Hamburger Beiträge
zur Angewandten Mathematik

On the Minimum Norm Solution of Linear Programs

Christian Kanzow, Houduo Qi, Liquin Qi

Reihe A
Preprint 155
July 2000

Hamburger Beiträge zur Angewandten Mathematik

Reihe A	Preprints
Reihe B	Berichte
Reihe C	Mathematische Modelle und Simulation
Reihe D	Elektrische Netzwerke und Bauelemente
Reihe E	Scientific Computing
Reihe F	Computational Fluid Dynamics and Data Analysis

ON THE MINIMUM NORM SOLUTION OF LINEAR PROGRAMS[1]

Christian Kanzow[2], Houduo Qi[3], and Liqun Qi[3]

[2] University of Hamburg
Department of Mathematics
Center for Optimization and Approximation
Bundesstrasse 55
20146 Hamburg
Germany
e-mail: kanzow@math.uni-hamburg.de

[3] University of New South Wales
School of Mathematics
Sydney 2052, NSW
Australia
e-mail: hdqi@maths.unsw.edu.au
L.Qi@unsw.edu.au

July 18, 2000

Abstract. This paper describes a new technique to find the minimum norm solution of a linear program. The main idea is to reformulate this problem as an unconstrained minimization problem with a convex and smooth objective function. The minimization of this objective function can be carried out by a Newton-type method which is shown to be globally convergent. Furthermore, under certain assumptions, this Newton-type method converges in a finite number of iterations to the minimum norm solution of the underlying linear program.

Key Words. Linear programs, minimum norm solution, unconstrained minimization, Newton method, finite termination.

[1]This research was supported by the Australian Research Council. It was carried out while the first author was visiting the University of New South Wales.

1 Introduction

Consider the linear program in primal form

$$\min\ c^T x \quad \text{s.t.} \quad Ax = b,\ x \geq 0 \tag{1}$$

together with its dual

$$\max\ b^T \lambda \quad \text{s.t.} \quad A^T \lambda \leq c, \tag{2}$$

where $A \in \mathbb{R}^{m \times n}, c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$ are the given data, and A is assumed to have full rank. Let us denote by

$$\inf(P) := \inf\{c^T x \mid Ax = b, x \geq 0\}$$

the optimal value of the primal problem (1). Throughout this manuscript, we assume that

$$\inf(P) \in \mathbb{R}. \tag{3}$$

This is equivalent to saying that the primal (and hence also the dual) linear program has a nonempty solution set. In particular, this means that the feasible set of (1) is nonempty.

The aim of this paper is to find the minimum norm solution of the primal program (1), i.e., we want to find the solution x^* of the program

$$\min \frac{1}{2}\|x\|^2 \quad \text{s.t.} \quad Ax = b,\ c^T x = \inf(P),\ x \geq 0. \tag{4}$$

Note that this problem has a unique solution under the assumption (3). Since the minimum norm solution could be a vertex as well as a point belonging to the relative interior of the solution set (depending on the particular linear program under consideration), neither the simplex method (see, e.g., [1]) nor the class of interior-point methods (see, e.g., [12]) will be assured to find the minimum norm solution of (1).

The standard method for finding a minimum norm solution of a convex program is based on the Tikhonov regularization, see [11]. Specialized to our linear program (1), the Tikhonov regularization generates a sequence of iterates $\{x^k\}$ with x^k being the unique solution of the regularized program

$$\min\ c^T x + \frac{\varepsilon_k}{2}\|x\|^2 \quad \text{s.t.} \quad Ax = b,\ x \geq 0, \tag{5}$$

where $\varepsilon_k > 0$ is a positive parameter and the sequence $\{\varepsilon_k\}$ tends to zero. Hence the Tikhonov regularization is, in general, quite costly since it has to solve a sequence of quadratic programs. On the other hand, due to special properties of linear programs, it is known that a solution of a single quadratic program (5) with a sufficiently small but positive parameter ε_k already gives a solution of (4). This follows, e.g., from Mangasarian and Meyer [7, Corollary 2].

