ERROR BOUNDS FOR STRONGLY CONVEX PROGRAMS AND (SUPER)LINEARLY CONVERGENT ITERATIVE SCHEMES FOR THE LEAST 2-NORM SOLUTION OF LINEAR PROGRAMS

by

O. L. Mangasarian & R. De Leone

Computer Sciences Technical Report #631

February 1986

University of Wisconsin-Madison
Computer Sciences Department

Error Bounds for Strongly Convex Programs and (Super)Linearly Convergent Iterative Schemes for the Least 2-Norm Solution of Linear Programs

O. L. Mangasarian & R. De Leone[1)]

Technical Report #631

February 1986

ABSTRACT

Given an arbitrary point (x, u) in $R^n \times R^m_+$, we give bounds on the Euclidean distance between x and the unique solution $\bar{x}$ to a strongly convex program in terms of the violations of the Karush-Kuhn-Tucker conditions by the arbitrary point (x, u). These bounds are then used to derive linearly and superlinearly convergent iterative schemes for obtaining the unique least 2-norm solution of a linear program. These schemes can be used effectively in conjunction with the successive overrelaxation (SOR) methods for solving very large sparse linear programs.

AMS(MOS) Subject Classifications: 90C25, 90C05

Key Words: Convex programs, linear programs, error bounds, iterative methods

---

Sponsored by the United States Army under contract No. DAAG29-80-C-0041. This material is based on research sponsored by National Science Foundation Grant DCR-8420963 and Air Force Office of Scientific Research Grant AFOSR-ISSA-85-00080.

[1)] On leave from CRAI, via Bernini 5, Rende, Cosenza, Italy.

Error Bounds for Strongly Convex Programs and (Super)Linearly Convergent Iterative Schemes for the Least 2-Norm Solution of Linear Programs

O. L. Mangasarian & R. De Leone

1. Introduction

We consider the problem

$$\min_x \; f(x) \quad \text{subject to} \;\; x \in S := \{x | x \geq 0,\; g(x) \leq 0\} \tag{1.1}$$

where $f: R^n \to R$ and $g: R^n \to R^m$ are differentiable and convex functions on R^n, S is nonempty and in addition f is strongly convex on R^n, that is

$$(\nabla f(y) - \nabla f(x))(y-x) \geq k \left\|y - x\right\|_2^2 \tag{1.2}$$

for all x, y in R^n and some $k > 0$, where $\|\cdot\|_2$ denotes the 2-norm. It follows immediately that (1.1) has a unique solution $\bar{x}$ in S. Our purpose here is that given any x in R^n to obtain a bound on the distance $\|x - \bar{x}\|_2$, in terms of the violations of the Karush-Kuhn-Tucker conditions for (1.1) by x and any nonnegative u in R^m (Theorem 2.2), or by x and an "optimal" u chosen by solving a single linear program (Remark 2.6). The error bound (2.7) of Theorem 2.2, which is also a Lipschitz continuity result of order $\frac{1}{2}$ (see (2.13)), involves 3 parameters α, β, γ which may not be readily computable. In Theorem 2.5 we replace these parameters by corresponding upper bounds $\alpha(x_0)$, $\beta(\hat{x}, \hat{u})$, $\gamma(\hat{x}, \hat{u})$ which are readily computable from any primal feasible x_0 and any primal-dual feasible point $(\hat{x}, \hat{u})$ which satisfies the primal Slater constraint qualification. Related Lipschitz continuity results are given by Daniel in [3] for positive definite quadratic programs. Stronger local Lipschitz continuity results for more general programs are given by Robinson in [17,18].

In Section 3 of the paper we turn our attention to what motivated the paper originally, namely computing the least 2-norm solution of a linear program. Determination of the least 2-norm solution of a linear program has been the keystone of the successive overrelaxation (SOR) methods for solving very large sparse linear programs not solvable by standard

Sponsored by the United States Army under contract No. DAAG29-80-C-0041. This material is based on research sponsored by National Science Foundation Grant DCR-8420963 and Air Force Office of Scientific Research Grant AFOSR-ISSA-85-00080.

pivotal packages [9,10]. The first result of Section 3 is that the 2-norm $\|\hat{x}\|_2$ of any solution $\hat{x}$ of a linear program bounds the Euclidean distance $\|\hat{x}-\bar{x}\|_2$ between $\hat{x}$ and the least 2-norm solution of the linear program. This inequality, $\|\hat{x}-\bar{x}\|_2 \le \|\hat{x}\|_2$, which is obviously valid for any two points $\hat{x}$ and $\bar{x}$ in the nonnegative orthant R^n_+ if $\hat{x} \ge \bar{x}$, is not valid if we merely have $\|\hat{x}\|_2 \ge \|\bar{x}\|_2$ as can be seen from the simple example in R^2 of $\hat{x} = \begin{pmatrix}1\\0\end{pmatrix}$, $\bar{x} = \begin{pmatrix}0\\1\end{pmatrix}$ where $\|\bar{x}\|_2 = \|\hat{x}\|_2 < \|\hat{x}-\bar{x}\|_2$. Theorem 3.2 gives an improved bound on $\|\hat{x}-\bar{x}\|_2$ by solving a linear program. The final and computationally important results of this paper, contained in Theorems 3.7 and 3.8, are linearly and superlinearly convergent schemes for determining the least 2-norm solution of a linear program. We give the essence of these results. In solving very large sparse linear programs one solves by an SOR technique [8,9,10] a quadratic perturbation (3.3) of the linear program (3.1) for "sufficiently small" value ε of the perturbation parameter ε, that is $\varepsilon \in (0,\bar{\varepsilon}]$ for some $\bar{\varepsilon} > 0$. Until now there was no simple way of determining when $\varepsilon \le \bar{\varepsilon}$. Theorems 3.7 and 3.8 do this as follows. Given a value ε_i of the perturbation parameter, we approximately solve the quadratic perturbation problem (3.3) for $x(\varepsilon_i)$ by an SOR or any other procedure to a residual accuracy $r(\varepsilon_i)$ defined by (3.14). Then we decrease ε_i to $\varepsilon_{i+1} = \mu\varepsilon_i$, $\mu \in (0,1)$ and solve (3.3) to a residual accuracy $r(\varepsilon_{i+1})$ such that

