The Linear Convergence of a Successive Linear Programming Algorithm*

Michael C. Ferris[†]   Sergei K. Zavriev[‡]

December 3, 1996

Abstract

We present a successive linear programming algorithm for solving constrained nonlinear optimization problems. The algorithm employs an Armijo procedure for updating a trust region radius. We prove the linear convergence of the method by relating the solutions of our subproblems to standard trust region and gradient projection subproblems and adapting an error bound analysis due to Luo and Tseng. Computational results are provided for polyhedrally constrained nonlinear programs.

1 Introduction

In this paper we develop a first order technique for constrained optimization problems based on subproblems with a linear objective function. There have been many attempts to generate robust successive linear programming algorithms for nonlinear programming (see, for example [1, 14, 15, 19, 24, 26, 28]). This paper specifies a new algorithm that has the same convergence properties as the gradient projection method. We modify the conditional gradient method [3] (sometimes called the Frank-Wolfe method [2, 16]) for polyhedrally constrained optimization problems to obtain a linearly convergent algorithm. The only known linear rate of convergence is given in [12, 13], but this is for problems where the constraints are not polyhedral, but have "positive curvature"; there are even simple examples [3, p. 199]

*This work is partially supported by the National Science Foundation under grant CCR-9157632.

[†]Computer Sciences Department, University of Wisconsin, Madison, WI 53706

[‡]Operations Research Department, Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia, 119899

that show the conditional gradient method has a sublinear rate of convergence.

The main idea is to use a polyhedral trust region constraint in the standard linear programming subproblem of the conditional gradient method. This maintains the attractive computational feature that all subproblems are linear programs, for which a wealth of highly effective and efficient software has been developed. It also guarantees that every subproblem we generate has an optimal solution. We develop an adaptive mechanism to update the trust region parameter that guarantees the linear convergence of the method under the standard assumptions given by Luo and Tseng [20].

Our analysis is somewhat involved and relies upon estimating descent properties of the directions generated by our algorithm in terms of standard quadratic subproblems arising for example in gradient projection algorithms. To our knowledge, every linearly convergent technique for such problems is based on such quadratic subproblems. The main contribution of this paper is that the solutions of our linear subproblems are considered as approximate solutions of the quadratic subproblems and analysis is carried out to precisely quantify this relationship.

We will consider the following optimization problem

$$\mathcal{P} : \inf_{x \in X} f(x),$$

where $X \subset \mathbb{R}^n$ is a nonempty closed convex set, and the objective function f is assumed to be continuously differentiable. Furthermore, it will be assumed that f has a Lipschitz continuous gradient ∇f on X, that is

$$\forall x, y \in X, \ \ \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| .$$

Here L is a positive scalar and $\|\cdot\|$ represents the Euclidean norm on $\mathbb{R}^n$. We also suppose that

$$f_{\text{opt}} = \inf_{x \in X} f(x) > -\infty,$$

and the optimal solution set for $\mathcal{P}$ is not empty.

Let $\mathcal{N}_X(x)$ denote the normal cone to X at $x \in X$. The stationary set for the problem $\mathcal{P}$ is the set of all points satisfying the first order optimality conditions for $\mathcal{P}$, that is

$$\begin{aligned} X_{\text{stat}} &:= \{x \in X \mid 0 \in \nabla f(x) + \mathcal{N}_X(x)\} \\ &= \{x \in X \mid \forall y \in X, \langle \nabla f(x), y - x \rangle \geq 0\} . \end{aligned}$$

Our algorithm will converge under standard assumptions to a stationary point. Conditions relating stationary points to solutions of $\mathcal{P}$ are well known.

Our method resembles a trust region method (see [7, 9]) by virtue of the fact that the subproblems are of the form:

$$\mathcal{LP}(x,r): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h\rangle \\ \text{subject to} & x+h \in X, \\ & \|h\|_* \leq r, \end{array}$$

where $\|\cdot\|_*$ represents an arbitrary norm on $\mathbb{R}^n$ and r is the trust region radius. We actually prove convergence of our algorithm based on X being a convex set and the trust region constraint being determined by an arbitrary norm. When specialized to the case where X polyhedral and $\|\cdot\|_*$ is a polyhedral norm, the subproblems $\mathcal{LP}(x,r)$ are linear programs and our algorithm is a successive linear programming algorithm.

There are two key approximations in our analysis. The first is that we relate solutions of our linear programming descent problem $\mathcal{LP}(x,r)$ to the following subproblem

$$\mathcal{QP}(x,a): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h\rangle + \frac{1}{2a}\|h\|^2 \\ \text{subject to} & x+h \in X, \end{array}$$

where a is a corresponding stepsize. When X is polyhedral, $\mathcal{QP}(x,a)$ is a quadratic programming problem. It is important that we guarantee in our algorithm that the trust region radius r is not too small. We therefore employ a special Armijo type procedure for choosing r at each iteration. The essential difficulty in the proof relates to finding an expression for the inverse $a(x,r)$ of $r(x,a)$, and this is explored in Section 2. The key relationship is between solutions of $\mathcal{LP}(x,r)$ and $\mathcal{QP}(x,a)$ appears as Lemma 4.1.

Usually rapid convergence techniques (i.e., at least a linear rate) for solving $\mathcal{P}$ require the solution of a subproblem with a quadratic objective function at each iteration $k = 1, 2, \ldots$. For example, the gradient projection algorithm solves subproblems of the form $\mathcal{QP}(x,a)$. The second piece of our analysis, given as Theorem 2.10 in Section 2 of this paper, relates the solutions of $\mathcal{QP}(x,r)$ to solutions of the following (standard) trust region problem

$$\mathcal{TP}(x,r): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h\rangle \\ \text{subject to} & x+h \in X, \\ & \|h\| \leq r. \end{array}$$

Note that the objective function of this problem is linear and that the trust region constraint is defined in terms of the Euclidean norm.

In Section 3 we state the general form of the method which we call the Sequential Linearization Method, and prove its linear convergence to the stationary set $X_{\rm stat}$ in Section 4. The special case of polyhedral constraints is considered in Section 5. The resulting algorithm is a linearly convergent sequential linear programming (SLP) technique, for which we also provide some numerical results using Matlab, Cplex and the Cute suite of problems.

2 Technical Preliminaries

For every $x \in X$, consider the following auxiliary problem

$$\mathcal{TP}(x,r): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h \rangle \\ \text{subject to} & x+h \in X, \\ & \|h\| \le r, \end{array}$$

where $r \geq 0$ is a parameter of the problem. Let $\mathcal{H}(x,r)$ and $v(x,r)$ denote the optimal set and the optimal value of $\mathcal{TP}(x,r)$ respectively. We shall relate this to the following problem:

$$\mathcal{TP}(x,\infty): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h \rangle \\ \text{subject to} & x+h \in X. \end{array}$$

Note that it is possible for $v(x,\infty) = -\infty$ and $\mathcal{H}(x,\infty) = \emptyset$.

The standard optimality conditions (see [21, 25]) for the problem $\mathcal{TP}(x,r)$ provide the following lemma.

Lemma 2.1 *i. For every $x \in X$, $r > 0$ and $h \in \mathbb{R}^n$ (a) and (b) are equivalent:*

(a) $h \in \mathcal{H}(x,r)$;

(b) there exists $\lambda \geq 0$ such that

$$\begin{array}{l} 0 \in \nabla f(x) + \lambda h + \mathcal{N}_X\left(x+h\right), \\ x+h \in X, \\ \lambda(\|h\| - r) = 0, \ \|h\| \le r. \end{array}$$

ii. For every $x \in X$ and $h \in \mathbb{R}^n$ (a) and (b) are equivalent:

(a) $h \in \mathcal{H}(x,\infty)$;

(b) $0 \in \nabla f(x) + \mathcal{N}_X\left(x+h\right)$, $x+h \in X$.

The following corollary is immediate from Lemma 2.1.