In this paper, we suggest an alternative approach for finding the minimum norm solution of the linear program (1). Based on a reformulation of the minimum norm problem (4) as a nonsmooth system of equations by Smith and Wolkowicz [10], we rewrite the problem (4) as an unconstrained minimization problem with a smooth and convex objective function. This reformulation also depends on a certain parameter which has to be sufficiently large but finite.

The details of this idea are given in Section 2. Section 3 then describes a Newton-type method for the minimization of our unconstrained objective function. This method will be shown to be globally convergent and finitely terminating under suitable assumptions. We conclude this paper with some final remarks in Section 4.

A few words about our notation: We denote the n-dimensional real space by $\mathbb{R}^n$. For a vector $x \in \mathbb{R}^n$, we write $\|x\|$ for its Euclidean norm, and x_+ or $[x]_+$ for the vector $\max\{0, x\}$, where the maximum is taken componentwise, i.e., x_+ is the projection of x onto the nonnegative orthant. The rows of a matrix B will be denoted by B_i, whereas we write b_{ij} for the (i,j)th element of the matrix B. Finally, $\mathcal{B}_\varepsilon(\bar{x}) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| \leq \varepsilon\}$ denotes the closed ball with center $\bar{x} \in \mathbb{R}^n$ and radius $\varepsilon > 0$.

2 Unconstrained Minimization Reformulation

In order to find an unconstrained minimization reformulation for the minimum norm solution of the primal linear program (1), we exploit the following result by Smith and Wolkowicz [10] which is essentially based on some related results given by Mangasarian [6] and Mangasarian and Meyer [7].

Theorem 2.1 *A vector $x^* \in \mathbb{R}^n$ is the minimum norm solution of the primal linear program (1) if and only if there exists a positive number $R > 0$ such that, for each $r \geq R$, we have*

$$x^* = [A^T \lambda_r^* - rc]_+,$$

where λ_r^* denotes a solution of the nonlinear system

$$A[A^T \lambda_r - rc]_+ = b.$$

Note that the characterization stated in Theorem 2.1 depends on the parameter r which has to be sufficiently large. Due to the dependence on r we denote our variable by λ_r. According to another result stated in [10], λ_r is related in some way to the dual variable λ from the linear program (2); however, this relation does not play any role within this paper, so we omit the details here.

Motivated by the characterization from Theorem 2.1, let us now introduce the function

$$f(\lambda_r) := \frac{1}{2}\|[A^T \lambda_r - rc]_+\|^2 - b^T \lambda_r. \tag{6}$$

Note that the function f also depends on the parameter r. Some elementary properties of this function are stated in our next result.

Lemma 2.2 *The function f from (6) is convex and continuously differentiable with gradient*

$$\nabla f(\lambda_r) = A[A^T \lambda_r - rc]_+ - b.$$

The proof of Lemma 2.2 is straightforward and therefore omitted here. Comparing the statements of Theorem 2.1 and Lemma 2.2, we see that, for $r > 0$ sufficiently large, the vector $x^* := [A^T\lambda_r^* - rc]_+$ is the minimum norm solution of the primal linear program (1) if and only if λ_r^* is a stationary point of the function f. Due to the convexity of f, this is equivalent to saying that x^* is the minimum norm solution of (1) if and only if λ_r^* is a minimum of the function f. Hence we get the unconstrained minimization reformulation

$$\min f(\lambda_r), \quad \lambda_r \in \mathbb{R}^m, \tag{7}$$

for the minimum norm problem. The advantage of the reformulation (7) is that this is a smooth reformulation to which we can apply any first order minimization method in order to solve this problem, whereas the reformulation provided by Theorem 2.1 gives a nonsmooth nonlinear system of equations, and it is usually much more difficult to find globally convergent methods for solving such a kind of problems. Moreover, in Section 3 we will describe a second order method for the minimization of f (note that f is once but not twice continuously differentiable).

Our next result gives a sufficient condition for the function f to have compact level sets.

Lemma 2.3 *Consider the set*

$$X := \{\lambda_r \in \mathbb{R}^m \mid A^T\lambda_r \leq 0,\, b^T\lambda_r \geq 0\}.$$

Assume that $X = \{0\}$. Then f is coercive.