(1.3) $$r(\varepsilon_{i+1}) \le \nu r(\varepsilon_i) \quad \text{for some} \quad \nu < \mu^{1/2} \quad \text{for linear convergence}$$

and

(1.4) $$r(\varepsilon_i) \le \xi \varepsilon_i^{1/2} \eta^{\rho^i} \quad \text{for some} \quad \xi > 0,\ \eta \in (0,1),\ \rho > 1 \quad \text{for superlinear convergence}$$

Theorem 3.7 shows that the sequence of approximate solutions $\{x(\varepsilon_i)\}$ thus generated converges to the unique least 2-norm solution of the linear program (3.1) at a linear rate under (1.3), while Theorem 3.8 establishes ρ-rate superlinear convergence under (1.4).

We briefly describe now our notation and some basic concepts used. For a vector x in the n-dimensional real space R^n, $|x|$ and x_+ will denote the vectors in R^n with components $|x|_i = |x_i|$ and $(x_+)_i = \max\{x_i, 0\}$, $i = 1,\dots,n$ respectively. For a norm $\|x\|_\beta$ on R^n, the dual norm $\|x\|_{\beta^*}$ on R^n will be defined by $\|x\|_{\beta^*} := \max_{\|y\|_\beta = 1} xy$, where xy denotes the scalar product. The generalized Cauchy-Schwarz inequality $|xy| \le \|x\|_\beta \cdot \|y\|_{\beta^*}$, for x, y in R^n, follows immediately from this definition of the dual norm. For

$1 \leq p,\ q \leq \infty$, and $\frac{1}{p} + \frac{1}{q} = 1$, the p-norm $\left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ and the q-norm are dual norms on R^n [6]. If $\|\cdot\|_\beta$ is a norm on R^n, we shall, with a slight abuse of notation, let $\|\cdot\|_\beta$ also denote the corresponding norm on R^m for $m \neq n$. R^n_+ will denote the nonnegative orthant or the set of points in R^n with nonnegative components, while $R^{m\times n}$ will denote the set of all $m \times n$ real matrices. For $A \in R^{m\times n}$, A^T will denote the transpose, A_i will in general denote the ith row, while $\|A\|_\beta$ will denote the matrix norm [1,13] subordinate to the vector norm $\|\cdot\|_\beta$, that is $\|A\|_\beta := \max_{\|x\|_\beta=1} \|Ax\|_\beta$. The consistency condition $\|Ax\|_\beta \leq \|A\|_\beta \|x\|_\beta$ follows immediately from this definition of a matrix norm. We shall also use $\|\cdot\|$ to denote an arbitrary vector norm and its subordinate matrix norm. For an x in R^n we shall make use of some of the following norm-equivalence inequalities [19]

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \leq n\|x\|_\infty \tag{1.5}$$

A vector of ones in R^n for any integer n will be denoted by e. For a differentiable function $g: R^n \to R^m$, $\nabla g(x)$ will denote that $m \times n$ Jacobian matrix at x. Similarly for a differentiable function $L(x,u): (x,u) \in R^{n+m} \to R$, $\nabla_x L(x,u)$ will denote the n-dimensional gradient vector with respect to x, while $\nabla_u L(x,u)$ will denote the m-dimensional gradient vector with respect to u.

2. Error Bounds for Strongly Convex Programs

We first need a preliminary lemma which is essentially Lemma 2.1 of [11] for the case when f is strongly convex. Consider the dual of our nonlinear program (1.1) [7]

$$\begin{array}{ll} \max\limits_{x,u} & L(x,u) - x\nabla_x L(x,u) \\ \text{subject to} & (x,u) \in T := \{(x,u) | u \geq 0,\ \nabla_x L(x,u) \geq 0\} \end{array} \tag{2.1}$$

where $L(x,u)$ is the standard Lagrangian

$$L(x,u) := f(x) + ug(x).$$

The Karush-Kuhn-Tucker (KKT) optimality conditions for (1.1) are [7]

$$\begin{array}{l} v = \nabla_x L(x,u) = \nabla f(x) + u\nabla g(x) \geq 0,\ x \geq 0,\ xv = 0, \\ y = -\nabla_u L(x,u) = -g(x) \geq 0,\ u \geq 0,\ uy = 0 \end{array} \tag{2.2}$$

If we make the definitions

$$z := \begin{pmatrix} x \\ u \end{pmatrix},\quad w := \begin{pmatrix} v \\ y \end{pmatrix},\quad F(z) := \begin{pmatrix} \nabla_x L(x,u) \\ -\nabla_u L(x,u) \end{pmatrix} \tag{2.3}$$

then the Karush-Kuhn-Tucker conditions take on the equivalent complementarity formulation [2]

$$z \geq 0,\ w = F(z) \geq 0,\ zw = 0 \tag{2.4}$$

Our preliminary lemma establishes the strong monotonicity of the "twisted" derivative $F(z)$ under the strong convexity of f and convexity of g.

2.1 Lemma Let f and g be differentiable on R^n, let g be convex on R^n and let f be strongly convex on R^n with positive constant k, then $F(z)$ as defined in (2.3) is continuous and strongly monotone with respect to x on $R^n \times R^m_+$, that is for all $z := \begin{pmatrix} x \\ u \end{pmatrix}$ and $\bar{z} := \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix}$ in $R^n \times R^m_+$

$$(z - \bar{z})(F(z) - F(\bar{z})) \geq k\,\|x - \bar{x}\|_2^2 \tag{2.5}$$

Proof Just replace the last inequality of the proof of Lemma 2.1 of [11] by the inequality of (1.2) above. ∎

We can now state and prove two error bound results.

