This is to certify that the

dissertation entitled

An Algorithm For Non-Negative Least Error
Minimal Norm Solutions

presented by

Panagiotis V. Nikolopoulos

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Mathematics

V. P. Sreedharan
Major professor

Date 7/28/89

AN ALGORITHM FOR NON-NEGATIVE LEAST ERROR MINIMAL NORM SOLUTIONS

By

Panagiotis Vasilios Nikolopoulos

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1989

ABSTRACT

AN ALGORITHM FOR NON–NEGATIVE LEAST ERROR MINIMAL NORM SOLUTIONS

By

Panagiotis Vasilios Nikolopoulos

In this thesis we consider non–negative solutions of a system of m real linear equations, $Ax = b$, in n unknowns which minimize the residual error when $\mathbb{R}^m$ is equipped with a strictly convex norm. Out of these solutions we seek the one which is of least norm when $\mathbb{R}^n$ is equipped with a strictly convex and smooth norm. An implementable iterative algorithm accomplishing this is given. The algorithm is globally convergent and it does not require that a non–negative minimal error solution be found first. As a special case, we test the algorithm for the ℓ^p–norms $(1 < p < \infty)$. Numerical results are also included.

ACKNOWLEDGMENTS

I would like to express my sincere thanks to Professor V.P. Sreedharan, without whose direction and encouragement I would not have written this thesis. I also thank the other members of my committee, Professors C. Seebeck, C. Weil, D. Dunninger and J. Plotkin. Finally, special thanks are reserved for Cathy Sparks for typing the final manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

CHAPTER 1

1.1 INTRODUCTION

In this chapter we assume that the system of m real linear equations in n unknowns $Ax = b$ has a non–negative solution. We give an implementable iterative algorithm converging to the least norm $\|\cdot\|$ solution of $Ax = b, x \geq 0$ for a strictly convex norm $\|\cdot\|$ on $\mathbb{R}^n$. This algorithm is modeled after a similar algorithm in [6]. Before we state the algorithm we formulate some duality theorems, again analogous to these in [6]. These theorems will be needed to show that the algorithm is convergent and is properly formulated.

1.2 NOTATION AND SOME PRELIMINARIES

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, $n \geq 1$ and $<\cdot,\cdot>$ denote the standard Euclidean inner product, with $\|\cdot\|_2$, the corresponding Euclidean length. For x , $y \in \mathbb{R}^n$, we write $x \geq y$ iff $x_j \geq y_j$, $\forall j = 1,\ldots,n$ where $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$, x_j, $y_j \in \mathbb{R}$. The definition of $\leq$ is now clear. In formulating the duality theorems we will need the well known notion of the dual norm on $\mathbb{R}^n$: Given the norm $\|\cdot\|$ on $\mathbb{R}^n$, we define the *dual norm* $\|\cdot\|'$ by

(1.2.1) $\quad \|y\|' = \max \{<x , y> \mid \|x\| = 1, x \in \mathbb{R}^n\}$.

Given $y \neq 0$, we define (see [7], [8]) y' a $\|\cdot\|$–dual and y^* a $\|\cdot\|'$–dual by the equations

(1.2.2) $\quad \|y'\| = 1, <y', y> = \|y\|'$,

and

(1.2.3) $\quad \|y^*\|' = 1, <y^*, y> = \|y\|$.

The norm $\|\cdot\|$ is said to be *strictly convex* iff the unit sphere

$S = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ has no line segments on it. The norm $\|\cdot\|$ is said to be *smooth* iff through each point of unit norm in $\mathbb{R}^n$, there passes precisely one hyperplane supporting the closed unit ball $B = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$. One easily sees that if the norm $\|\cdot\|$ is strictly convex (smooth), then $\|\cdot\|$ $(\|\cdot\|')$–duals are unique and

(1.2.4) $\quad x^{*\,\prime} = x \mid \|x\| \quad (x'^{\,*} = x \mid \|x\|'),\ \forall\ x \neq 0.$

Furthermore, the map $x \mapsto x'$ $(x \mapsto x^*)$ of $\mathbb{R}^n \backslash \{0\}$ into the $\|\cdot\|$ $(\|\cdot\|')$–unit sphere is continuous, and positively homogeneous of degree zero.