Proof. The proof is by contradiction. Assume there is a constant $\gamma \in \mathbb{R}$ and a sequence $\{\lambda_r^k\} \subseteq \mathbb{R}^m$ with $\|\lambda_r^k\| \to \infty$ and $f(\lambda_r^k) \leq \gamma$. Without loss of generality, we can assume that the sequence $\{\lambda_r^k/\|\lambda_r^k\|\}$ converges to some vector λ_r^* satisfying $\|\lambda_r^*\| = 1$. Since

$$\frac{1}{2}\left\| \left[A^T\lambda_r^k/\|\lambda_r^k\| - rc/\|\lambda_r^k\|\right]_+ \right\|^2 - b^T(\lambda_r^k/\|\lambda_r^k\|)/\|\lambda_r^k\| = \frac{f(\lambda_r^k)}{\|\lambda_r^k\|^2} \leq \frac{\gamma}{\|\lambda_r^k\|^2},$$

we obtain

$$\frac{1}{2}\|[A^T\lambda_r^*]_+\|^2 \leq 0$$

by taking the limit $k \to \infty$. This implies

$$A^T\lambda_r^* \leq 0.$$

Since $f(\lambda_r^k) \leq \gamma$ and the expression $\frac{1}{2}\|[A^T\lambda_r^k - rc]_+\|^2$ is obviously nonnegative, we also have

$$-b^T\lambda_r^k \leq \gamma.$$

Dividing both sides by $\|\lambda_r^k\|$ and taking limits, we obtain

$$b^T\lambda_r^* \geq 0.$$

Hence λ_r^* is an element of the set X. In view of our assumption, this implies $\lambda_r^* = 0$, a contradiction to $\|\lambda_r^*\| = 1$. This shows that f is indeed coercive. $\square$

Our next result gives a sufficient condition for the set X defined in Lemma 2.3 to contain only the zero vector. The proof of this result makes use of the well-known Farkas-Lemma which states that the system

$$Ax = b, \; x \geq 0$$

has a solution if and only if the implication

$$A^T \lambda_r \leq 0 \Longrightarrow b^T \lambda_r \leq 0$$

holds, see [5].

Theorem 2.4 *Assume that (1) is strictly feasible, i.e., assume there is a vector $\hat{x} \in \mathbb{R}^n$ such that $A\hat{x} = b$ and $\hat{x} > 0$. Then f is coercive.*

Proof. Following Lemma 2.3, let us define the set

$$X := \{\lambda_r \in \mathbb{R}^m \mid A^T \lambda_r \leq 0, \, b^T \lambda_r \geq 0\}.$$

In view of Lemma 2.3, we have to show that X contains only the zero vector. Assume this is not true. Then there exists a vector $\hat{\lambda}_r \in \mathbb{R}^m$ with

$$A^T \hat{\lambda}_r \leq 0, \, b^T \hat{\lambda}_r \geq 0, \, \hat{\lambda}_r \neq 0. \tag{8}$$

Since the linear program (1) has a nonempty feasible set, it follows from (8) and Farkas' Lemma that the vector $\hat{\lambda}_r$ satisfies

$$A^T \hat{\lambda}_r \leq 0, \, b^T \hat{\lambda}_r = 0, \, \hat{\lambda}_r \neq 0. \tag{9}$$

Let $\hat{x} \in \mathbb{R}^n$ be the strictly feasible vector for the primal linear program (1). Then it follows from (9) that

$$0 = b^T \hat{\lambda}_r = \hat{x}^T A^T \hat{\lambda}_r.$$

Since $\hat{x}_i > 0$ and $[A^T \hat{\lambda}_r]_i \leq 0$ for all $i = 1, \ldots, n$, this shows that $A^T \hat{\lambda}_r = 0$. However, A is assumed to have full rank, hence we get $\hat{\lambda}_r = 0$, a contradiction to $\hat{\lambda}_r \neq 0$ in view of (9). □

Theorem 2.4 guarantees that any descent method for the minimization of f will generate a bounded sequence. In particular, this sequence will have an accumulation point. For any reasonable descent method, such an accumulation point will be a stationary point and hence a global minimum of f due to the convexity of this function. A particular descent method with an additional finite termination property will be discussed in the next section.