2.2 Theorem (Error bound in terms of KKT residuals) Let $f: R^n \to R$, $g: R^n \to R^m$ be differentiable on R^n, let f be strongly convex on R^n with positive constant k and let g be convex on R^n. Let either g be linear and $S \neq \phi$, or let g satisfy the Slater constraint qualification, that is

$$g(\hat{x}) < 0,\ \hat{x} > 0 \tag{2.6}$$

for some $\hat{x} \in R^n$. Then for any $(x, u) \in R^n \times R^m_+$ the distance $\|x - \bar{x}\|_2$ to the unique solution $\bar{x}$ of (1.1) is bounded by

$$\begin{aligned}\|x - \bar{x}\|_2 \leq k^{-1/2}\big[x \nabla_x L(x,u) - ug(x) + \alpha\|(-\nabla_x L(x,u))_+\|_1 \\ + \beta\|(g(x))_+\|_\infty + \gamma\|(-x)_+\|_\infty\big]^{1/2}\end{aligned} \tag{2.7}$$

where

$$\alpha := \min_{x \in S}(\|x\|_\infty + \|\nabla f(x)\|_1 / k) \tag{2.8}$$

$$\beta := \min_{(u,v) \in W} \|u\|_1 \tag{2.9}$$

$$\gamma := \min_{(u,v) \in W} \|v\|_1 \tag{2.10}$$

where $W \subset R^{m+n}_+$ is the nonempty closed convex polyhedral set of optimal multipliers (u, v) of the convex program (1.1) associated with the constraints $g(x) \leq 0,\ x \geq 0$.

Proof Since $S \neq \phi$ and f is strongly convex, the program (1.1) has a unique solution $\bar{x}$. Since either g is linear or the Slater constraint qualification (2.6) is satisfied there exist optimal Lagrange multipliers $(\bar{u}, \bar{v}) \in R^{m+n}_+$ such that $(\bar{x}, \bar{u}, \bar{v})$ satisfy the KKT conditions (2.2) [7] and hence the set W of optimal Lagrange multipliers $(\bar{u}, \bar{v})$ is nonempty, closed and convex and in fact polyhedral here. Now for any $x \in S$ we have

$$\|\nabla f(x)\|_1 \cdot \|x - \bar{x}\|_\infty \geq \nabla f(x)(x - \bar{x}) \geq (\nabla f(x) - \nabla f(\bar{x}))(x - \bar{x}) \geq k\|x - \bar{x}\|_2^2 \geq k\|x - \bar{x}\|_\infty^2$$

where the second inequality follows from the minimum principal [7]. Hence

$$\|\bar{x}\|_\infty - \|x\|_\infty \leq \|x - \bar{x}\|_\infty \leq \|\nabla f(x)\|_1 / k$$

and since x is an arbitrary point in S it follows that

$$\|\bar{x}\|_\infty \le \min_{x \in S} \left(\|x\|_\infty + \|\nabla f(x)\|_1 / k\right) = \alpha \tag{2.11}$$

where the minimum exists because of the continuity of the minimand on R^n and the compactness of its level sets. Now let $z := (x, u) \in R^n \times R^m_+$, $(\bar{u}, \bar{v}) \in W$ and let $\bar{z} := (\bar{x}, \bar{u})$. Then by Lemma 2.1 we have that

$$\begin{aligned}
k\|x - \bar{x}\|_2^2 &\le (z - \bar{z})(F(z) - F(\bar{z})) \\
&= zF(z) - \bar{z}F(z) - zF(\bar{z}) \qquad (\text{Since } \bar{z}F(\bar{z}) = 0) \\
&\le zF(z) + \bar{z}(-F(z))_+ + \nabla_x L(\bar{x}, \bar{u})(-x)_+ \\
&\qquad\qquad (\text{Since } ug(\bar{x}) \le 0 \text{ and } \xi \le \xi_+) \\
&= x \nabla_x L(x, u) - ug(x) + \bar{x}(-\nabla_x L(x, u))_+ + \bar{u}(g(x))_+ + \bar{v}(-x)_+ \\
&\le x \nabla_x L(x, u) - ug(x) + \|\bar{x}\|_\infty \cdot \|(-\nabla_x L(x, u))_+\|_1 \\
&\qquad + \|\bar{u}\|_1 \cdot \|(g(x))_+\|_\infty + \|\bar{v}\|_1 \cdot \|(-x)_+\|_\infty
\end{aligned}$$

Since $(\bar{u}, \bar{v})$ is an arbitrary point in W it follows that in the last expression above, $\|\bar{u}\|_1$ and $\|\bar{v}\|_1$ can be replaced by their respective minima over W, while $\|\bar{x}\|_\infty$ can be replaced by its upper bound α given by (2.11). Using the definitions (2.9) and (2.10) we have then

$$\begin{aligned}
k\|x - \bar{x}\|_2^2 \le x \nabla_x L(x, u) - ug(x) + \alpha\|(-\nabla_x L(x, u))_+\|_1 \\
+ \beta\|(g(x))_+\|_\infty + \gamma\|(-x)_+\|_\infty
\end{aligned}$$

from which (2.7) follows immediately. ∎

2.3 Remark Note that the error bound of (2.7) is zero, if and only if x satisfies the Karush-Kuhn-Tucker conditions (2.2) for some $u \in R^m_+$. In fact if we define a perturbation vector $p = (p_1, p_2, p_3, p_4) \in R^{1+n+m+n}$ and define $x(p) \in R^n$ to be a solution of the perturbed Karush-Kuhn-Tucker conditions

$$\begin{aligned}
x \nabla_x L(x, u) - ug(x) &= p_1 \\
(-\nabla_x L(x, u))_+ &= p_2 \\
(g(x))_+ &= p_3 \\
(-x)_+ &= p_4
\end{aligned} \tag{2.12}$$

for some u in R^m_+, then $x(0) = \bar{x}$, the unique solution of (1.1). It follows then from (2.7) that

$$\|x(p) - x(0)\|_2 \leq \lambda \|p\|_1^{1/2} \tag{2.13}$$

where

$$\lambda = (\max\{1, \alpha, \beta, \gamma\}/k)^{1/2} \tag{2.14}$$

The relation (2.13) shows that $x(p)$ is Lipschitzian of order $\frac{1}{2}$, with a Lipschitz constant λ, at $p = 0$.