A non-empty convex subset K of $\mathbb{R}^n$ is said to be a *convex cone* iff $\lambda y \in K, \forall y \in K$ and $\forall \lambda \geq 0$.

1.3 SOME DUALITY THEORY

Let $a \in \mathbb{R}^n$. With the primal problem (P)

(P) $\quad \min\{\|x-a\| \mid x \in \mathbb{R}^n,\ Ax = b,\ x \geq 0\}$

we associate a "dual" problem (P'):

(P') $\quad \max\ \{\langle y,b\rangle - \langle A^T y + \xi, a\rangle \mid \xi \in \mathbb{R}^n,\ \xi \geq 0,\ y \in \mathbb{R}^m, \|A^T y + \xi\|' \leq 1\}$

where A^T denotes the transpose of A.

We have the following basic theorems:

1.3.1 THEOREM. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. Assume that $K = \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ is nonempty. Then the problems (P) and (P') have the same value.

Proof. For any $x \geq 0$ such that $Ax = b$, $\xi \in \mathbb{R}^n$, $\xi \geq 0$, $y \in \mathbb{R}^m$ with $\|A^T y + \xi\|' \leq 1$, we have

$$\begin{aligned}
\langle y,b\rangle - \langle A^T y + \xi, a\rangle &= \langle y, Ax\rangle - \langle A^T y + \xi, a\rangle \\
&= \langle A^T y, x\rangle - \langle A^T y + \xi, a\rangle \\
&\le \langle A^T y + \xi, x\rangle - \langle A^T y + \xi, a\rangle \\
&= \langle A^T y + \xi, x - a\rangle \\
&\le \|A^T y + \xi\|' \, \|x - a\| \\
&\le \|x - a\|. \qquad (1.3.1.1)
\end{aligned}$$

This proves that value of (P') $\le$ value of (P).

Now value of (P) $= d(a,K)$

$$= \inf\{\|x - a\| \mid x \in K\}, \qquad (1.3.1.2)$$

where $K = \{x \in \mathbf{R}^n \mid Ax = b,\ x \ge 0\}$.

By the duality theorem of Nirenberg [4], we have

$$d(a,K) = \max_{\|z\|' \le 1} (\langle z,a\rangle - \sigma(z)), \qquad (1.3.1.3)$$

where σ is the support function of the polyhedrally convex set K, i.e.

$$\sigma(z) = \sup\{\langle z,x\rangle \mid x \in K\}. \qquad (1.3.1.4)$$

So

$$d(a,K) = \max_{\|z\|' \le 1} (\langle z,a\rangle - \sup_{x \in K} \langle z,x\rangle). \qquad (1.3.1.5)$$

Now

$$-\sup\{\langle z,x\rangle \mid x \in K\} = \inf\{\langle -z,x\rangle \mid Ax = b, x \ge 0\}. \qquad (1.3.1.6)$$

The standard linear program on the right of the above equation is feasible by hypothesis. So by the well known strong duality theorem of linear programming [3],

$$\inf\{\langle -z,x\rangle \mid Ax = b, x \ge 0\} = \max\{\langle y,b\rangle \mid A^T y \le -z\}, \qquad (1.3.1.7)$$

with the convention that maximum over the empty set is $-\infty$. Rewriting (1.3.1.7) and combining it with (1.3.1.6) yields

$$-\sup\{\langle z,x\rangle \mid x \in K\} = \max\{\langle y,b\rangle \mid A^T y + \xi = -z,\ y \in \mathbf{R}^m,\ \xi \in \mathbf{R}^n,\ \xi \ge 0\}.$$

Inserting this in (1.3.1.5) we get

(1.3.1.8) $$d(a,K)=\max_{\|z\|'\leq 1}(\langle z,a\rangle + \max\{\langle y,b\rangle \mid A^T y + \xi = -z, \xi \geq 0\}).$$

Given z, with $\|z\|' \leq 1$, if there does not exist $y \in \mathbf{R}^m$, $\xi \in \mathbf{R}^n$, $\xi \geq 0$ such that $A^T y + \xi = -z$, then the linear program occuring in (1.3.1.8) has the value $-\infty$. Since $0 \leq d(a,K) < \infty$, we may therefore consider only z's of the form