3 Newton-type Method

In this section, we want to develop a Newton-type method for the solution of the unconstrained minimization problem

$$\min \; f(\lambda_r), \quad \lambda_r \in \mathbb{R}^m.$$

Since the objective function f is once but not twice continuously differentiable, we first review a few concepts from nonsmooth analysis. To this end, we will often write $F := \nabla f$. Since $F = \nabla f$ is locally Lipschitzian, we can define the so-called B-subdifferential of $F = \nabla f$ at a point λ_r by

$$\partial_B F(\lambda_r) := \{H \in \mathbb{R}^{m \times m} \,|\, \exists \{\lambda_r^k\} \subseteq D_F : \lambda_r^k \to \lambda_r \text{ and } F'(\lambda_r^k) \to H\},$$

where D_F denotes the set of points in $\mathbb{R}^m$ where $F = \nabla f$ is differentiable, cf. Qi [8]. The convex hull of this set is Clarke's [2] generalized Jacobian (which, in our case, should better be called the generalized Hessian of f):

$$\partial F(\lambda_r) := \text{conv}\{\partial_B F(\lambda_r)\}.$$

Our first result gives an overestimate for the generalized Jacobian of F at an arbitrary point. We omit its proof here since it follows relatively easily from standard calculus rules in nonsmooth analysis.

Lemma 3.1 *Each element $H \in \partial(\nabla f)(\lambda_r)$ is of the form*

$$H = ADA^T$$

with $D = diag(d_1, \ldots, d_n)$ being a diagonal matrix with entries d_i such that

$$d_i \begin{cases} = 1, & \text{if } [A^T\lambda_r - rc]_i > 0, \\ \in [0,1], & \text{if } [A^T\lambda_r - rc]_i = 0, \\ = 0, & \text{if } [A^T\lambda_r - rc]_i < 0. \end{cases}$$

Based on the previous result, we can prove the following nonsingularity result which will play an important role in our subsequent analysis.

Theorem 3.2 *Let $r > 0$ be sufficiently large such that Theorem 2.1 applies, and let λ_r^* be a stationary point of f. Assume further that the corresponding minimum norm solution $x^* = [A^T\lambda_r^* - rc]_+$ of the primal linear program (1) has at least m positive components, and that we can find a subset of m positive components of x^* such that the corresponding columns of A are linearly independent. Then all elements $H \in \partial(\nabla f)(\lambda_r^*)$ are positive definite.*

Proof. Let $H \in \partial(\nabla f)(\lambda_r^*)$ be arbitrarily given. Let us introduce the index sets

$$\begin{aligned} I_+ &:= \{i \,|\, [A^T\lambda_r^* - rc]_i > 0\}, \\ I_0 &:= \{i \,|\, [A^T\lambda_r^* - rc]_i = 0\}, \\ I_- &:= \{i \,|\, [A^T\lambda_r^* - rc]_i < 0\}. \end{aligned}$$

In view of Lemma 3.1, we then have

$$H = ADA^T$$

with $D = \operatorname{diag}(d_1, \ldots, d_n)$ and

$$d_i \begin{cases} = 1, & \text{if } i \in I_+, \\ \in [0,1], & \text{if } i \in I_0, \\ = 0, & \text{if } i \in I_-. \end{cases}$$

Since the minimum norm solution x^* of (1) is supposed to have at least m positive components and since we have $x^* = [A^T\lambda_r^* - rc]_+$, it follows that the index set I_+ contains at least m indices. Hence the diagonal matrix D contains at most $n-m$ zero entries.

We now claim that this implies that H is positive definite. To this end, assume without loss of generality that the first m components of x^* are positive such that the corresponding column vectors of the matrix A are linearly independent. Then we can write

$$D = \begin{pmatrix} D_m & \\ & D_{n-m} \end{pmatrix} \quad \text{and} \quad A = (A_m \;\; A_{n-m})$$

with $D_m = I_m$, D_{n-m} being a diagonal matrix with entries in $[0,1]$, and $A_m \in \mathbb{R}^{m\times m}$, $A_{n-m} \in \mathbb{R}^{m\times(n-m)}$ such that A_m is nonsingular. This implies

$$\begin{aligned} H &= ADA^T \\ &= (A_m \;\; A_{n-m}) \begin{pmatrix} I_m & \\ & D_{n-m} \end{pmatrix} \begin{pmatrix} A_m^T \\ A_{n-m}^T \end{pmatrix} \\ &= A_m A_m^T + A_{n-m} D_{n-m} A_{n-m}^T, \end{aligned}$$

i.e., H is positive definite since $A_m A_m^T$ is positive definite and $A_{n-m}D_{n-m}A_{n-m}^T$ is at least positive semi-definite. □