If the point (x, u) of Theorem 2.1 is both primal and dual feasible, the bound (2.7) of Theorem 2.1 simplifies considerably as indicated in the following.

2.4 Corollary (Error bound for primal-dual feasible points) If in addition to the assumptions of Theorem 2.2, x is primal feasible and (x, u) is dual feasible, that is $x \in S$ and $(x, u) \in T$, then

$$\|x - \bar{x}\|_2 \leq \left(\left(x \nabla_x L(x, u) - ug(x) \right) / k \right)^{1/2} \tag{2.15}$$

This corollary partially extends a result of [12, Equation 2.15] for error bounds for positive semidefinite quadratic programs to strongly monotone convex programs. Pang has given related error bounds for nonlinear complementarity problems [15] and linearly constrained variational inequalities [16].

We note that the error bound of (2.7) contains 3 parameters α, β, and γ which may not be easy to compute. These parameters can be replaced by bounds which are more easily computable. In particular, if we let x^0 be any primal feasible point, and let $\hat{x}$ satisfy, in addition to the Slater constraint qualification (2.6), the dual feasibility condition $(\hat{x}, \hat{u}) \in T$ for some $\hat{u}$, then we have:

$$\alpha \leq \alpha(x^0) := \|x^0\|_\infty + \|\nabla f(x^0)\|_1 / k \tag{2.16}$$

$$\beta \leq \beta(\hat{x}, \hat{u}) := \left(\hat{x} \nabla_x L(\hat{x}, \hat{u}) - \hat{u}g(\hat{x}) \right) / \min_i \; -g_i(\hat{x}) \tag{2.17}$$

$$\gamma \leq \gamma(\hat{x}, \hat{u}) := \left(\hat{x} \nabla_x L(\hat{x}, \hat{u}) - \hat{u}g(\hat{x}) \right) / \min_i \; \hat{x}_i \tag{2.18}$$

where the inequality of (2.16) follows immediately from the definition (2.8) of α and the inequalities (2.17) and (2.18) from Theorem 2.2 of [11]. We therefore have the following.

2.5 Theorem (Explicit error bound in terms of KKT residuals) Let the assumptions of Theorem 2.2 hold including (2.6), let $x^0 \in S$ and let $(\hat{x}, \hat{u}) \in T$ for some $\hat{u} \in R^m_+$. Then for any $(x, u) \in R^n \times R^m_+$ the distance $\|x - \bar{x}\|_2$ to the unique solution $\bar{x}$ of (1.1) is bounded by

$$\begin{aligned} \|x - \bar{x}\|_2 \leq k^{-1/2}\big[x \nabla_x L(x,u) - ug(x) + \alpha(x^0)\|(-\nabla_x L(x,u))_+\|_1 \\ + \beta(\hat{x},\hat{u})\|(g(x))_+\|_\infty + \gamma(\hat{x},\hat{u})\|(-x)_+\|_\infty\big]^{1/2} \end{aligned} \tag{2.19}$$

where $\alpha(x^0)$, $\beta(\hat{x}, \hat{u})$ and $\gamma(\hat{x}, \hat{u})$ are defined by (2.16)-(2.18).

2.6 Remark We note that for a fixed x, x^0, $\hat{x}$ and $\hat{u}$, the choice of u in the bound (2.19) can be optimized by solving the following linear program in order to obtain the best bound on $\|x - \bar{x}\|_2$:

$$\begin{aligned} \min_{(u,s) \in R^m \times R^n} \quad & u(\nabla g(x)x - g(x)) + \alpha(x^0)es \\ & -\nabla f(x) - u\nabla g(x) \leq s \\ & u, s \geq 0 \end{aligned} \tag{2.20}$$

Under the assumptions of Theorem 2.5, the objective function of the feasible linear program (2.20) is bounded below and hence is solvable. Any solution (u, s) of (2.20) will provide an optimal u which will give the best bound in (2.19) for the given fixed x, x^0, $\hat{x}$ and $\hat{u}$.

3. Application to Least 2-Norm Solution for Linear Programs

In this section we use the error bound for strongly convex programs to derive two simple bounds (Theorems 3.1 and 3.2) for the least 2-norm solution of a linear program in terms of any other solution of the linear program. More importantly we give in Theorems 3.7 and 3.8 linearly and superlinearly convergent iterative procedures for determining the least 2-norm solution of a linear program. The proposed schemes should be very helpful in precisely determining the manner in which the perturbation parameter ε and its corresponding error residual $r(\varepsilon)$ (3.14) should be decreased in the highly effective successive overrelaxation methods for solving very large sparse programs [8,9,10].

We consider the linear program

$$\min_{x} \; cx \quad \text{subject to} \;\; Ax \geq b, \; x \geq 0 \tag{3.1}$$

where $c \in R^n$, $b \in R^m$ and $A \in R^{m \times n}$, and its dual

$$\max_{u} \; bu \quad \text{subject to} \;\; A^T u \leq c, \; u \geq 0 \tag{3.2}$$

It is known [9,10] that $\bar{x}$ is the unique least 2-norm solution to (3.1) if and only if $\bar{x}$ is the unique solution to the quadratic program

$$\min_{x} \; cx + \frac{\varepsilon}{2} xx \quad \text{subject to} \;\; Ax \geq b, \; x \geq 0 \tag{3.3}$$

for all $\varepsilon \in (0, \bar{\varepsilon}]$ for some $\bar{\varepsilon} > 0$. The dual to the quadratic program (3.3) is [7]

$$\max_{x,u,v} \; -\frac{\varepsilon}{2} xx + bu \quad \text{subject to} \;\; v = \varepsilon x - A^T u + c, \; (u, v) \geq 0 \tag{3.4}$$

Now if $(\hat{x}, \hat{u})$ is an arbitrary <u>optimal</u> point for the dual linear programs (3.1)-(3.2), then for any $\varepsilon \in (0, \bar{\varepsilon}]$, the point $(\hat{x}, \hat{u}, \hat{v} := \varepsilon \hat{x} - A^T \hat{u} + c)$ is <u>feasible</u> for the dual quadratic program (3.3)-(3.4) and hence by Corollary 2.4

$$\|\hat{x} - \bar{x}\|_2 \leq \left(\frac{c\hat{x} + \varepsilon \hat{x}\hat{x} - b\hat{u}}{\varepsilon} \right)^{1/2} = \|\hat{x}\|_2 \tag{3.5}$$

where $\bar{x}$ is the least 2-norm solution of the linear program (3.1). Hence we have established the following.