(1.3.1.9) $$z = -A^T y - \xi,\ y \in \mathbf{R}^m,\ \xi \in \mathbf{R}^n,\ \xi \geq 0,\ \|z\|' \leq 1.$$

From (1.3.1.8), we see that there exists $\bar{z}$, $\|\bar{z}\|' \leq 1$, such that

(1.3.1.10) $$d(a,K) = \langle \bar{z},a\rangle + \max\{\langle y,b\rangle \mid A^T y + \xi = -\bar{z},\ \xi \geq 0\}.$$

By remarks above, $\exists$ $\bar{y} \in \mathbf{R}^m$ and $\bar{\xi} \in \mathbf{R}^n$, $\bar{\xi} \geq 0$ such that

(1.3.1.11) $$\bar{z} = -A^T \bar{y} - \bar{\xi}$$

and

(1.3.1.12) $$\begin{aligned} d(a,K) &= \langle \bar{z},a\rangle + \langle \bar{y},b\rangle \\ &= -\langle A^T\bar{y} + \bar{\xi},a\rangle + \langle \bar{y},b\rangle \\ &\leq \max\{\langle y,b\rangle - \langle A^T y + \xi,a\rangle \mid y \in \mathbf{R}^m,\ \xi \in \mathbf{R}^n,\ \xi \geq 0, \\ &\qquad \|A^T y + \xi\|' \leq 1\} \end{aligned}$$

(1.3.1.13) $$= \text{value of } (P').$$

We have now shown that value of (P) $\leq$ value of (P') and so the theorem is completely proved.

REMARK. The easy half in the above theorem, viz: value of (P') $\leq$ value of (P) will be referred to as the *weak duality principle.*

1.3.2 THEOREM. Let the norm $\|\cdot\|$ on $\mathbf{R}^n$ be strictly convex. Assume also that $K = \{x \in \mathbf{R}^n \mid Ax = b,\ x \geq 0\}$ is nonempty and $a \in \mathbf{R}^n \setminus K$. Then $\bar{x} \in \mathbf{R}^n$ solves (P) iff $A\bar{x} = b$, $\bar{x} \geq 0$ and $\exists y \in \mathbf{R}^m$, $\xi \in \mathbf{R}^n$, $\xi \geq 0$

such that

$$(1.3.2.1) \qquad \|A^T y + \xi\|' = 1, \quad \langle y,b\rangle - \langle A^T y + \xi,a\rangle > 0,$$

and

$$(1.3.2.2) \qquad \bar{x} - a = (\langle y,b\rangle - \langle A^T y + \xi,a\rangle)(A^T y + \xi)'.$$

Furthermore,

$$(1.3.2.3) \qquad \langle \xi,\bar{x}\rangle = 0.$$

Proof. "If" part: We have by (1.3.2.1) and (1.3.2.3)

$$\|\bar{x} - a\| = \langle y,b\rangle - \langle A^T y + \xi,a\rangle,$$

which by Theorem 1.3.1 shows that $\|\bar{x} - a\|$ is the value of (P) and so $\bar{x}$ solves (P).

"Only if" part: Let $y \in \mathbb{R}^m$, $\xi \geq 0$ be a solution of (P'), so that by Theorem 1.3.1

$$\|\bar{x} - a\| = \langle y,b\rangle - \langle A^T y + \xi,a\rangle = \rho > 0.$$

Now

$$(1.3.2.4) \qquad \langle A^T y + \xi, \bar{x} - a\rangle \leq \|A^T y + \xi\|' \, \|\bar{x} - a\| = \rho.$$