Corollary 2.2 *For every $x \in X$ and $r > 0$ the following assertion holds:*

$$x \in X_{\text{stat}} \iff v(x, r) = 0.$$

For a subset $S \subseteq \mathbb{R}^n$, let $\Pi_S(x)$ denote the set of all orthogonal projections of x on S, that is

$$\Pi_S(x) := \{y \in S \mid \|x - y\| = \operatorname{dist}(x, S)\}.$$

Here $\operatorname{dist}(x, S)$ represents the distance from x to S

$$\operatorname{dist}(x, S) := \inf_{y \in S} \|x - y\|.$$

Let $\pi_S(x)$ be a typical element of $\Pi_S(x)$; it is clear that if S is closed and convex then

$$\Pi_S(x) = \{\pi_S(x)\}.$$

For $x \in X$ and $a \geq 0$ define

$$p_0(x, a) := \pi_X(x - a\nabla f(x)) - x.$$

Clearly, for $a > 0$, $p_0(x, a)$ is the unique solution of the following problem:

$$\mathcal{QP}(x, a): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h \rangle + \frac{1}{2a}\|h\|^2 \\ \text{subject to} & x + h \in X. \end{array}$$

The following two simple results prove to be useful in the sequel.

Lemma 2.3 ([25]) *Let $x \in X$ and $a > 0$ be given. The following assertions hold.*

i. $h = p_0(x, a)$ *if and only if*

$$0 \in \nabla f(x) + \frac{1}{a}h + \mathcal{N}_X(x + h), \; x + h \in X;$$

ii. $\langle \nabla f(x), p_0(x, a) \rangle \leq -\frac{1}{a}\|p_0(x, a)\|^2$.

Corollary 2.4 *Let $h \in \mathbb{R}^n$ be a limit point of a sequence $\{p_0(x, a_k)\}$, $a_k \to \infty$, then $h \in \mathcal{H}(x, \infty)$.*

We now define a residual function $r_0(x,\cdot):\mathbb{R}_+\to\mathbb{R}_+$ by

$$r_0(x,a) := \|p_0(x,a)\|,$$

where $\mathbb{R}_+ := \{x\in\mathbb{R} \mid x\ge 0\}$. We collect some well known properties of the function $r_0(x,\cdot):\mathbb{R}_+\to\mathbb{R}_+$ in the following lemma. The proofs can be found in [10, 17, 27].

Lemma 2.5 *For every $x\in X$ the following properties hold:*

i. $r_0(x,\cdot)\in\mathcal{C}(\mathbb{R}_+)$;

ii. $\forall a'\ge a>0,\ r_0(x,a')\ge r_0(x,a)$.

iii. $\forall a'\ge a>0,\ \frac{r_0(x,a')}{a'}\le\frac{r_0(x,a)}{a}$.

iv. $\forall a>0,\ r_0(x,a)=0 \iff x\in X_{\text{stat}}$.

The following lemma bounds the residual in terms of any solution of $\mathcal{TP}(x,r)$.

Lemma 2.6 *Let $x\in X$ and suppose $\mathcal{H}(x,\infty)\neq\emptyset$. Then for every $h\in\mathcal{H}(x,\infty)$ and every $a\in[0,\infty)$,*

$$r_0(x,a)=\|p_0(x,a)\|\le\|h\|.$$

Proof The result is trivial for $a=0$. Otherwise, let us fix arbitrary $x\in X$, $h\in\mathcal{H}(x,\infty)$, and $a>0$. Suppose that

$$\|h\|<\|p_0(x,a)\|. \tag{1}$$

By the definition of $\mathcal{H}(x,\infty)$ we have

$$\langle\nabla f(x),h\rangle\le\langle\nabla f(x),p_0(x,a)\rangle.$$

Hence, applying (1) we obtain

$$\langle\nabla f(x),h\rangle+\frac{1}{2a}\|h\|^2<\langle\nabla f(x),p_0(x,a)\rangle+\frac{1}{2a}\|p_0(x,a)\|^2,$$

which is a contradiction to the fact that $p_0(x,a)$ is the solution of $\mathcal{QP}(x,a)$. $\square$

We now investigate some limiting properties of this residual function. For every $x \in X$ we let

$$r_0(x, \infty) := \lim_{a \to \infty} r_0(x, a).$$

Obviously, it is possible that $r_0(x, \infty) = \infty$ for some $x \in X$. The following lemma allows us to identify certain limits of the functions $p_0(x, \cdot)$ and $r_0(x, \cdot)$. It is really just a technical result used for proving the ensuing result.

Lemma 2.7 *Let $x \in X$ and suppose for some $a' \in (0, \infty)$*

$$r_0(x, a') = r_0(x, \infty). \tag{2}$$

Then

$$\forall a \in [a', \infty), p_0(x, a) = p_0(x, a'). \tag{3}$$

In this case,

$$\lim_{a \to \infty} p_0(x, a) = p_0(x, a') \in \mathcal{H}(x, \infty).$$

Proof Let $a > a'$. Lemma 2.5(ii) and (2) show that

$$\|p_0(x, a)\| = \|p_0(x, a')\|. \tag{4}$$

We now derive a contradiction to the supposition that

$$p_0(x, a) \neq p_0(x, a').$$

Since $p_0(x, a)$ and $p_0(x, a')$ are the unique solutions of the problems $\mathcal{QP}(x, a)$ and $\mathcal{QP}(x, a')$ respectively, we have

$$\begin{aligned}
\langle \nabla f(x), p_0(x, a) \rangle + \frac{1}{2a} \|p_0(x, a)\|^2 &< \langle \nabla f(x), p_0(x, a') \rangle + \frac{1}{2a} \|p_0(x, a')\|^2 \\
&< \langle \nabla f(x), p_0(x, a') \rangle + \frac{1}{2a'} \|p_0(x, a')\|^2 \\
&< \langle \nabla f(x), p_0(x, a) \rangle + \frac{1}{2a'} \|p_0(x, a)\|^2 .
\end{aligned}$$

Hence, by (4)

$$\langle \nabla f(x), p_0(x, a) \rangle < \langle \nabla f(x), p_0(x, a') \rangle < \langle \nabla f(x), p_0(x, a) \rangle,$$

which is a contradiction, and so (3) is proved. It now follows from (3) that $\lim_{a \to \infty} p_0(x, a) = p_0(x, a')$, so that Corollary 2.4 implies $p_0(x, a') \in \mathcal{H}(x, \infty)$. $\square$

For $x \in X$ with $\mathcal{H}(x,\infty) \neq \emptyset$, we define $p_0(x,\infty)$ as follows:

$$p_0(x,\infty) := \operatorname*{arg\,min}_{h \in \mathcal{H}(x,\infty)} \|h\| .$$

In particular, note the difference between the definition of $p_0(x,\infty)$ and $r_0(x,\infty)$ to that of $p_0(x,a)$ and $r_0(x,a)$ given above.

Lemma 2.8 *For every $x \in X$ the following assertions hold:*

i. $r_0(x,\infty) < \infty \iff \mathcal{H}(x,\infty) \neq \emptyset$.

ii. If $r_0(x,\infty) < \infty$ *then*

$$r_0(x,\infty) = \|p_0(x,\infty)\| , \tag{5}$$

and for all $r \geq r_0(x,\infty)$

$$\mathcal{H}(x,r) = \mathcal{H}(x,\infty) \cap \mathbb{B}_r ,$$

where $\mathbb{B}_r := \{x \in \mathbb{R}^n \mid \|x\| \leq r\}$.