3.1 Theorem (Bound for the distance between an LP solution and the least 2-norm LP solution) For the linear program (3.1)

$$\|\hat{x} - \bar{x}\|_2 \leq \|\hat{x}\|_2 \tag{3.6}$$

where $\hat{x}$ is any optimal solution to (3.1) and $\bar{x}$ is the unique optimal solution to (3.1) with least 2-norm.

We can improve on the bound (3.6) if instead of using $\hat{u}$ which is a solution of the linear program (3.2) we use $u(\hat{x}, \varepsilon)$, which minimizes the bound of (2.15) for the given linear program solution $\hat{x}$, and such that $(\hat{x}, u(\hat{x}, \varepsilon))$ is feasible for the dual quadratic program (3.3). Hence we take $u(\hat{x}, \varepsilon)$ as a solution of the linear program

$$\max_u \ bu \quad \text{subject to } A^T u \leq c + \varepsilon\hat{x}, \ u \geq 0 \tag{3.7}$$

This linear program is solvable because it is feasible (its feasible region contains that of (3.2)) and its objective function is bounded above by $(c + \varepsilon\hat{x})\hat{x}$. Hence $bu(\hat{x}, \varepsilon) \geq b\hat{u} = c\hat{x}$ and the bound (3.5) is improved as follows

$$\|\hat{x} - \bar{x}\|_2 \leq \left(\hat{x}\hat{x} - \frac{bu(\hat{x}, \varepsilon) - c\hat{x}}{\varepsilon}\right)^{1/2} \leq \|\hat{x}\|_2 \tag{3.8}$$

Since the bound of (3.8) is valid for all $\varepsilon \in (0, \bar{\varepsilon}]$ and $bu(\hat{x}, \varepsilon)$ is a bounded nonincreasing function of ε we can take its limit as $\varepsilon \downarrow 0$. We summarize this result in the following.

3.2 Theorem (Optimal bound for the distance between an LP solution and the least 2-norm LP solution) For the linear program (3.1) the following bound holds where $\hat{x}$ is any optimal solution to (3.1), $\bar{x}$ is the least 2-norm solution of (3.1) and $u(\hat{x})$ is a solution of the linear program (3.7):

$$\|\hat{x} - \bar{x}\|_2 \leq \lim_{\varepsilon \downarrow 0} \left(\hat{x}\hat{x} - \frac{bu(\hat{x}, \varepsilon) - c\hat{x}}{\varepsilon}\right)^{1/2} \leq \|\hat{x}\|_2 \tag{3.9}$$

The following example illustrates the bounds (3.6) and (3.9).

3.3 Example min x_2 s.t. $-x_1 \geq -2, \ x_2 \geq 1, \ (x_1, x_2) \geq 0$

Problem (3.7) for this LP with $\hat{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is

$$\max \ -2u_1 + u_2 \quad \text{s.t. } -u_1 \leq \varepsilon, \ u_2 \leq 1 + \varepsilon, \ (u_1, u_2) \geq 0 \tag{3.10}$$

The primal solution set is $\bar{S} = \{x \in R^2 | 0 \le x_1 \le 2,\ x_2 = 1\}$ and the least 2-norm solution is $\bar{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We then have

$$\|\bar{x}\|_2 = 1 = \|\hat{x} - \bar{x}\|_2 < \|\hat{x}\|_2 = \sqrt{2}$$

which is the bound (3.6). The solution to (3.10) is $u(\hat{x}, \varepsilon) = \begin{pmatrix} 0 \\ 1+\varepsilon \end{pmatrix}$ and hence $bu(\hat{x}, \varepsilon) = 1 + \varepsilon$ and the bound (3.9) gives

$$1 = \|\hat{x} - \bar{x}\|_2 \le \lim_{\varepsilon \downarrow 0} \left(2 - \frac{1 + \varepsilon - 1}{\varepsilon}\right)^{1/2} = 1$$

which is a sharp bound for this problem.

3.4 Remark We note that under certain assumptions, such as the strong second order sufficient optimality condition and linear independence of the gradients of the active constraints [4, p.44] the function $bu(\hat{x}, \varepsilon)$ is differentiable with respect to ε at $\varepsilon = 0$ and $\frac{d}{d\varepsilon}\left(bu(\hat{x}, \varepsilon)\right)\big|_{\varepsilon=0} = \hat{x}\hat{x}$. For such a case the bound (3.9) degenerates to $\big($since $c\hat{x} = bu(\hat{x}, 0)\big)$

$$\left\|\hat{x} - \bar{x}\right\|_2^2 \le \hat{x}\hat{x} - \frac{d}{d\varepsilon}\left(bu(\hat{x}, \varepsilon)\right)\big|_{\varepsilon=0} = 0$$

and hence $\hat{x} = \bar{x}$, which of course is the consequence of the second order sufficient optimality condition which implies that $\hat{x}$ is a locally and hence globally unique solution of the linear program (3.1).