On the other hand,

$$\begin{aligned} \langle A^T y + \xi, \bar{x} - a\rangle &= \langle A^T y + \xi,\bar{x}\rangle - \langle A^T y + \xi,a\rangle \\ &= \langle A^T y,\bar{x}\rangle + \langle \xi,\bar{x}\rangle - \langle A^T y + \xi,a\rangle \\ &= \langle y,A\bar{x}\rangle + \langle \xi,\bar{x}\rangle - \langle A^T y + \xi,a\rangle \\ &= \langle y,b\rangle - \langle A^T y + \xi,a\rangle + \langle \xi,\bar{x}\rangle \\ (1.3.2.5) \qquad &= \rho + \langle \xi,\bar{x}\rangle \geq \rho, \end{aligned}$$

since both the vectors ξ and $\bar{x}$ are ≥ 0. From (1.3.2.4) and (1.3.2.5) we see that

$$(1.3.2.6) \qquad \langle A^T y + \xi, \bar{x} - a\rangle = \rho,$$

and that $\langle \xi,\bar{x}\rangle = 0$, proving (1.3.2.3). Since $\|A^T y + \xi\|' = 1$ and $\|(\bar{x} - a)/\rho\| = 1$, we see from (1.3.2.6)

$$\langle (\bar{x} - a)/\rho, \; A^T y + \xi\rangle = \|A^T y + \xi\|'.$$

In view of the strict convexity of the norm $\|\cdot\|$, we get $(\bar{x} - a)/\rho = (A^T y + \xi)'$, or $\bar{x} - a = \rho(A^T y + \xi)'$, which is (1.3.2.2), completing the proof of the theorem.

1.3.3. If we assume that the norm $\|\cdot\|$ is both smooth and strictly convex we can get more symmetric results as stated below.

1.3.4 THEOREM. Let the norm $\|\cdot\|$ be both smooth and strictly convex. Assume that $K = \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ is nonempty and $a \in \mathbb{R}^n \backslash K$. Then $\bar{x} \in \mathbb{R}^n$ solves (P) iff $A\bar{x} = b$, $\bar{x} \geq 0$ and $\exists y \in \mathbb{R}^n$, $\xi \in \mathbb{R}^n$, $\xi \geq 0$ such that

(1.3.4.1) $$u^* = A^T y + \xi,$$

where $u = \bar{x} - a$, and

(1.3.4.2) $$<\xi, \bar{x}> = 0.$$

Furthermore, (y,ξ) solves (P').

Proof. "Only if" part: By Theorem 1.3.2, there exists $\bar{x}$ such that $A\bar{x} = b$, $\bar{x} \geq 0$ and (y,ξ) satisfying (1.3.2.1), (1.3.2.2) and (1.3.2.3). By (1.3.2.2), $u = \|u\|(A^T y + \xi)'$, so that $u^* = A^T y + \xi$, due to the smoothness of $\|\cdot\|$.

"If" part: We shall show that $\bar{x}$ solves (P) and also (y,ξ) solves (P'). Observe that by the strict convexity of the norm $\|\cdot\|$,

(1.3.4.3) $$u/\|u\| = u^{*\prime} = (A^T y + \xi)'.$$

Now

(1.3.4.4) $$1 = \|u^*\|' = \|A^T y + \xi\|'$$
$$= <(A^T y + \xi)', A^T y + \xi>.$$

By (1.3.4.4) and (1.3.4.3) we see that

(1.3.4.5) $$\|u\| = <u, A^T y + \xi>$$

or

$$\|\bar{x} - a\| = \langle \bar{x} - a, A^T y + \xi \rangle$$

(1.3.4.6) $$= \langle y,b \rangle - \langle A^T y + \xi, a \rangle$$

due to the calculation in (1.3.2.5) and the fact $\langle \xi, \bar{x} \rangle = 0$. Equation (1.3.4.6) in view of Theorem 1.3.1 completes the proof.

1.3.5 COROLLARY. Let $\bar{x}$ solve (P) under the hypotheses of the theorem. Then

(1.3.5.1) $$\langle y,b \rangle = \langle u^*, \bar{x} \rangle.$$

Proof. By (1.3.4.6),

$$\begin{aligned} \langle y,b \rangle &= \|u\| + \langle A^T y + \xi, a \rangle \\ &= \langle A^T y + \xi, a + u \rangle, \text{ by (1.3.4.5)} \\ &= \langle A^T y + \xi, \bar{x} \rangle \\ &= \langle u^*, \bar{x} \rangle. \end{aligned}$$

1.3.6. The special case of Theorem 1.3.4 with $\|\cdot\| = \|\cdot\|_2$, the standard Euclidean norm, is important for later applications. So we record it explicitly.