Proof i. Let $\mathcal{H}(x,\infty) \neq \emptyset$. Applying Lemma 2.6 we immediately get that $r_0(x,\infty) < \infty$. Conversely, let $r_0(x,\infty) < \infty$. Then for an arbitrary sequence $\{a_k\} \to \infty$ the sequence $\{p_0(x,a_k)\}$ is bounded and has a limit point $p \in \mathbb{R}^n$. By Corollary 2.4, we have that $p \in \mathcal{H}(x,\infty)$, hence $\mathcal{H}(x,\infty) \neq \emptyset$.

ii. Suppose that $r_0(x,\infty) < \infty$. Lemma 2.6 yields $\|p_0(x,\infty)\| \geq r_0(x,\infty)$. On the other hand, by Corollary 2.4

$$\infty > r_0(x,\infty) = \lim_{a \to \infty} \|p_0(x,a)\| \geq \min_{h \in \mathcal{H}(x,\infty)} \|h\| = \|p_0(x,\infty)\| .$$

For the final assertion of the lemma, let us fix an r with

$$r \geq r_0(x,\infty) , \tag{6}$$

and take an arbitrary $h \in \mathcal{H}(x,r)$. We need to show that

$$h \in \mathcal{H}(x,\infty). \tag{7}$$

Applying Lemma 2.1(i), obtain that there exists $\lambda \geq 0$ such that

$$0 \in \nabla f(x) + \lambda h + \mathcal{N}_X(x+h) , \tag{8}$$

$$x + h \in X, \tag{9}$$

$$\lambda(\|h\| - r) = 0, \|h\| \leq r. \tag{10}$$

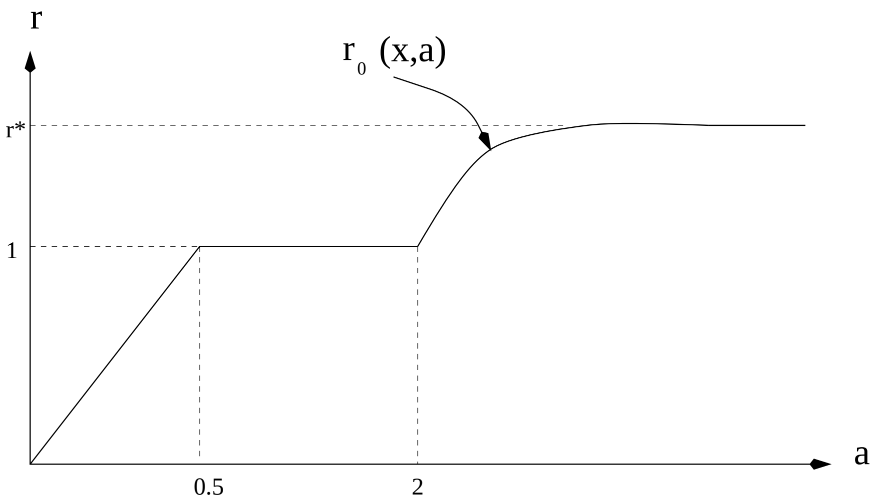

Figure 1: $r_0(x,a)$: $a_0(x,r)$ is essentially the inverse of this function

If $\lambda = 0$ then (7) follows immediately from (8), (9) and Lemma 2.1(ii). Assume therefore that $\lambda > 0$. Then (10) implies $\|h\| = r$. By Lemma 2.3(i) obtain $h = p_0(x, \lambda^{-1})$. Therefore, $r = \|p_0(x, \lambda^{-1})\| \leq r_0(x, \infty)$. Hence, by (6) and Lemma 2.5(ii) we have

$$\forall a \geq \lambda^{-1} > 0, \; r_0(x,a) = r_0(x, \infty).$$

Now applying Lemma 2.7 we obtain (7).

$\square$

We now define a function $a_0(x,r)$ that acts like the inverse of the function $r_0(x,a)$, see Figure 1. The inverse is generally set valued (see $r = 1$); hence for every $x \in X$ and $r \in [0, \infty)$ define

$$a_0(x,r) := \min \{a \geq 0 \mid r_0(x,a) \geq r\},$$

where we formally set $\min \emptyset = \infty$. Note that this is important in the example of Figure 1 when $r > r^*$. Furthermore, in that example, note that $r_0(x,a) = 1$ for any $a \in [1/2, 2]$ but that $a_0(x,1) = 1/2$ is well defined. The main properties of the function $a_0(x, \cdot) : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ are presented in the following lemma.

Lemma 2.9 *For every $x \in X$ the following properties hold.*

i. For all $r' \geq r \geq 0$, $a_0(x,r') \geq a_0(x,r) \geq a_0(x,0) = 0$.

ii. $\forall r \in (0,\infty)$, $a_0(x,r) > 0$ *if* $x \notin X_{\text{stat}}$, *and* $a_0(x,r) = \infty$ *if* $x \in X_{\text{stat}}$.

iii. $\forall r \in (0,\infty), a_0(x,r) < \infty \iff \mathcal{H}(x,\infty) = \emptyset$.

iv. For all $r \in [0,\infty)$, $r_0(x, a_0(x,r)) \le r$.

Proof i. Follows from Lemma 2.5(ii).

ii. Follows from Lemma 2.5(iv).

iii. Follows from Lemma 2.8(i).

iv. Obviously, for every $x \in X$ and $r > 0$ we have the following: if there exists $a' \geq 0$ such that $r \le r_0(x,a')$ then

$$r_0(x, a_0(x,r)) = r; \tag{11}$$

Otherwise, $r \ge r_0(x,\infty)$, with $r_0(x,\infty) < \infty$ and $a_0(x,r) = \infty$.

□

We now present an important relationship that exists between the problems $\mathcal{TP}(x,r)$ and $\mathcal{QP}(x,a)$. Essentially, it states that the solution of $\mathcal{QP}(x,a)$ is also a solution of $\mathcal{TP}(x,r)$.

Theorem 2.10 *For every* $x \in X$ *and* $r \ge 0$, $p_0(x, a_0(x,r)) \in \mathcal{H}(x,r)$.

Proof Let $x \in X$ and $r > 0$ be arbitrary. We can suppose that $x \notin X_{\text{stat}}$, since otherwise $0 = p_0(x, a_0(x,r)) \in \mathcal{H}(x,r)$. It follows that $a_0(x,r) > 0$ and we consider two possible cases.

In the first case, suppose that there exists $a' > 0$ such that $r \le r_0(x,a')$. Then $a_0(x,r) < \infty$, $r_0(x, a_0(x,r)) = r$ by (11) and so $\|p_0(x, a_0(x,r))\| = r$. Also, by Lemma 2.3(i) we have

$$0 \in \nabla f(x) + \tfrac{1}{a_0(x,r)} p_0(x, a_0(x,r)) + \mathcal{N}_X\left(x + p_0(x, a_0(x,r))\right),$$

$$x + p_0(x, a_0(x,r)) \in X.$$

Applying Lemma 2.1(i) we obtain $p_0(x, a_0(x,r)) \in \mathcal{H}(x,r)$.

Otherwise, $r > r_0(x,a')$ for all $a' > 0$. It follows that $r \ge r_0(x,\infty)$, $r_0(x,\infty) < \infty$, $\mathcal{H}(x,\infty) \ne \emptyset$, and $a_0(x,r) = \infty$. Now, from Lemma 2.8(ii) we have

$$p_0(x, a_0(x,r)) = p_0(x,\infty) \in \mathcal{H}(x,r),$$

completing the proof of the theorem. □

The following result that provides a useful bound on the optimal value of $\mathcal{TP}(x,r)$ is now easily established using Lemma 2.3 and Lemma 2.9(iv).

Corollary 2.11 *For every $x \in X \setminus X_{\text{stat}}$ and $r > 0$ the following bound holds:*

$$v(x,r) = \min_{x+h \in X, \|h\| \leq r} \langle \nabla f(x), h \rangle \leq -\frac{1}{a_0(x,r)} r^2,$$

where we formally set $\frac{1}{\infty} := 0$.

3 The Sequential Linearization Method

Let $\|\cdot\|_*$ be an arbitrary norm in $\mathbb{R}^n$. It is well known that there are positive constants c_1 and c_2 such that for all $x \in \mathbb{R}^n$

$$c_1 \|x\|_* \leq \|x\| \leq c_2 \|x\|_* .$$

Note that for every $x \in X$ and $r > 0$, the following inclusions hold:

$$\begin{aligned} \{h \mid x+h \in X, \|h\| \leq c_1 r\} &\subseteq \{h \mid x+h \in X, \|h\|_* \leq r\} \\ &\subseteq \{h \mid x+h \in X, \|h\| \leq c_2 r\}. \end{aligned} \tag{12}$$

For $x \in X$ and $r > 0$ consider the following problem:

$$\mathcal{LP}(x,r): \quad \begin{array}{ll} \min_h & \langle \nabla f(x), h \rangle \\ \text{subject to} & x+h \in X, \\ & \|h\|_* \leq r. \end{array}$$

Clearly, when $\|\cdot\|_*$ is a polyhedral norm and X is a polyhedral set, $\mathcal{LP}(x,r)$ is a linear programming problem. Denote by $\mathcal{H}_*(x,r)$ and $v_*(x,r)$ the solution set and the optimal value of $\mathcal{LP}(x,r)$ respectively. The inclusions (12) imply that

$$v(x, c_2 r) \leq v_*(x,r) \leq v(x, c_1 r) \tag{13}$$

for every $x \in X$ and $r \geq 0$. The next result follows immediately from Corollary 2.2 and (13).