Note that the assumptions of Theorem 3.2 imply that λ_r^* is the unique global minimum of our convex objective function f. However, the assumptions do not imply that the solution set of the primal linear program (1) reduces to a singleton.

We further note that the assumptions of Theorem 3.2 are satisfied in case the minimum norm solution x^* of the linear program (1) is a nondegenerate vertex. This follows from the fact that, by definition, a nondegenerate vertex has exactly m positive components, together with a standard result from linear programming which says that the columns of A corresponding to the positive components are linearly independent, cf. [1]. In addition, the assumption may, in many cases, also be satisfied if the minimum norm solution x^* is not a vertex of (1).

We next describe our Newton-type algorithm for the unconstrained minimization of f.

Algorithm 3.3 *(Newton-type method)*

(S.0) Choose $\lambda_r^0 \in \mathbb{R}^m, \rho > 0, \beta \in (0,1), p > 1, \varepsilon \geq 0$, *and set* $k := 0$.

(S.1) If $\|\nabla f(\lambda_r^k)\| \leq \varepsilon$: *STOP.*

(S.2) Choose $H_k \in \partial(\nabla f)(\lambda_r^k)$, *and calculate a solution* $\Delta\lambda_r^k \in \mathbb{R}^m$ *of the linear system*

$$H_k \Delta\lambda_r = -\nabla f(\lambda_r^k). \tag{10}$$

Consider the following cases:

(a) *If the system (10) has no solution, set* $\Delta\lambda_r^k := -\nabla f(\lambda_r^k)$ *and go to (S.3).*

(b) *If (10) has a solution satisfying* $\|\nabla f(\lambda_r^k + \Delta\lambda_r^k)\| = 0$*, set* $\lambda_r^{k+1} := \lambda_r^k + \Delta\lambda_r^k$ *and STOP.*

(c) *If (10) has a solution* $\Delta\lambda_r^k$ *satisfying*

$$\nabla f(\lambda_r^k)^T \Delta\lambda_r^k \leq -\rho\|\Delta\lambda_r^k\|^p,$$

then go to (S.3), else set $\Delta\lambda_r^k := -\nabla f(\lambda_r^k)$ *and go to (S.3).*

(S.3) *Compute a stepsize* $t_k = \max\{\beta^\ell \mid \ell = 0, 1, 2, \dots\}$ *such that*

$$f(\lambda_r^k + t_k\Delta\lambda_r^k) \leq f(\lambda_r^k) + \sigma t_k \nabla f(\lambda_r^k)^T \Delta\lambda_r^k.$$

(S.4) *Set* $\lambda_r^{k+1} := \lambda_r^k + t_k\Delta\lambda_r^k, k \leftarrow k+1$*, and go to (S.1).*

Note that Algorithm 3.3 looks pretty much like a standard Newton method for the minimization of an unconstrained optimization problem: In Step (S.2), we try to solve the Newton equation (10) with H_k being a substitute for the usually not existing Hessian of the function f at the current iterate. We then check whether we are already at a solution or whether we have to switch to a steepest descent direction in case the linear system (10) is not solvable or does not provide a search direction satisfying the sufficient decrease condition from Step (S.2) (c). Step (S.3) is then a standard Armijo line search applied to the objective function f. In our analysis, we will always assume that Algorithm 3.3 does not terminate in Step (S.1) after a finite number of iterations (though we will show that finite termination will eventually happen in Step (S.2) (b)).

Our first result contains the global convergence properties of Algorithm 3.3.