1.3.7 THEOREM. Assume that the set $K = \{x \epsilon \mathbb{R}^n \mid Ax = b, x \geq 0\}$ is nonempty and that $a \epsilon \mathbb{R}^n \setminus K$. Then $\bar{x} \epsilon \mathbb{R}^n$ is the solution of (P), with $\|\cdot\| = \|\cdot\|_2$, iff there exist $y \epsilon \mathbb{R}^m$, $\xi \epsilon \mathbb{R}^n$, $\xi \geq 0$ such that

(1.3.7.1) $$u = A^T y + \xi,$$

where

(1.3.7.2) $$u = \bar{x} - a,$$

and

(1.3.7.3) $$\|u\|_2^2 = \langle y,b \rangle - \langle u,a \rangle.$$

Proof. By Theorem 1.3.5, $\bar{x}$ is the solution of (P) iff there exist $\bar{y}\in\mathbb{R}^m$, $\bar{\xi}\in\mathbb{R}^n$, $\bar{\xi} \geq 0$ such that $u^* = A^T\bar{y} + \bar{\xi}$ and $<\bar{y},b> = <u^*,\bar{x}>$. Since $\|\cdot\| = \|\cdot\|_2$, $u^* = u/\|u\|$. Setting $y = \|u\|\bar{y}$, $\xi = \|u\|\bar{\xi}$, we see that (1.3.7.1) and (1.3.7.2) hold and conversely. Also,

$$\begin{aligned} <y,b> &= <\|u\|\bar{y},b> \\ &= \|u\| <\bar{y},b> \\ &= \|u\| <u^*,\bar{x}>, \text{ by (1.3.5.1)} \\ &= <u,\bar{x}> \\ &= <u,u + a>, \end{aligned}$$

which is (1.3.7.3).

1.3.8. Theorems 1.3.1, 1.3.2, 1.3.4 and 1.3.6 correspond to Theorems 3.2, 3.4, 3.7 and 3.8 of [6] respectively. In [6] a simple geometric proof of Theorem 3.8 of [6] was sketched. In the same spirit, it would be worthwhile, to give a direct geometric proof of Theorem 1.3.6 without relying on results for general norms. Let us recall a couple of well known facts.

Fact (*i*). If K is a nonempty closed convex subset of $\mathbb{R}^n$ with $a\in\mathbb{R}^n$, then $\bar{x}\in K$ is the unique point nearest to a in K, for the Euclidean norm iff

(1.3.8.1) $\qquad <a - \bar{x}, x - \bar{x}> \leq 0, \ \forall x\in K.$

Fact (*ii*). If K is a convex cone in $\mathbb{R}^n$, then its ***negative polar*** K^o is defined by

(1.3.8.2) $\qquad K^o = \{y\in\mathbb{R}^n \mid <y,x> \leq 0, \ \forall x\in K\}.$

It will be convenient to state Theorem 1.3.6 in an equivalent form before proving it.

1.3.9 THEOREM. Assume that the set $K = \{x \epsilon \mathbb{R}^n | Ax = b,\ x \geq 0\}$ is nonempty with $a \epsilon \mathbb{R}^n \backslash K$. Then $\bar{x}$ solves (P) for the ℓ^2–norm $\|\cdot\|_2$ iff $A\bar{x} = b$, $\bar{x} \geq 0$ and $\exists y \epsilon \mathbb{R}^m$, $\xi \epsilon \mathbb{R}^n$, $\xi \geq 0$ such that

(1.3.9.1) $$\bar{x} - a = A^T y + \xi,$$

and

(1.3.9.2) $$<\xi, \bar{x}> = 0.$$

Proof. $\bar{x}$ solves (P) iff $\bar{x} \epsilon K$ is the point in K nearest to a, for the Euclidean norm $\|\cdot\|_2$. By *Fact* (*i*), this is the case iff

(1.3.9.3) $$<a - \bar{x},\ x - \bar{x}> \leq 0,\ \forall x \epsilon K.$$

Now suppose that $\bar{x}$ solves (P). Let $J = \{j \epsilon [1,n] | \bar{x}_j = 0\}$ and

(1.3.9.4) $$H = (\ker A) \cap \{v \epsilon \mathbb{R}^n | v_j \geq 0,\ \forall j \epsilon J\}.$$

Due to the fact that $\bar{x}_j > 0,\ \forall j \notin J$ and $\bar{x}_j = 0,\ \forall j \epsilon J$, one easily sees that given $v \epsilon H$, $\exists \eta > 0$ such that $\bar{x} + \eta v \epsilon K$, and so from (1.3.9.3) we see that $<a - \bar{x}, \eta v> \leq 0$; i.e.