Lemma 3.1 *For every $x \in X$ and $r > 0$, the following assertion holds:*

$$x \in X_{\text{stat}} \iff v_*(x,r) = 0.$$

In order to justify the technique that we use to update the trust region radius, we will also need a simple technical result.

Lemma 3.2 *For every $\delta > 0$ and $x \in X \setminus X_{\text{stat}}$, there exist $\underline{r}$, $\overline{r}$ with $0 < \underline{r} < \overline{r}$, such that the following properties hold.*

i. $\forall r \in [\overline{r}, \infty), \forall h \in \mathcal{H}_*(x, r),\ f(x) - f(x+h) \leq \frac{\delta}{2} r^2$;

ii. $\forall r \in [0, \underline{r}], \forall h \in \mathcal{H}_*(x, r),\ f(x) - f(x+h) \geq \frac{\delta}{2} r^2$.

Proof Let $\delta > 0$ and $x \in X \setminus X_{\text{stat}}$.

i. Define $\overline{r} := \sqrt{2\delta^{-1}(f(x) - f_{\text{opt}})}$, so that for all $r \in [\overline{r}, \infty)$

$$f(x) - f(x+h) \leq f(x) - f_{\text{opt}} = \frac{\delta}{2}\overline{r}^2 \leq \frac{\delta}{2} r^2.$$

ii. Let

$$\underline{a} := 2c_1^2(Lc_2^2 + \delta)^{-1}, \tag{14}$$

so that by Lemma 2.5, there exists $\underline{r} > 0$ with $a_0\left(x, c_1\underline{r}\right) \leq \underline{a}$. Take an arbitrary $r \in (0, \underline{r}]$ and $h \in \mathcal{H}_*(x, r)$. By Lemma 2.9(i)

$$a_0\left(x, c_1 r\right) \leq a_0\left(x, c_1\underline{r}\right) \leq \underline{a},$$

then Corollary 2.11 implies

$$v(x, c_1 r) \leq -\frac{1}{a_0\left(x, c_1 r\right)}(c_1 r)^2 \leq -\frac{c_1^2}{\underline{a}} r^2. \tag{15}$$

It follows from (13) and (15) that

$$\langle \nabla f(x), h \rangle = v_*(x, r) \leq v(x, c_1 r) \leq -\frac{c_1^2}{\underline{a}} r^2.$$

Therefore, by (14)

$$\begin{aligned}
f(x+h) - f(x) &\leq \langle \nabla f(x), h \rangle + \frac{L}{2}\|h\|^2 \\
&\leq -\frac{c_1^2}{\underline{a}} r^2 + \frac{Lc_2^2}{2}\|h\|_*^2 \\
&\leq -\frac{c_1^2}{\underline{a}} r^2 + \frac{Lc_2^2}{2} r^2 \\
&= -\frac{1}{2}\left(\frac{2c_1^2}{\underline{a}} - Lc_2^2\right) r^2 \\
&= -\frac{\delta}{2} r^2.
\end{aligned}$$

$\square$

The algorithm we propose will have an Armijo style component to it. To describe this component precisely, let us fix arbitrary parameters θ and r_* with $0 < \theta < 1$ and $r_* > 0$, and define

$$\mathcal{R} := \left\{ r > 0 \mid r = \theta^k r_*,\ k = 0, \pm 1, \pm 2, \ldots \right\}.$$

The next result now follows immediately from Lemma 3.1 and Lemma 3.2 and shows there is an $r \in \mathcal{R}$ that satisfies the Armijo test given below in Algorithm SLP, but that r/θ does not.

Corollary 3.3 *Let $x \in X$ and for each $r' \in \mathcal{R}$, let $h_*(x, r') \in \mathcal{H}(x, r')$. The following assertions hold.*

i. If $x \in X_{\text{stat}}$, then

$$\forall r' \in \mathcal{R},\ v_*(x, r') = \left\langle \nabla f(x), h_*(x, r') \right\rangle = 0.$$

ii. If $x \notin X_{\text{stat}}$, then

$$\forall r' \in \mathcal{R},\ v_*(x, r') = \left\langle \nabla f(x), h_*(x, r') \right\rangle < 0,$$

and moreover, there exists $r \in \mathcal{R}$ such that

$$f(x) - f(x + h_*(x, \frac{r}{\theta})) < \frac{\delta}{2} \left(\frac{r}{\theta} \right)^2$$

and

$$f(x) - f(x + h_*(x, r)) \geq \frac{\delta}{2} r^2.$$

We now state the main algorithm. The parameter k counts the number of iterations, while μ_k represents the number of subproblems that we have solved at the kth iteration.

Algorithm SLP

Parameters δ, $r_* > 0$, $0 < \theta < 1$.

Input $x^1 \in X$.

Step 0 (Initialization): Set $k := 0$, $r_k := r_*$, $k_{\text{stop}} := \infty$.

Step 1 Set $k := k+1$, $\mu := 1$, $r := r_{k-1}$. Solve the problem $\mathcal{LP}(x^k, r)$ to obtain a solution $h_*(x^k, r) \in \mathcal{H}(x^k, r)$. Set

$$\nu := \begin{cases} 0 & \text{if } v_*(x^k, r) = \left\langle \nabla f(x^k), h_*(x^k, r) \right\rangle = 0; \\ -1 & \text{if } v_*(x^k, r) \neq 0 \text{ and } f(x^k) - f(x^k + h_*(x^k, r)) \geq \frac{\delta}{2} r^2; \\ 1 & \text{if } v_*(x^k, r) \neq 0 \text{ and } f(x^k) - f(x^k + h_*(x^k, r)) < \frac{\delta}{2} r^2; \end{cases}$$

If $\nu = 0$, set $k_{\text{stop}} = k$, $x^{k_{\text{stop}}} = x^k$ and exit.

Step 2 Set $\mu := \mu + 1$ and $r := r\theta^\nu$. Solve the problem $\mathcal{LP}(x^k, r)$ to obtain a solution $h_*(x^k, r) \in \mathcal{H}(x^k, r)$.

Step 3 If $\nu = -1$ and $f(x^k) - f(x^k + h_*(x^k, r)) < \frac{\delta}{2} r^2$ or $\nu = 1$ and $f(x^k) - f(x^k + h_*(x^k, r)) \geq \frac{\delta}{2} r^2$ then go to Step 4, else go to Step 2.

Step 4 If $\nu = -1$, set $r_k = \theta r$, else $r_k = r$. Let $\mu_k = \mu$ and $x^{k+1} = x^k + h_*(x^k, r_k)$ and go to Step 1.

Note that in Step 1, the parameter ν can take on one of three values. If $\nu = 0$, a stationary point has been found and the algorithm terminates. If $\nu = 1$, then sufficient decrease was not achieved and thus the subproblem needs to be resolved with a smaller value of r. If $\nu = -1$, then sufficient decrease was achieved, but the subproblem is solved again, this time with a larger value of r. Both of these resolves take place in Step 2.

If the algorithm terminates after a finite number of iterations, we can define the iterates $x^{k_{\text{stop}}+1}, x^{k_{\text{stop}}+2}, \ldots$ as equal to $x^{k_{\text{stop}}}$ for the purposes of stating convergence results. Furthermore, it follows from Corollary 3.3 that for every sequence of iterates $\{x^k\}$ generated by the algorithm, $\mu_k < \infty$, $k = 1, 2, \ldots$, and hence even if $k_{\text{stop}} = \infty$, $r_k > 0$, $k = 1, 2, \ldots$. Therefore, the algorithm is well defined. We now proceed to prove the convergence and associated rates as outlined in the introduction.

