly independent for every m and span some K plane $\mathsf{T}^{(m)}$. As $m \to \infty$, we have that $\mathsf{T}^{(\infty)} \subset \mathcal{T}_{z^{(\infty)}}\mathcal{M}$. From (4.2) we see that $\mathsf{T}^{(\infty)} = \mathcal{T}_{z^{(\infty)}}\mathcal{M}$. Hence, $f(z^{(\infty)})$ is a stationary point on $\mathcal{M}$. Since $z^{(\infty)}$ was an arbitrary cluster point, we have the result.

□

The opposing simplex was used in (5.3) in order to construct a positive spanning set. The aligning of the simplices was used to derive (5.4). As we approach a cluster point, the aligning procedure will typically only involve a simple translation. This is because in the neighborhood $\mathcal{V}_\varepsilon(z^{(\infty)})$, $\mathcal{T}_{\mathbf{x}}\mathcal{M} \approx \mathcal{T}_{z^{(\infty)}}\mathcal{M}$ for $\mathbf{x} \in \mathcal{V}_\varepsilon(z^{(\infty)})$. This suggests that we only perform the aligning procedure when $\delta^{(k)}$ is sufficiently small, i.e., $\delta^{(k)} < \delta^*$ for some user specified constant $\delta^* > 0$. As we approach $z^{(\infty)}$, our aligning procedure will likely involve an initial translation and rotation to make our simplices approximately normal to $\mathcal{M}$. After that we will typically only need to perform translations in order to keep the simplices aligned if we remain in $\mathcal{V}_\varepsilon(z^{(\infty)})$ because the simplices will already be approximately normal to $\mathcal{M}$. Also, by delaying the alignment procedure, the probability that the aligned simplex is completely labeled increases because we are remaining closer to $\mathcal{M}$.

In addition to delaying the alignment procedure, we do not need to initially find the exact $\widetilde{\chi}_i^{(k)}$ in Step 1 of Algorithm 4.2 by solving (3.3). Instead, we can take the barycenter of each completely labeled N-face as being an approximation to $\widetilde{\chi}_i^{(k)}$. As the size of our simplices decreases, this approximation becomes better. When we begin aligning our simplices, we can then begin to find the exact $\widetilde{\chi}_i^{(k)}$. This guarantees that we are only looking at how $f(\mathbf{x})$ can change on $\mathcal{M}$ itself since, in the limit, we are finding the change in $f(\mathbf{x})$ restricted to $\mathcal{T}_{z^{(\infty)}}\mathcal{M}$ for some cluster point $z^{(\infty)}$. Additionally, by finding the $\widetilde{\chi}_i^{(k)}$ when the simplices are of a sufficiently small diameter, we can speed up the solution of (3.3).

6 Discussion

We've developed a direct search algorithm that is designed to solve the same class of problems as the Lewis and Torczon method in [11]. In contrast to using an augmented Lagrangian to 'pull' us onto the feasible set, we use a simplex method that guarantees we are never further away from the feasible set than some *a priori* specified distance. This makes our method similar to the filter based algorithm of Fletcher and Leyffer [7]. In the limit, it becomes similar to the frame based method of Coope and Price [5] restricted to the feasible set $\mathcal{M}$. As in [11], we will delay any extensive numerical testing of our algorithm until well developed software is constructed. The results of future numerical work will be reported as they become available.

7 Acknowledgments

The author would like to thank Eugene L. Allgower (Colorado State University) for his helpful discussions regarding the simplicial continuation method. Additional appreciations go to John E. Dennis, JR. (Rice University) for his help in developing the direct search algorithm.

References

[1] E. L. Allgower and K. Georg. Generation of triangulations by reflections. *Utilitas Mathematica*, 16:123–129, 1979.

[2] E. L. Allgower and K. Georg. Numerical path following. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, volume 5, pages 3–207. North-Holland, 1997.

[3] E. L. Allgower and P. H. Schmidt. An algorithm for piecewise-linear approximation of an implicitly defined manifold. *SIAM Journal on Numerical Analysis*, 22:322–346, 1985.

[4] A. R. Conn, N. I. M. Gould, and P. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28:545–572, 1991.

[5] I. D. Coope and C. J. Price. Frame based methods for unconstrained optimization. *Journal of Optimization Theory and Applications*, 107:261–274, 2000.

[6] D. W. Dreisigmeyer. Equality constraints, Riemannian manifolds and direct search methods. *Submitted to SIOPT*.

[7] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming, Series A*, 91:239–269, 2002.

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

[9] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins, 3rd edition, 1996.

[10] J. M. Lee. *Introduction to Smooth Manifolds*. Springer, 2003.

[11] 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*, 12:1075–1089, 2002.

[12] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.