(1.3.9.5) $$<a - \bar{x},\ v> \leq 0,\quad \forall v \epsilon H.$$

By *Fact* (*ii*), this means $a - \bar{x} \epsilon H^{\circ}$. Again, one easily sees that

(1.3.9.6) $$H^{\circ} = (\mathrm{Ker} A)^{\perp} + \{v \epsilon \mathbb{R}^n | v_j = 0,\ \forall j \notin J,\ v_j \leq 0,\ \forall j \epsilon J\}.$$

Since $(\ker A)^{\perp} = \mathrm{Im}(A^T)$, it follows that $\exists y \epsilon \mathbb{R}^m$, $\xi \epsilon \mathbb{R}^n$, $\xi \geq 0$, $\xi_j = 0,\ \forall j \notin J$ such that

(1.3.9.7) $$\bar{x} - a = A^T y + \xi,$$

which is (1.3.9.1). Since $\xi_j = 0,\ \forall j \notin J$, (1.3.9.2) also holds.

To prove the converse, assume that $A\bar{x} = b$, $\bar{x} \geq 0$ and that (1.3.9.1), (1.3.9.2) hold. We shall show that (1.3.9.3) holds $\forall x \epsilon K$, which would prove that $\bar{x}$ solves (P). Now $\forall x \epsilon K$,

$$\begin{aligned} <\bar{x} - a, x - \bar{x}> &= <A^T y + \xi, x - \bar{x}> \\ &= <A^T y, x - \bar{x}> + <\xi, x> - <\xi, \bar{x}> \\ &= <y, Ax - A\bar{x}> + <\xi, x> \end{aligned}$$

$$= \langle y, b - b \rangle + \langle \xi, x \rangle$$
$$= \langle \xi, x \rangle \geq 0,$$

which is (1.3.9.3).

1.3.10 REMARK. The condition (1.3.9.2) is equivalent to each of the following:

(1.3.10.1) $\langle y, b \rangle = \langle u, \bar{x} \rangle$

and

(1.3.10.2) $\|u\|_2^2 = \langle y, b \rangle - \langle u, a \rangle$, where $u = \bar{x} - a$.

This is so, since

$$\langle u, \bar{x} \rangle = \langle A^T y + \xi, \bar{x} \rangle = \langle A^T y, \bar{x} \rangle$$
$$= \langle y, A\bar{x} \rangle = \langle y, b \rangle,$$

and

$$\|u\|_2^2 = \langle u, u \rangle = \langle u, \bar{x} - a \rangle$$
$$= \langle u, \bar{x} \rangle - \langle u, a \rangle = \langle y, b \rangle - \langle u, a \rangle.$$

1.3.11 COROLLARY. For the ℓ^2–norm, if $\bar{x}, y, \xi$, etc. are as in Theorem 1.3.9, then $(d^{-1}y, d^{-1}\xi)$, where $d = \|\bar{x} - a\|_2$, solves the dual (P') for the ℓ^2–norm i.e.

(P'): $\max\{\langle z, b \rangle - \langle A^T z + \zeta, a \rangle \,|\, \|A^T z + \zeta\|_2 = 1,\ z \in \mathbb{R}^m, \zeta \in \mathbb{R}^n, \zeta \geq 0\}$

Proof. Due to (1.3.9), $\|A^T(d^{-1}y) + d^{-1}\xi\|_2 = 1$. By (1.3.10.1) for $u = \bar{x} - \alpha$, $\langle d^{-1}y, b \rangle = d^{-1}\langle u, \bar{x} \rangle$, whereas by (1.3.9.1)

$$\langle A^T(d^{-1}y) + d^{-1}\xi, a \rangle = d^{-1}\langle \bar{x} - a, a \rangle = d^{-1}\langle u, a \rangle.$$

Thus,

$$\langle d^{-1}y, b \rangle - \langle A^T(d^{-1}y) + d^{-1}\xi, a \rangle = d^{-1}\langle u, \bar{x} - a \rangle = d,$$

which completes the proof.