4 General Case: Convergence Theory

It is easily seen that every infinite sequence of iterates $\{x^k\}$ of the algorithm has the following properties:

$$f(x^k) - f(x^k + h_*(x^k, \frac{r_k}{\theta})) < \frac{\delta}{2} \left(\frac{r_k}{\theta} \right)^2, \tag{16}$$

$$f(x^k) - f(x^{k+1}) = f(x^k) - f(x^k + h_*(x^k, r_k)) \geq \frac{\delta}{2} r_k^2, \tag{17}$$

$$r_k > 0,\ k = 1, 2, \ldots .$$

We also note that

$$\left\| x^k - x^{k+1} \right\| = \left\| h_*(x^k, r_k) \right\| \leq c_2 \left\| h_*(x^k, r_k) \right\|_* \leq c_2 r_k.$$

Hence, for $k = 1, 2, \ldots$

$$f(x^k) - f(x^{k+1}) \geq \frac{\delta}{2c_2^2} \left\| x^k - x^{k+1} \right\|^2 . \tag{18}$$

To analyze convergence properties of this sequence, we need the following lemma.

Lemma 4.1 *Suppose that for some* $x \in X \setminus X_{\text{stat}}$, $r > 0$ *and* $h_*(x, \frac{r}{\theta}) \in \mathcal{H}_*(x, \frac{r}{\theta})$ *the following condition holds:*

$$f(x) - f(x + h_*(x, \frac{r}{\theta})) < \frac{\delta}{2} \left(\frac{r}{\theta} \right)^2 . \tag{19}$$

Then

$$v_*(x, r/\theta) \geq -\frac{Lc_2^2 + \delta}{2\theta^2} r^2; \tag{20}$$

$$a_0\,(x, r) \geq a_0\,(x, \tilde{c}_1 r) \geq \underline{a} := \frac{2\theta^2 \tilde{c}_1{}^2}{Lc_2^2 + \delta}; \tag{21}$$

$$r \geq \underline{\tilde{a}} r_0\,(x, 1) = \underline{\tilde{a}} \left\| x - \pi_X\,(x - \nabla f(x)) \right\| , \tag{22}$$

where $\tilde{c}_1 := \min(1, c_1)$, $\underline{\tilde{a}} = \min(1, \underline{a})$.

Proof It follows from (19) that

$$\begin{aligned} -\frac{\delta}{2} \left(\frac{r}{\theta} \right)^2 &< f(x + h_*(x, \frac{r}{\theta})) - f(x) \\ &\leq \langle \nabla f(x), h_*(x, r/\theta) \rangle + \frac{L}{2} \left\| h_*(x, r/\theta) \right\|^2 \\ &\leq \langle \nabla f(x), h_*(x, r/\theta) \rangle + \frac{Lc_2^2}{2} \left(r/\theta \right)^2 . \end{aligned}$$

Hence

$$v_*(x, r/\theta) = \langle \nabla f(x), h_*(x, r/\theta) \rangle \geq -\frac{Lc_2^2 + \delta}{2\theta^2} r^2,$$

and so (20) holds.

If $a_0(x,\tilde{c_1}r)=\infty$ then $a_0(x,r)=\infty$ and

$$r \geq \tilde{c_1}r \geq r_0(x,1) \geq \underline{\tilde{a}}r_0(x,1).$$

Thus, (21) and (22) are valid.

Suppose that $a_0(x,\tilde{c_1}r)<\infty$. We also note from Lemma 2.9(ii) that $0<a_0(x,\tilde{c_1}r)$. By Theorem 2.10 $p_0(x,a_0(x,\tilde{c_1}r)) \in \mathcal{H}(x,\tilde{c_1}r)$; therefore

$$\langle \nabla f(x), p_0(x,a_0(x,\tilde{c_1}r))\rangle = v(x,\tilde{c_1}r) \geq v(x,c_1r).$$

Applying (13) and (20) we obtain

$$\langle \nabla f(x), p_0(x,a_0(x,\tilde{c_1}r))\rangle \geq v(x,c_1r) \geq v_*(x,r) \geq v_*(x,r/\theta) \geq -\frac{Lc_2^2+\delta}{2\theta^2}r^2.$$

Now by Lemma 2.3(ii), we have

$$\begin{aligned} -\frac{Lc_2^2+\delta}{2\theta^2}r^2 &\leq \langle \nabla f(x), p_0(x,a_0(x,\tilde{c_1}r))\rangle \\ &\leq -\frac{1}{a_0(x,\tilde{c_1}r)}\|p_0(x,a_0(x,\tilde{c_1}r))\|^2 \\ &= -\frac{1}{a_0(x,\tilde{c_1}r)}r_0(x,a_0(x,\tilde{c_1}r))^2 \\ &\leq -\frac{1}{a_0(x,\tilde{c_1}r)}\tilde{c_1}^2r^2, \end{aligned}$$

the last inequality following from Lemma 2.9(iv). Hence, by Lemma 2.9(i)

$$a_0(x,r) \geq a_0(x,\tilde{c_1}r) \geq \frac{2\theta^2\tilde{c_1}^2}{Lc_2^2+\delta} = \underline{a}$$

and (21) is proved.

Lemma 2.9(iv) and Lemma 2.5(ii) imply

$$r \geq r_0(x,a_0(x,r)) \geq r_0(x,\underline{a}),$$

so applying Lemma 2.5(ii) and (iii)

$$r_0(x,\underline{a}) \geq \begin{cases} \underline{a}r_0(x,1) & \text{if } \underline{a} \leq 1; \\ r_0(x,1) & \text{if } \underline{a} \geq 1. \end{cases}$$

Thus (22) is proved. $\square$

Note that the following convergence theorem does not require any assumptions apart from the solvability of $\mathcal{P}$ and the convexity of X.

Theorem 4.2 *Every sequence of iterates $\{x^k\}$ produced by SLP has the property that*

$$r_0\left(x^k, 1\right) = \left\|x^k - \pi_X\left(x^k - \nabla f(x^k)\right)\right\| \to 0.$$

Moreover, if X is compact, then every sequence of iterates converges to X_{stat}.

Proof It follows from (16), (17) and Lemma 4.1 that for every sequence of iterates produced by SLP

$$f(x^k) - f(x^{k+1}) \geq \frac{\delta}{2}\underline{\tilde{a}}^2 \left\|x^k - \pi_X\left(x^k - \nabla f(x^k)\right)\right\|^2, \tag{23}$$

for $k = 1, 2, \ldots$. However, $\{f(x^k)\}$ is bounded below, hence converges, implying the first statement of the theorem. If the second statement is not valid, then there is a subsequence $\{x^k\}_{k\in\mathcal{K}}$ and an $\epsilon > 0$ such that $\operatorname{dist}\left(x^k, X_{\text{stat}}\right) > \epsilon$ for $k \in \mathcal{K}$. Since $\{x^k\}$ is bounded, $\{x^k\}_{k\in\mathcal{K}}$ has a limit point $\hat{x}$ for which $\operatorname{dist}(\hat{x}, X_{\text{stat}}) \geq \epsilon > 0$. But this is a contradiction to the first statement of the theorem. $\square$

Our convergence analysis is based in part of that given by Luo and Tseng [20]. The following assumptions are part of that work and will be used in our main convergence result.

Assumption A1. For every $\nu \geq f_{\text{opt}}$ there exist scalars $\kappa > 0$ and $\eta > 0$ such that

$$\operatorname{dist}(x, X_{\text{stat}}) \leq \kappa \left\|x - \pi_X\left(x - \nabla f(x)\right)\right\|,$$

for all $x \in X$ with $f(x) \leq \nu$ and $\|x - \pi_X(x - \nabla f(x))\| \leq \eta$.

Assumption A2. For every $\nu \geq f_{\text{opt}}$ the set

$$f(\{x \in X_{\text{stat}} \mid f(x) \leq \nu\}) \subset \mathbb{R}^1$$

is finite.

Theorem 2.1 of [20] gives sufficient conditions to guarantee that the above assumptions hold.