Theorem 3.4 *Every accumulation point* λ_r^* *of a sequence* $\{\lambda_r^k\}$ *generated by Algorithm 3.3 is a stationary point and hence a global minimum of* f*. If, in addition,* $r > 0$ *satisfies the assumption of Theorem 2.1 and* $x^* = [A^T\lambda_r^* - rc]_+$*, then* x^* *is the minimum norm solution of the primal linear program (1).*

Proof. The fact that every accumulation point λ_r^* is a stationary point of f is rather standard and can more or less be verified by following the proof of Theorem 11 (a) in [3]. Since f is convex by Lemma 2.2, we then obtain that λ_r^* is also a global minimum of f. Finally, the last statement follows from Theorem 2.1. □

Our next aim is to show that Algorithm 3.3 has a finite termination property. To this end, we follow the technique of proof from [4], where the finite termination of the minimum function approach for the solution of linear complementarity problems is established.

We begin with a technical result.

Lemma 3.5 *Let* λ_r^* *be a stationary point of* f*. Then there is an* $\varepsilon > 0$ *such that*

$$\nabla f(\lambda_r) - H(\lambda_r - \lambda_r^*) = 0$$

holds for all $\lambda_r \in \mathcal{B}_\varepsilon(\lambda_r^*)$ *and all* $H \in \partial(\nabla f)(\lambda_r)$*.*

Proof. First recall from Lemma 2.2 that

$$\nabla f(\lambda_r) = A[A^T\lambda_r - rc]_+ - b.$$

Furthermore, any $H \in \partial(\nabla f)(\lambda_r)$ is given by

$$H = ADA^T$$

for some diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n)$ with entries (depending on λ_r) as specified in Lemma 3.1. Since λ_r^* is a stationary point of f, it follows that

$$b_i = A_i[A^T\lambda_r^* - rc]_+ \quad \forall i = 1, \dots, m.$$

Using these relations, we obtain for any fixed component $i \in \{1, \dots, m\}$:

$$\begin{aligned}
& [\nabla f(\lambda_r) - H(\lambda_r - \lambda_r^*)]_i \\
= \; & A_i[A^T\lambda_r - rc]_+ - b_i - H_i(\lambda_r - \lambda_r^*) \\
= \; & A_i[A^T\lambda_r - rc]_+ - A_i[A^T\lambda_r^* - rc]_+ - A_iDA^T(\lambda_r - \lambda_r^*) \\
= \; & \sum_{j=1}^{n} a_{ij}\Big(\max\{0, [A^T\lambda_r - rc]_j\} - \max\{0, [A^T\lambda_r^* - rc]_j\} - d_j[A^T(\lambda_r - \lambda_r^*)]_j\Big) \\
=: \; & \sum_{j=1}^{n} a_{ij}s_j.
\end{aligned}$$

We now consider five cases in order to show that each s_j in the above expression is equal to zero provided that λ_r is sufficiently close to λ_r^*.

(a) If $[A^T\lambda_r^* - rc]_j > 0$, then we also have $[A^T\lambda_r - rc]_j > 0$ for λ_r close to λ_r^*. In view of Lemma 3.1, this implies $d_j = 1$. Hence we immediately obtain $s_j = 0$ in this case.

(b) If $[A^T\lambda_r^* - rc]_j < 0$, then we also have $[A^T\lambda_r - rc]_j < 0$ for λ_r close to λ_r^*. Lemma 3.1 then implies $d_j = 0$, and this shows that we have $s_j = 0$ also in this case.

(c) If $[A^T\lambda_r^* - rc]_j = 0$ and $[A^T\lambda_r - rc]_j > 0$, we have $d_j = 1$ and therefore also $s_j = 0$.

(d) If $[A^T\lambda_r^* - rc]_j = 0$ and $[A^T\lambda_r - rc]_j < 0$, we have $d_j = 0$. This also implies $s_j = 0$.

(e) If $[A^T\lambda_r^* - rc]_j = 0$ and $[A^T\lambda_r - rc]_j = 0$, we have $d_j \in [0, 1]$ by Lemma 3.1. However, since, in this case, we have

$$[A^T(\lambda_r - \lambda_r^*)]_j = [A^T\lambda_r - rc]_j - [A^T\lambda_r^* - rc]_j = 0 - 0 = 0,$$

we see that $s_j = 0$ holds regardless of the precise value of d_j.