1.4 ALGORITHM

We assume $b \neq 0$ and that the system $Ax = b$, $x \geq 0$ is feasible. Here then is the algorithm for finding $\bar{x}$ solving (P)

(P): $\min\{\|x\| \,|\, Ax = b,\ x \geq 0\}$, which corresponds to $a = 0$ in Section 1.3. We assume that the norm $\|\cdot\|$ is strictly convex.

1.4.1 ALGORITHM.

Step 1. Fid x_0, the solution of the problem

$$Ax = b,\ x \geq 0,\ \|x\|_2 \text{ (min)}.$$

Set $g_0 = x_0/\|x_0\|'$, $\beta_0 = <g_0, x_0>$ and $k = 0$.

Step 2. Define $a_k = \beta_k g_k'$ and find x_{k+1} solving the problem

$$Ax = b,\ x \geq 0,\ \|x - a_k\|_2 \text{ (min)}.$$

Let $u_k = x_{k+1} - a_k$. If $u_k = 0$, STOP; x_{k+1} solves (P), *else* proceed.

Step 3. Let $\gamma_k = <u_k, x_{k+1}>$. If $\gamma_k \leq 0$, set $\bar{\alpha}_k = \beta_k/(\beta_k - \gamma_k)$ and

GO TO *Step 5*, *else* set $\bar{\alpha}_k = 1$ and proceed.

Step 4. If

$$\gamma_k <u', g_k> \geq \beta_k \|u_k\|'$$

set $\alpha_k = 1$ and GO TO *Step 6*, *else* proceed.

Step 5. Find, if exists, $\alpha_k \epsilon (0, \bar{\alpha}_k)$ such that

$$\begin{aligned}&(\alpha_k \gamma_k + (1 - \alpha_k)\beta_k)\\ &\qquad < (\alpha_k u_k + (1 - \alpha_k) g_k)', u_k - g_k>\\ &\qquad = (\gamma_k - \beta_k)\|\alpha_k u_k + (1 - \alpha_k) g_k\|'.\end{aligned}$$

(It will be proven that such an α_k always exists and it is unique, if the norm $\|\cdot\|$ is also smooth).

Step 6. Define

$$g_{k+1} = (\alpha_k u_k + (1 - \alpha_k) g_k)/\|\alpha_k u_k + (1 - \alpha_k) g_k\|'$$

and

$$\beta_{k+1} = (\alpha_k \gamma_k + (1 - \alpha_k)\beta_k)/\|\alpha_k u_k + (1 - \alpha_k) g_k\|'.$$

Increase k by 1 and RETURN to *Step 2.*

1.4.2 REMARK. As it can easily be seen from the subsequent proof of convergence of the above algorithm, *Step 1* is only a convenient initialization. In fact, it can be replaced by

Step 1. Pick any g_0 such that there are $y_0 \epsilon \mathbf{R}^m$, $\xi_0 \epsilon \mathbf{R}^n$, $\xi_0 \geq 0$ with

$$g_0 = A^T y_0 + \xi_0,\ \|g_0\|' = 1 \quad \text{and} \quad \langle y_0, b\rangle \geq 0.$$

1.5 FEASIBILITY OF THE ALGORITHM

In this section we show that the various steps in the algorithm are properly formulated.

1.5.1 LEMMA. Let g_0, etc. be as in *Step 1* of the algorithm. Then $\exists y_0 \epsilon \mathbf{R}^m$, $\xi_0 \epsilon \mathbf{R}^n$, $\xi_0 \geq 0$ such that

$$g_0 = A^T y_0 + \xi_0,\ \|g_0\|' = 1$$

and

$$\beta_0 = \langle y_0, b\rangle > 0.$$

Proof. By Theorem 1.3.9, $\exists z \epsilon \mathbf{R}^m$, $\zeta \epsilon \mathbf{R}^n$, $\zeta \geq 0$ such that $x_0 = A^T z + \zeta$ and $\langle \zeta, x_0\rangle = 0$. Let $y_0 = z/\|x_0\|'$ and $\xi_0 = \zeta/\|x_0\|'$. By *Step 1* of the algorithm, $g_0 = x_0/\|x_0\|' = A^T y_0 + \xi_0$.