We conclude by giving linearly and superlinearly convergent procedures for obtaining the least 2-norm solution of the linear program (3.1) based on the error bound (2.7). These procedures should be very useful in the successive overrelaxation (SOR) procedure for solving (3.3) [8,9]. The usefulness comes in determining a method for cutting the size of the parameter ε in (3.3) and the accuracy to which (3.3) is solved for each ε. This results in a precise scheme that drives ε below the value $\bar{\varepsilon}$, which in general is unknown and very difficult to compute. We first outline how the proposed procedure is applied. To solve (3.3) for a fixed ε, we apply an SOR procedure [8,9] or any other procedure to its dual (3.4) with the variable x eliminated through the dual constraint

(3.11)
$$x = \frac{1}{\varepsilon}\left(A^T u + v - c\right)$$

and thus obtaining the dual problem

$$\text{(3.12)} \qquad \min_{(u,v)\geq 0} \theta(u,v) := \min_{(u,v)\geq 0} \frac{1}{2}\left\|A^T u + v - c\right\|_2^2 - \varepsilon bu$$

which would have to be solved for a sufficiently small $\varepsilon \in (0, \bar{\varepsilon}]$. Since we do not know a priori how small ε need be, we consequently need to solve (3.12) for a decreasing sequence of ε values. If an iterative procedure such as SOR is used to solve (3.12), as in the case of very large sparse linear programs [8,9], we would have a procedure with an infinite inner loop. Our present proposed approach now eliminates the need to solve (3.12) exactly and consists of solving (3.12) only to an explicit finite accuracy after which ε is decreased sufficiently to generate a linear or superlinear rate of convergence of the overall procedure.

To define our procedures we need to define approximate and exact solutions to (3.12) and (3.3). For that purpose we first give the necessary and sufficient Karush-Kuhn-Tucker optimality conditions for (3.12):

$$\text{(3.13)} \qquad \begin{array}{lrcl} (a) & \nabla_u \theta(u,v) & = & A(A^T u + v - c) - \varepsilon b \geq 0 \\ (b) & u \nabla_u \theta(u,v) & = & 0 \\ (c) & u & \geq & 0 \\ (d) & \nabla_v \theta(u,v) & = & A^T u + v - c \geq 0 \\ (e) & v \nabla_v \theta(u,v) & = & 0 \\ (f) & v & \geq & 0 \end{array}$$

Now we make the following definitions.

3.5 Definition (Exact solutions to (3.12) & (3.3)) For a fixed positive ε an <u>exact</u> solution to the dual quadratic program (3.12) is designated by $(\bar{u}(\varepsilon), \bar{v}(\varepsilon))$ and hence must satisfy (3.13). The corresponding $\bar{x}(\varepsilon)$ in R^n defined by (3.11) with $(u,v) = (\bar{u}(\varepsilon), \bar{v}(\varepsilon))$ is an <u>exact</u> solution to the quadratic program (3.3). The set of all $(\bar{u}(\varepsilon), \bar{v}(\varepsilon))$ which are exact solutions to (3.12) for a fixed positive ε is designated by $W(\varepsilon)$.

3.6 Definition (Approximate solutions to (3.12) & (3.3)) For a fixed positive ε <u>any</u> point in R_+^{m+n} is an <u>approximate</u> solution to the dual quadratic program (3.12) and is designated by $(u(\varepsilon), v(\varepsilon))$. The corresponding $x(\varepsilon)$ in R^n defined by (3.11) with

$(u,v) = (u(\varepsilon),\ v(\varepsilon))$ is an approximate solution to the quadratic program (3.3). The residual $r(\varepsilon)$ associated with $(u(\varepsilon),\ v(\varepsilon),\ x(\varepsilon))$ is defined by

$$(3.14) \qquad r(\varepsilon) := \left[\left| x(\varepsilon)v(\varepsilon) + u(\varepsilon)\big(Ax(\varepsilon) - b\big) \right| + \left\| \big(b - Ax(\varepsilon)\big)_+ \right\|_\infty + \left\| \big(-x(\varepsilon)\big)_+ \right\|_\infty \right]^{1/2}$$

Note that for an $\varepsilon > 0$ and an approximate solution $(u(\varepsilon),\ v(\varepsilon))$ to (3.12) and a corresponding approximate solution $x(\varepsilon)$ to (3.3), $r(\varepsilon) = 0$ if and only if $(u(\varepsilon),\ v(\varepsilon)) \in W(\varepsilon)$ and $x(\varepsilon) = \bar{x}(\varepsilon)$. We also have that for $\varepsilon \in (0, \bar{\varepsilon}]$ for some $\bar{\varepsilon} > 0$, $\bar{x}(\varepsilon) = \bar{x}$, where $\bar{x}$ is the least 2-norm solution of the linear program (3.1) [9,10].

We are prepared now to state and prove our linearly and superlinearly convergent procedures for computing the least 2-norm solution of the linear program (3.1) and we begin with the former.

3.7 Theorem (Linearly convergent procedure for least 2-norm solution of a linear program) Assume that the linear program (3.1) is solvable and that $b \neq 0$. Let $\{\varepsilon_0,\ \varepsilon_1, \ldots\}$ be a decreasing sequence of positive numbers such that

$$(3.15) \qquad \varepsilon_{i+1} = \mu\varepsilon_i \text{ for some } \mu \in (0,1)$$

and let $\{u(\varepsilon_i),\ v(\varepsilon_i),\ x(\varepsilon_i)\}$ be a corresponding sequence of approximate solutions to (3.12) and (3.3) satisfying Definition 3.6 and such that their residuals as defined by (3.14) satisfy

$$(3.16) \qquad r(\varepsilon_{i+1}) \leq \nu r(\varepsilon_i)$$

for some $\nu > 0$ and such that

$$(3.17) \qquad \nu < \mu^{1/2}$$

Then the sequence $\{x(\varepsilon_i)\}$ converges to $\bar{x}$, the least 2-norm solution of the linear program (3.1), at the linear root-rate [14]

$$(3.18) \qquad \|x(\varepsilon_i) - \bar{x}\|_2 \leq \delta(\nu/\mu^{1/2})^i \text{ for } i \geq \bar{i}$$

for some constant δ and some integer $\bar{i}$.