Since the component $i \in \{1, \dots, m\}$ was chosen arbitrarily, the claim follows. $\square$

We are now able to state a finite termination property for Algorithm 3.3.

Theorem 3.6 *Let $r > 0$ be sufficiently large such that Theorem 2.1 applies, and let λ_r^* be a stationary point of f satisfying the assumptions of Theorem 3.2. Then there is an $\varepsilon > 0$ such that, for all λ_r^k sufficiently close to λ_r^*, the next iterate λ_r^{k+1} generated by Algorithm 3.3 is equal to λ_r^*, i.e., $x^{k+1} := [A^T\lambda_r^{k+1} - rc]_+$ is equal to the minimum norm solution of the primal linear program (1).*

Proof. Let $\varepsilon_1 := \varepsilon$ with ε being the constant from Lemma 3.5. In view of Theorem 3.2 and a standard perturbation result for the generalized Jacobian (see [9]), it follows that there is a positive number ε_2 such that any matrix $H \in \partial(\nabla f)(\lambda_r)$ is nonsingular for all $\lambda_r \in \mathcal{B}_r(\lambda_r^*)$. Now define $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$ and assume that $\lambda_r^k \in \mathcal{B}_\varepsilon(\lambda_r^*)$. Then the matrix $H_k \in \partial(\nabla f)(\lambda_r^k)$ chosen in Step (S.2) of Algorithm 3.3 is nonsingular. Hence Algorithm 3.3 and Lemma 3.5 imply

$$(\lambda_r^k + \Delta\lambda_r^k) - \lambda_r^* = \lambda_r^k - H_k^{-1}\nabla f(\lambda_r^k) - \lambda_r^* = -H_k^{-1}\left(\nabla f(\lambda_r^k) - H_k(\lambda_r^k - \lambda_r^*)\right) = 0.$$

Consequently, Algorithm 3.3 takes $\lambda_r^{k+1} = \lambda_r^k + \Delta\lambda_r^k = \lambda_r^*$ as the next iterate and terminates in Step (S.2) (b) since λ_r^* is a stationary point of f. $\square$

4 Concluding Remarks

In this paper, we introduced an alternative way to find the minimum norm solution of a linear program. In contrast to the classical Tikhonov regularization, which has to solve a quadratic program, we have to solve an unconstrained convex minimization problem in order to find the minimum norm solution. However, similarly to the Tikhonov regularization, our approach depends on a certain parameter which has to be sufficiently large but whose precise value is usually unknown a priori. Further work has to be done in order to get a better understanding on the choice of this parameter. Finally, we note an interesting difference between the Tikhonov regularization and our reformulation: While the former has to solve a quadratic program in n variables, our unconstrained minimization reformulation depends on m variables only, and m can be significantly smaller than n.

References

[1] D. Bertsimas and J.N. Tsitsiklis: *Introduction to Linear Optimization.* Athena Scientific, Belmont, MA, 1997.

[2] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, NY, 1983 (reprinted by SIAM, Philadelphia, PA, 1990).

[3] T. De Luca, F. Facchinei and C. Kanzow: *A semismooth equation approach to the solution of nonlinear complementarity problems.* Mathematical Programming 75, 1996, pp. 407–439.

[4] A. Fischer and C. Kanzow: *On finite termination of an iterative method for linear complementarity problems.* Mathematical Programming 74, 1996, pp. 279–292.

[5] O.L. Mangasarian: *Nonlinear Programming.* McGraw-Hill, New York, NY, 1969 (reprinted by SIAM, Philadelphia, PA, 1994).

[6] O.L. Mangasarian: *Normal solutions of linear programs.* Mathematical Programming 22, 1984, pp. 206–216.

[7] O.L. Mangasarian and R.R. Meyer: *Nonlinear perturbation of linear programs.* SIAM Journal on Control and Optimization 17, 1979, pp. 745–757.

[8] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[9] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–367.

[10] P.W. Smith and H. Wolkowicz: *A nonlinear equation for linear programming.* Mathematical Programming 34, 1986, pp. 235–238.

[11] A.N. Tikhonov and V.Y. Arsenin: *Solutions of ill-posed problems.* Halsted Press, Wiley, New York, NY, 1977.

[12] S.J. Wright: *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, PA, 1997.