To obtain more precise convergence properties of SLP, we need the following lemma. This result essentially provides a local error bound.

Lemma 4.3 *Let Assumption A1 hold. Then for every $v > f_{\rm opt}$ and $\underline{a} > 0$ there exists $\Psi_1 = \Psi_1(v, \underline{a}, \kappa, L) > 0$ such that for all $r \geq r_0(x, \underline{a})$,*

$$f(x + p_0(x, a_0(x, r))) - f(\pi_{X_{\rm stat}}(x)) \leq \Psi_1 r^2$$

for all $x \in X$ with $f(x) \leq \nu$ and $r_0(x, 1) = \|x - \pi_X(x - \nabla f(x))\| \leq \eta$.

Proof Let $v > f_{\rm opt}$ and $\underline{a}$ be arbitrary, and consider any $x \in X$ with $f(x) \leq v$, $r_0(x, 1) \leq \eta$, and $r > r_0(x, \underline{a})$. Fix an arbitrary

$$\bar{x} = \pi_{X_{\rm stat}}(x) \in \Pi_{X_{\rm stat}}(x).$$

Note that $X_{\rm stat}$ may not be convex. It follows from A1 that

$$\|x - \bar{x}\| = \text{dist}(x, X_{\rm stat}) \leq \kappa r_0(x, 1),$$

hence, by Lemma 2.5(iii)

$$\|x - \bar{x}\| \leq \tilde{\kappa} r_0(x, a_0(x, r)) = \tilde{\kappa} \|p_0(x, a_0(x, r))\|, \tag{24}$$

where $\tilde{\kappa} = \kappa \max(1, \underline{a}^{-1})$. Obviously, if $x \in X_{\rm stat}$, then $\bar{x} = x$ and

$$x + p_0(x, a_0(x, r)) = x = \bar{x},$$

and the desired bound holds.

Suppose therefore that $x \notin X_{\rm stat}$, then $a_0(x, r) > 0$ (it is possible that $a_0(x, r) = \infty$). For simplicity, denote $p_0(x, a_0(x, r))$ by $\bar{p}_0$. Note that by the definition of $a_0(x, r)$ and Lemma 2.8

$$\|\bar{p}_0\| = r_0(x, a_0(x, r)) \begin{cases} = r & \text{if } r \leq r_0(x, \infty), \\ < r & \text{if } r > r_0(x, \infty). \end{cases}$$

Thus,

$$\|\bar{p}_0\| \leq r. \tag{25}$$

Let us show that

$$\langle \nabla f(x), x + \bar{p}_0 - \bar{x} \rangle \leq \frac{1}{2\underline{a}} \|x - \bar{x}\|^2. \tag{26}$$

If $a_0(x, r) = \infty$, then $\bar{p}_0 \in \mathcal{H}(x, \infty)$ and

$$\langle \nabla f(x), \bar{p}_0 \rangle \leq \langle \nabla f(x), \bar{x} - x \rangle,$$

which implies (26).

If $a_0(x,r) < \infty$, then by the definition of $p_0(x, a_0(x,r))$ we have

$$\langle \nabla f(x), \bar{p}_0 \rangle + \frac{1}{2a_0(x,r)} \|\bar{p}_0\|^2 \le \langle \nabla f(x), \bar{x} - x \rangle + \frac{1}{2a_0(x,r)} \|\bar{x} - x\|^2 ;$$

hence

$$\langle \nabla f(x), x + \bar{p}_0 - \bar{x} \rangle \le \frac{1}{2\underline{a}} \|\bar{x} - x\|^2 .$$

It follows from (26), (24) and (25) that

$$\langle \nabla f(x), x + \bar{p}_0 - \bar{x} \rangle \le \frac{\tilde{\kappa}^2}{2\underline{a}} r^2. \tag{27}$$

Applying the Mean Value Theorem we have

$$f(x + \bar{p}_0) - f(\bar{x}) = \langle \nabla f(\xi), x + \bar{p}_0 - \bar{x} \rangle \tag{28}$$

for some $\xi \in [\bar{x}, x + \bar{p}_0]$. Clearly,

$$\|\xi - x\| \le \|\bar{x} = x\| + \|\bar{p}_0\| \le \tilde{\kappa}_1 r, \tag{29}$$

where $\tilde{\kappa}_1 := \tilde{\kappa} + 1$. It now follows from (28) that

$$\begin{aligned} f(x + \bar{p}_0) - f(\bar{x}) &= \langle \nabla f(\xi), x + \bar{p}_0 - \bar{x} \rangle \\ &= \langle \nabla f(\xi) - \nabla f(x), x + \bar{p}_0 - \bar{x} \rangle + \langle \nabla f(x), x + \bar{p}_0 - \bar{x} \rangle \\ &\le L \|x - \xi\| (\|x - \bar{x}\| + \|\bar{p}_0\|) + \langle \nabla f(x), x + \bar{p}_0 - \bar{x} \rangle . \end{aligned}$$

Using the relations (27) and (29), we obtain

$$f(x + \bar{p}_0) - f(\bar{x}) \le (L\tilde{\kappa}_1^2 + \tilde{\kappa}^2/2\underline{a}) r^2,$$

as required. $\square$

Lemma 4.4 *Let A1 hold. Then for every $v > f_{\rm opt}$ there exists $\Psi = \Psi(v, \underline{a}, \kappa, L) > 0$ such that if $x \in X \setminus X_{\rm stat}$ and $r > 0$ satisfies $f(x) \le v$, $r_0(x,1) \le \eta$ and (19), then*

$$\Psi r^2 \ge f(x) - f(\pi_{X_{\rm stat}}(x)).$$

Proof Lemma 4.1 and (19) imply that

$$v_*(x, r/\theta) \ge -\frac{Lc_2^2 + \delta}{2\theta^2} r^2,$$

so that Theorem 2.10 shows

$$p_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right) \in \mathcal{H}(x, \tilde{c}_1 r).$$

Applying (13) we obtain

$$\begin{aligned} \langle \nabla f(x), p_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right)\rangle &= v(x, \tilde{c}_1 r) \geq v(x, c_1 r) \\ &\geq v_*(x, r) \geq v_*(x, r/\theta) \geq -\frac{Lc_2^2+\delta}{2\theta^2} r^2. \end{aligned}$$

Therefore

$$\begin{aligned} & f(x + p_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right)) - f(x) \\ \geq\ & \langle \nabla f(x), p_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right)\rangle - \frac{L}{2}\left\|p_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right)\right\|^2 \\ \geq\ & -\frac{Lc_2^2+\delta}{2\theta^2} r^2 - \frac{L\tilde{c}_1{}^2}{2} r^2 \\ =\ & -\Psi_2 r^2, \end{aligned} \tag{30}$$

where

$$\Psi_2 := \frac{Lc_2^2 + \theta^2 L\tilde{c}_1{}^2 + \delta}{2\theta^2}.$$

By Lemma 4.1, $a_0\left(x, \tilde{c}_1 r\right) \geq \underline{a}$, so that Lemma 2.9(iv) implies

$$\tilde{c}_1 r \geq r_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right) \geq r_0\left(x, \underline{a}\right).$$

Applying Lemma 4.3 we see that

$$f(x + p_0\left(x, a_0\left(x, \tilde{c}_1 r\right)\right)) - f(\pi_{X_{\text{stat}}}(x)) \leq \Psi_1 r^2, \tag{31}$$

where $\Psi_1 = \Psi_1(v, \underline{a}, \kappa, L) > 0$. Combining (30) and (31) yields

$$f(x) - f(\pi_{X_{\text{stat}}}(x)) \leq (\Psi_1 + \Psi_2) r^2.$$

$\square$

Now the proof of the main convergence theorem is similar to that given by Luo and Tseng [20, Theorem 3.1] for the linear convergence of a class of descent techniques.