Proof By Theorem 2.2 we have

$$\begin{aligned}\|x(\varepsilon_i)-\bar{x}(\varepsilon_i)\|_2 \le \varepsilon_i^{-1/2}\big[\big|x(\varepsilon_i)v(\varepsilon_i)+u(\varepsilon_i)\big(Ax(\varepsilon_i)-b\big)\big| \\ +\beta(\varepsilon_i)\|\big(b-Ax(\varepsilon_i)\big)_+\|_\infty+\gamma(\varepsilon_i)\|\big(-x(\varepsilon_i)\big)_+\|_\infty\big]^{1/2}\end{aligned} \tag{3.19}$$

where

$$\beta(\varepsilon_i) := \min_{(u,v)\in W(\varepsilon_i)} \|u\|_1 = \min_{(u,v)\ge 0}\left\{ eu \,\middle|\, \begin{array}{lcll} A^T u + v &=& \varepsilon_i\bar{x}(\varepsilon_i) & +\ c \\ bu &=& c\bar{x}(\varepsilon_i) & +\ \varepsilon_i\bar{x}(\varepsilon_i)\bar{x}(\varepsilon_i)/2 \end{array}\right\} \tag{3.20a}$$

$$\gamma(\varepsilon_i) := \min_{(u,v)\in W(\varepsilon_i)} \|v\|_1 = \min_{(u,v)\ge 0}\left\{ ev \,\middle|\, \begin{array}{lcll} A^T u + v &=& \varepsilon_i\bar{x}(\varepsilon_i) & +\ c \\ bu &=& c\bar{x}(\varepsilon_i) & +\ \varepsilon_i\bar{x}(\varepsilon_i)\bar{x}(\varepsilon_i)/2 \end{array}\right\} \tag{3.20b}$$

By the fundamental theorem for the existence of basic feasible solutions for linear programs [5, Theorem 3.3], it follows that for each ε_i, there exist basis matrices $B_1(\varepsilon_i)$, $B_2(\varepsilon_i)$, that is $(n+1)\times(n+1)$ nonsingular submatrices of $\begin{bmatrix} A^T & I \\ b & 0 \end{bmatrix}$, such that

$$\beta(\varepsilon_i) = (e \quad 0)B_1(\varepsilon_i)^{-1}\begin{pmatrix} \varepsilon_i\bar{x}(\varepsilon_i)+c \\ c\bar{x}(\varepsilon_i)+\varepsilon_i\bar{x}(\varepsilon_i)\bar{x}(\varepsilon_i)/2 \end{pmatrix} \tag{3.21a}$$

$$\gamma(\varepsilon_i) = (0 \quad e)B_2(\varepsilon_i)^{-1}\begin{pmatrix} \varepsilon_i\bar{x}(\varepsilon_i)+c \\ c\bar{x}(\varepsilon_i)+\varepsilon_i\bar{x}(\varepsilon_i)\bar{x}(\varepsilon_i)/2 \end{pmatrix} \tag{3.21b}$$

Since there are only a finite number of basis matrices in $\begin{bmatrix} A^T & I \\ b & 0 \end{bmatrix}$, we have that upon taking B as that basis matrix with largest 1-norm,

$$\begin{aligned}\beta(\varepsilon_i) \text{ or } \gamma(\varepsilon_i) &\le \|B^{-1}\|_1\cdot\left\|\begin{matrix} \varepsilon_i\bar{x}(\varepsilon_i)+c \\ c\bar{x}(\varepsilon_i)+\varepsilon_i\bar{x}(\varepsilon_i)\bar{x}(\varepsilon_i)/2 \end{matrix}\right\|_1 \\ &= \|B^{-1}\|_1\cdot\big[\|\varepsilon_i\bar{x}(\varepsilon_i)+c\|_1+\big|c\bar{x}(\varepsilon_i)+\varepsilon_i\bar{x}(\varepsilon_i)\bar{x}(\varepsilon_i)/2\big|\big]\end{aligned} \tag{3.22}$$

Now

$$\bar{x}(\varepsilon_i) = \bar{x} \ \text{ for } \ \varepsilon\in(0,\bar{\varepsilon}] \tag{3.23}$$

where $\bar{x}$ is the unique least 2-norm solution of the linear program (3.1); and for $\varepsilon_i\ge\bar{\varepsilon}$ we have from (2.11) that

$$\begin{aligned}\|\bar{x}(\varepsilon_i)\|_1 \le n\|\bar{x}(\varepsilon_i)\|_\infty &\le n\ \min_x\left\{\|x\|_\infty+\frac{\|\varepsilon_i x+c\|_1}{\bar{\varepsilon}}\,\middle|\, Ax\ge b,\ x\ge 0\right\} \\ &\le n\ \min_x\left\{\|x\|_\infty+\frac{\varepsilon_0\|x\|_1+\|c\|_1}{\bar{\varepsilon}}\,\middle|\, Ax\ge b,\ x\ge 0\right\} =: \tau\end{aligned} \tag{3.24}$$

Using (3.23) and (3.24) in (3.22) gives

$$\beta(\varepsilon_i) \text{ or } \gamma(\varepsilon_i) \le \|B^{-1}\|_1 \big[\max\{\bar{\varepsilon}\|\bar{x}\|_1 + \|c\|_1,\ \varepsilon_0\tau + \|c\|_1\} + \max\{|c\bar{x}| + \bar{\varepsilon}\bar{x}\bar{x}/2,\ \|c\|_\infty\tau + \varepsilon_0\tau^2/2\}\big] =: \bar{\beta} \tag{3.25}$$

Hence combining (3.19) and (3.25) gives

$$\|x(\varepsilon_i) - \bar{x}(\varepsilon_i)\|_2 \le \varepsilon_i^{-1/2}\sigma r(\varepsilon_i) \tag{3.26}$$

where $r(\varepsilon_i)$ is the residual defined in (3.14) and

$$\sigma := (\max\{1, \bar{\beta}\})^{1/2} \tag{3.27}$$

From (3.15) we have that

$$\varepsilon_i = \mu^i \varepsilon_0 \quad , i = 0, 1, \ldots \tag{3.28}$$

and from (3.16) we have that

$$r(\varepsilon_i) \le \nu^i\, r(\varepsilon_0) \quad , i = 0, 1, \ldots \tag{3.29}$$