Theorem 4.5 *Let A1, A2 hold. Every infinite sequence of iterates produced by SLP converges to a point of* $x^* \in X_{\text{stat}}$. *Moreover, for every* $v \geq f_{\text{opt}}$, *there exist q_1, q_2 with $0 \le q_1, q_2 < 1$ such that for every infinite sequence $\{x^k\}$, $f(x^1) \le v$, the following estimations hold:*

$$\limsup_{k\to\infty} \frac{f(x^{k+1} - f(x^*)}{f(x^k) - f(x^*)} \le q_1;$$

$$\limsup_{k\to\infty} q_2^{-k} \mathrm{dist}\left(x^k, X_{\mathrm{stat}}\right) < \infty.$$

Every finite sequence of iterates $\{x^k\}$ of SLP terminates in the stationary set, $x^{k_{\mathrm{stop}}} \in X_{\mathrm{stat}}$.

Proof Let $v > f_{\mathrm{opt}}$ be chosen arbitrarily and consider the sequence of iterates $\{x^k\}$ of SLP with $f(x^1) \le v$. If $k_{\mathrm{stop}} < \infty$ then $v_+(x^{k_{\mathrm{stop}}}, r_{k_{\mathrm{stop}}}) = 0$ and Lemma 3.1 gives $x^{k_{\mathrm{stop}}} \in X_{\mathrm{stat}}$.

Suppose therefore that $k_{\mathrm{stop}} = \infty$. It is easily seen that for $k = 1, 2, \ldots$

$$-\infty < f_{\mathrm{opt}} \le f(x^{k+1}) \le f(x^k) \le v. \tag{32}$$

Hence, there exists $f_* \in [f_{\mathrm{opt}}, v]$ such that

$$\lim_{k\to\infty} f(x^k) = f_*. \tag{33}$$

It follows from (23) and (33) that

$$\lim_{k\to\infty} \left(f(x^k) = f(x^{k+1})\right) = \lim_{k\to\infty} r_0\left(x^k, 1\right) = 0, \tag{34}$$

and hence that for some $K_1 > 0$

$$r_0\left(x^k, 1\right) = \left\|x^k - \pi_x\left(x^k - \nabla f(x^k)\right)\right\| \le \eta,\ k = K_1, K_1 + 1, \ldots. \tag{35}$$

Since X_{stat} may not be convex, let us denote an arbitrary projection of x^k onto X_{stat} by $\bar{x}^k$, that is

$$\bar{x}^k = \pi_{X_{\mathrm{stat}}}\left(x^k\right) \in \Pi_{X_{\mathrm{stat}}}\left(x^k\right),\ k = 1, 2, \ldots.$$

Using (32) and (35), the assumption A1 implies that for $k = K_1, K_1 + 1, \ldots$

$$\kappa r_0\left(x^k, 1\right) \ge \mathrm{dist}\left(x^k, X_{\mathrm{stat}}\right) = \left\|x^k - \bar{x}^k\right\|. \tag{36}$$

Therefore, by (34)

$$\lim_{k\to\infty} \mathrm{dist}\left(x^k, X_{\mathrm{stat}}\right) = \lim_{k\to\infty} \left\|x^k - \bar{x}^k\right\| = 0. \tag{37}$$

It follows from (18) and (34) that

$$\lim_{k\to\infty}\left\|x^k - x^{k+1}\right\| = 0,$$

which combined with (37) yields

$$\lim_{k\to\infty}\left\|\bar{x}^k - \bar{x}^{k+1}\right\| = 0. \tag{38}$$

The assumption A2 postulates that there is some $f_{\text{stat}} \in \mathbb{R}^1$ and $K_2 \geq K_1$ such that

$$f(\bar{x}^k) = f_{\text{stat}} \in f(X_{\text{stat}}),\ k = K_2, K_2+1, \ldots. \tag{39}$$

The Mean Value Theorem implies that for $k \geq K_2$

$$f(\bar{x}^k) - f(x^k) = \left\langle \nabla f(\xi^k), \bar{x}^k - x^k \right\rangle$$

for some ξ^k lying on the line segment joining $\bar{x}^k$ and x^k. Hence

$$f_{\text{stat}} - f(x^k) = \left\langle \nabla f(\xi^k) - \nabla f(\bar{x}^k), \bar{x}^k - x^k \right\rangle + \left\langle \nabla f(\bar{x}^k), \bar{x}^k - x^k \right\rangle.$$

As $\bar{x}^k \in X_{\text{stat}}$, it follows that $\left\langle \nabla f(\bar{x}^k), \bar{x}^k - x^k \right\rangle \leq 0$, so that

$$f_{\text{stat}} - f(x^k) \leq L\left\|\xi^k - \bar{x}^k\right\|\left\|\bar{x}^k - x^k\right\| \leq L\left\|\bar{x}^k - x^k\right\|^2.$$

Therefore, by (37), $f_{\text{stat}} \leq \lim_{k\to\infty} f(x^k) = f_*$. Using (16), (32), (35), (39) and Lemma 4.4 we have

$$\Psi r_k^2 \geq f(x^k) - f_{\text{stat}},\ k = K_2, K_2+1, \ldots$$

so that using (17)

$$f(x^k) - f(x^{k+1}) \geq \frac{\delta}{2\Psi}(f(x^k) - f_{\text{stat}}),\ k = K_2, K_2+1, \ldots;$$

and

$$0 \leq f(x^{k+1}) - f_{\text{stat}} \leq q_1(f(x^k) - f_{\text{stat}}),\ k = K_2, K_2+1, \ldots,$$

where $q_1 = \max(0, 1 - \delta/2\Psi)$, $0 \leq q_1 < 1$. Thus

$$\lim_{k\to\infty} f(x^k) = f_* = f_{\text{stat}}, \tag{40}$$

and, moreover,

$$\frac{f(x^{k+1} - f_{\text{stat}}}{f(x^k) - f_{\text{stat}}} \le q_1, k = K_2, K_2 + 1, \ldots. \tag{41}$$

Now it follows from (18) that there exists $c_0 > 0$ such that

$$\left\| x^k - x^{k+1} \right\| \le c_0 \sqrt{q_1}^k, \; k = K_2, K_2 + 1, \ldots.$$

Hence, $\sum_{k=1}^{\infty} \left\| x^k - x^{k+1} \right\| < \infty$ and there exists an x^* such that

$$\lim_{k \to \infty} x^k = x^* \in X.$$

Moreover, by (37), $x^* \in X_{\text{stat}}$. It follows from (17) and (40) that

$$f(x^k) - f_{\text{stat}} \ge f(x^k) - f(x^{k+1}) \ge \frac{\delta}{2} r_k^2, \; k = 1, 2, \ldots$$

and by (16), (32), (35), Lemma 4.1 and A1 that

$$f(x^k) - f_{\text{stat}} \ge \frac{\delta \tilde{a}^2}{2\kappa^2} \text{dist} \left(x^k, X_{\text{stat}} \right)^2, \; k = K_2, K_2 + 1, \ldots.$$

Taking into account (41) we obtain that there exist $K' > 0$ and q_2 with $0 \le q_2 < 1$, such that

$$\text{dist} \left(x^k, X_{\text{stat}} \right)^2 \le K' q_2^k \; k = K_2, K_2 + 1, \ldots.$$

□

5 Special Case: Polyhedral Constraints

In the case where X is defined by linear equalities and inequalities, the problem $\mathcal{LP}(x, r)$ is a linear program, and hence the algorithm SLP is a successive linear programming algorithm.

We have implemented SLP in Matlab [23], using the Cplex [11] linear programming code for the trust region subproblems $\mathcal{LP}(x, r)$. Using the interface between Cute [4] and Matlab described in [6], we have tested the algorithm on a reasonable class of linearly constrained test problems. The results are summarized in Tables 1 and 2, which arbitrarily separates the problems in the Hock-Schittkowski collection [18] from the remaining ones we tested from the Cute collection. In the tables, the number of general linear constraints is denoted by "m" and "n" is used to denote the number of variables. Lower and upper bounds are treated explicitly by the linear programming code and thus are not reported here. The computations were carried out on a Sun SPARCstation 10 with 96MB RAM, running Matlab version 4.2c. The time reported in the table was derived using the "cputime" function of Matlab. No attempt was made to optimize the SLP code to allow for fast restarts in the Armijo test and so this is generally a considerable over estimate of the time that would be required for a sophisticated implementation of the algorithm. Of more importance are the number of iterations and the total number of LP's that needed to be solved. Note that the number of LP's is guaranteed to be at least twice the number of iterations due to the Armijo test.