Combining (3.26), (3.28) and (3.29) gives

$$\|x(\varepsilon_i) - \bar{x}(\varepsilon_i)\|_2 \le \sigma\varepsilon_0^{-1/2}\, r(\varepsilon_0)\ \ (\nu/\mu^{1/2})^i \tag{3.30}$$

Observing that $\nu/\mu^{1/2} < 1$ from (3.17), and that $\bar{x}(\varepsilon_i) = \bar{x}$ for $\varepsilon_i \in (0, \bar{\varepsilon}]$ it follows that $\lim_{i\to\infty} x(\varepsilon_i) = \bar{x}$. By defining

$$\delta := \sigma\varepsilon_0^{-1/2}\, r(\varepsilon_0) \tag{3.31}$$

and $\bar{i}$ as the smallest integer such that $\varepsilon_i \le \bar{\varepsilon}$, we have from (3.30) and the fact that $\bar{x}(\varepsilon_i) = \bar{x}$ for $\varepsilon_i \le \bar{\varepsilon}$, that for $i \ge \bar{i}$

$$\begin{aligned}\|x(\varepsilon_i) - \bar{x}\|_2 &\le \|x(\varepsilon_i) - \bar{x}(\varepsilon_i)\|_2 + \|\bar{x}(\varepsilon_i) - \bar{x}\|_2 \\ &= \|x(\varepsilon_i) - \bar{x}(\varepsilon_i)\|_2 \le \delta(\nu/\mu^{1/2})^i\end{aligned} \tag{3.32}$$

which establishes (3.18). ∎

We finally note that a superlinear root-rate of convergence [14] can be achieved in the procedure of Theorem 3.7 if we cut the residual $r(\varepsilon_i)$ more sharply than that given by (3.16)-(3.17). In particular we have the following.

3.8 Theorem (Superlinearly convergent procedure for least 2-norm solution of a linear program) Let the assumptions of Theorem 3.7 hold with (3.16) and (3.17) replaced by

$$r(\varepsilon_i) \leq \xi \varepsilon_i^{1/2} \eta^{\rho^i} \tag{3.33}$$

for some $\xi > 0$, $\eta \in (0,1)$ and $\rho > 1$. Then the sequence $\{x(\varepsilon_i)\}$ converges to $\bar{x}$, the least 2-norm solution of the linear program (3.1), at the superlinear root-rate of

$$\|x(\varepsilon_i) - \bar{x}\|_2 \leq \sigma \xi \eta^{\rho^i} \quad \text{for } i \geq \bar{i} \tag{3.34}$$

for some integer $\bar{i}$ and σ defined by (3.27).

Proof From (3.26) and (3.33) we obtain,

$$\|x(\varepsilon_i) - \bar{x}(\varepsilon_i)\|_2 \leq \sigma \xi \eta^{\rho^i} \tag{3.35}$$

Since $\bar{x}(\varepsilon_i) = \bar{x}$ for $i \geq \bar{i}$ for some $\bar{i}$, (3.34) follows from (3.35), and since $\eta^{\rho^i} \to 0$, $x(\varepsilon_i) \to \bar{x}$. ∎

Acknowledgement

We are indebted to our colleague Stephen M. Robinson for pointing out references [3,17,18].

References

1. S. D. Conte & C. de Boor: "Elementary numerical analysis", 3rd Edition, McGraw-Hill, New York 1980.

2. R. W. Cottle: "Nonlinear programs with positively bounded Jacobians", SIAM Journal of Applied Mathematics 14, 1966, 147-158.

3. J. W. Daniel: "Stability of the solution of definite quadratic programs", Mathematical Programming 5, 1973, 41-53.

4. A. V. Fiacco: "Introduction to sensitivity and stability analysis in nonlinear programming", Academic Press, New York 1983.

5. D. Gale: "Theory of linear economic models", McGraw-Hill, New York 1960.

6. A. S. Householder: "The theory of matrices in numerical analysis", Blaisdell Publishing, New York 1964.

7. O. L. Mangasarian: "Nonlinear programming", McGraw-Hill, New York 1969.

8. O. L. Mangasarian: "Solution of symmetric linear complementarity problems by iterative methods", Journal of Optimization Theory and Applications 22, 1977, 465-485.

9. O. L. Mangasarian: "Sparsity-preserving SOR algorithms for separable quadratic and linear programs", Computers and Operations Research 11, 1984, 105-112.

10. O. L. Mangasarian: "Normal solutions of linear programs", Mathematical Programming Study 22, 1984, 206-216.

11. O. L. Mangasarian & L. McLinden: "Simple bounds for solutions of monotone complementarity problems and convex programs", Mathematical Programming 32, 1985, 32-40.

12. O. L. Mangasarian & T.-H. Shiau: "Error bounds for monotone linear complementarity problems", University of Wisconsin Computer Sciences Report No. 606, July 1985, submitted to Mathematical Programming.

13. J. M. Ortega: "Numerical analysis a second course", Academic Press, New York 1972.

14. J. M. Ortega & W. C. Rheinboldt: "Iterative solution of nonlinear equations in several variables", Academic Press, New York 1970.

15. J.-S. Pang: "Inexact Newton methods for the nonlinear complementarity problem", School of Management, University of Texas at Dallas, Richardson, Texas 75083, April 1985.

16. J.-S. Pang: "A posteriori error bounds for the linearly-constrained variational inequality problem", School of Management, University of Texas at Dallas, Richardson, Texas 75083, June 1985.

17. S. M. Robinson: "Strongly regular generalized equations", Mathematics of Operations Research 5, 1980, 43-62.

18. S. M. Robinson: "Generalized equations and their solutions, Part II: Applications of nonlinear programming", Mathematical Programming Study 19, 1982, 200-221.

19. G. W. Stewart: "Introduction to matrix computations", Academic Press, New York 1973.