We have excluded the problems BOOTH, EXTRASIM, HIMMELBA, HS54, HS55, HYDROELM, HYDROELS, STANCMIN, SUPERSIM, and ZANGWIL3 from the tables of results for the sake of brevity, since in these examples a stationary point was found after one iteration with just two linear program solutions. Note that paradoxically more than one LP was required to solve some of the linear programs in the Cute collection. This is due to our default choice on the initial trust region radius.

The algorithm appears to be very robust. Since the LP solver is very fast, the resulting NLP code is also very efficient, even in the Matlab implementation.

Note however that the method will only give linear convergence and thus once an active set has been identified, the algorithm should apply a Newton process for solving the resulting reduced problem, such as the linear programming based technique given in [14]. Furthermore, conditions under which this method identifies the correct active set after a finite number of iterations could be derived in a similar fashion to that given by Burke and Moré [8]. Such a procedure would undoubtedly improve the performance of the algorithm on problems HS268 and S268. Similar finitely terminating successive linearization algorithms have been used for machine learning problems [5, 22]

References

[1] T. E. Baker and L. S. Lasdon. Successive linear programming at Exxon. *Management Science*, 31:264–274, 1985.

Problem	m	n	CPU	Iters	LPs
HS105	1	8	4.417	20	76
HS112	3	10	1.95	13	55
HS118	17	15	0.3333	2	10
HS119	8	16	1.75	12	40
HS21	1	2	0.1167	2	4
HS21MOD	1	7	0.3667	3	10
HS24	3	2	0.2167	2	6
HS268	5	5	59.55	600	1349
HS28	1	3	0.8333	6	25
HS35	1	3	1.333	13	35
HS35MOD	1	3	0.95	6	21
HS36	1	3	0.35	2	11
HS37	2	3	0.8667	8	23
HS41	1	4	0.85	9	21
HS44	6	4	0.5833	5	12
HS44NEW	6	4	0.6667	5	19
HS48	2	5	0.9667	8	25
HS49	2	5	3.15	22	56
HS50	3	5	2.15	17	51
HS51	3	5	0.9667	9	20
HS52	3	5	2.933	20	47
HS53	3	5	1.533	12	27
HS62	1	3	1.933	18	40
HS76	3	4	0.7833	8	21
HS86	10	5	0.9	8	20
HS9	1	2	0.6333	8	17

Table 1: SLP on polyhedrally constrained HS nonlinear programs

Problem	m	n	CPU	Iters	LPs
DEGENLPA	15	20	0.2	2	6
DEGENLPB	15	20	0.3667	2	8
GENHS28	8	10	1.4	14	42
GOULDQP2	349	699	5.217	2	14
GOULDQP3	349	699	36.43	25	58
HAGER1	10	21	1.883	16	34
HAGER2	10	21	1.3	12	25
HAGER3	10	21	1.55	13	27
HAGER4	10	21	0.6167	6	13
HATFLDH	7	4	0.25	2	8
HIMMELBI	12	100	3.333	23	53
HONG	1	4	0.7833	11	23
ODFITS	6	10	1.433	10	22
PENTAGON	15	6	0.15	2	5
QPCBLEND	74	83	2.467	8	22
QPCBOEI1	351	384	43.97	18	46
QPCBOEI2	166	143	15.38	19	50
QPCSTAIR	356	467	20.92	8	21
QPNBLEND	74	83	1.983	6	17
QPNBOEI1	351	384	104.2	32	74
QPNBOEI2	166	143	14.67	18	56
QPNSTAIR	356	467	22.22	8	21
READING2	200	303	4.167	7	15
S268	5	5	65.35	600	1349
S277-280	4	4	0.2333	2	6
SIMPLLPA	2	2	0.1667	2	4
SIMPLLPB	3	2	0.2167	2	5
SSEBLIN	72	194	2.067	2	17
TAME	1	2	0.15	2	4
TFI2	101	3	0.6333	4	10
TFI3	101	3	1.167	7	16

Table 2: SLP on remaining polyhedrally constrained nonlinear programs from Cute

[2] K. P. Bennett and O. L. Mangasarian. Bilinear separation of two sets in n-space. *Computational Optimization and Applications*, 2:207–227, 1993.

[3] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts, 1995.

[4] I. Bongartz, A. R. Conn, N. Gould, and Ph. L. Toint. CUTE: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software*, 21:123–160, 1995.

[5] P. S. Bradley, O. L. Mangasarian, and W. N. Street. Feature selection via mathematical programming. Mathematical Programming Technical Report 95-21, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1995.

[6] M. A. Branch. Getting CUTE with Matlab. Cornell Theory Center Report CTC94TR194, Cornell University, Ithaca, New York, 1994.

[7] J. V. Burke. A robust trust region method for constrained nonlinear programming problem. *SIAM Journal on Optimization*, 2:325–347, 1992.

[8] J. V. Burke and J. J. Moré. On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25:1197–1211, 1988.

[9] J. V. Burke, J. J. Moré, and G. Toraldo. Convergence properties of trust region methods for linear and convex constraints. *Mathematical Programming*, 47:305–336, 1990.

[10] P. H. Calamai and J. J. Moré. Projected gradient methods for linearly constrained problems. *Mathematical Programming*, 39:93–116, 1987.

[11] CPLEX Optimization Inc., Incline Village, Nevada. *Using the CPLEX(TM) Linear Optimizer and CPLEX(TM) Mixed Integer Optimizer (Version 2.0)*, 1992.

[12] J. C. Dunn. Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals. *SIAM Journal on Control and Optimization*, 17:187–211, 1979.

[13] J. C. Dunn and E. Sachs. The effect of perturbations on the convergence rates of optimization algorithms. *Applied Mathematics and Optimization*, 10:143–157, 1983.

[14] M. C. Ferris. Iterative linear programming solution of convex programs. *Journal of Optimization Theory and Applications*, 65:53–65, 1990.

[15] R. Fletcher and E. Sainz de la Maza. Nonlinear programming and nonsmooth optimization by successive linear programming. Numerical Analysis Report NA/100, Department of Mathematical Sciences, The University, Dundee, Scotland, 1987.

[16] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.

[17] E. M. Gafni and D. P. Bertsekas. Two-metric projection methods for constrained optimization. *SIAM Journal on Control and Optimization*, 22:936–964, 1984.

[18] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, Berlin, Germany, 1981.

[19] L. S. Lasdon, A. Waren, S. Sarkar, and F. Palacios. Solving the pooling problem using generalized reduced gradient and successive linear programming algorithms. *ACM SIGMAP Bulletin*, 27:9–15, 1979.

[20] Z.-Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: A general approach. *Annals of Operations Research*, 46:157–178, 1993.

[21] O. L. Mangasarian. *Nonlinear Programming*. McGraw–Hill, New York, 1969. SIAM Classics in Applied Mathematics 10, SIAM, Philadelphia, 1994.

[22] O. L. Mangasarian. Machine learning via polyhedral concave minimization. In H. Fischer, B. Riedmueller, and S. Schaeffler, editors, *Applied Mathematics and Parallel Computing - Festschrift for Klaus Ritter*, pages 175–188. Physica-Verlag A Springer-Verlag Company, Heidelberg, 1996.

[23] MATLAB. *User's Guide*. The MathWorks, Inc., 1992.

[24] F. Palacios-Gomez, L. Lasdon, and M. Engquist. Nonlinear optimization by successive linear programming. *Management Science*, 28:1106–1120, 1982.

[25] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

[26] T.-H. Shiau. *Iterative Linear Programming for Linear Complementarity and Related Problems*. PhD thesis, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1983. Technical Report 507.

[27] Ph. L. Toint. Global convergence of a class of trust region methods for nonconvex minimization in a Hilbert space. *IMA Journal of Numerical Analysis*, 8:231–252, 1988.

[28] J. Zhang, N.-H. Kim, and L. Lasdon. An improved successive linear programming algorithm. *Management Science*, 31:1312–1331, 1985.