Also

$$\beta_0 = \langle g_0, x_0\rangle = \|x_0\|_2^2/\|x_0\|' > 0$$

and

$$\beta_0\ \langle g_0, x_0\rangle \qquad = \langle A^T y_0 + \xi_0, x_0\rangle$$

$$= \langle y_0, Ax_0 \rangle = \langle y_0, b \rangle.$$

This completes the proof of the Lemma.

1.5.2 LEMMA. In the algorithm $\forall k \geq 1$, if $u_{k-1} \neq 0$, then

$$\alpha u_{k-1} + (1-\alpha) g_{k-1} \neq 0, \quad \forall \alpha \epsilon [0,1].$$

Furthermore, in this case, if we also assume that α_{k-1} has been determined, then

$\exists y_k \epsilon \mathbf{R}^m$, $\xi_k \epsilon \mathbf{R}^n$, $\xi_k \geq 0$ such that

$$g_k = A^T y_k + \xi_k, \quad \|g_k\|' = 1,$$

and

$$\beta_k = \langle y_k, b \rangle.$$

Proof. We shall prove this lemma by induction on k. Take $k = 1$. Since $u_0 \neq 0$, with $\|g_0\|' = 1$, $\alpha u_0 + (1-\alpha) g_0 = 0$ cannot hold for $\alpha = 1$ or $\alpha = 0$. So, if $\exists \alpha \epsilon [0,1]$, such that $\alpha u_0 + (1-\alpha) g_0 = 0$, then $\alpha \epsilon (0,1)$. Assume that such an α exists. We see that

$$\begin{aligned}
\|u_0\|_2^2 &= \|x_1 - a_0\|_2^2 = \langle x_1 - a_0, u_0 \rangle \\
&= \alpha^{-1} (\alpha - 1) \langle x_1 - a_0, g_0 \rangle \\
&= \alpha^{-1} (\alpha - 1) (\langle x_1, g_0 \rangle - \langle a_0, g_0 \rangle) \\
&= \alpha^{-1} (\alpha - 1) (\langle x_1, A^T y_0 + \xi_0 \rangle - \beta_0), \\
&\qquad \text{since } a_0 = \beta_0 g_0', \\
&= \alpha^{-1} (\alpha - 1) (\beta_0 + \langle x_1, \xi_0 \rangle - \beta_0), \\
&\qquad \text{since } \langle x_1, A^T y_0 \rangle = \langle b, y_0 \rangle, \\
&= \alpha^{-1} (\alpha - 1) \langle x_1, \xi_0 \rangle \leq 0,
\end{aligned}$$

contradicting our assumption that $u_0 \neq 0$.

Since x_1 is the solution of the problem

$$Ax = b, \ x \geq 0, \ \|x - a_0\|_2 \ (\min),$$

by Theorem 1.3.9, $\exists z \epsilon \mathbf{R}^m$, $\zeta \epsilon \mathbf{R}^n$, $\zeta \geq 0$ such that

$$u_0 = x_1 - a_0 = A^T z + \zeta \quad \text{and} \quad <z,b> = <u_0,x_1> = \gamma_0.$$

Since $\alpha u_0 + (1 - \alpha)g_0 \neq 0$, $\forall \alpha\epsilon[0,1]$, and since *Steps 4* and *5* of the algorithm have determined α_0 by assumption, we have $\alpha_0 u_0 + (1 - \alpha_0)g_0 \neq 0$. So *Step 6* is well defined. Define y_1, ξ_1 by

$$y_1 = (\alpha_0 z + (1 - \alpha_0)y_0)/\|\alpha_0 u_0 + (1 - \alpha_0)g_0\|'$$

and

$$\xi_1 = (\alpha_0 \zeta + (1 - \alpha_0)\xi_0)/\|\alpha_0 u_0 + (1 - \alpha_0)g_0\|'.$$

Then $\xi_1 \geq 0$, and

$$A^T y_1 + \xi_1 = (\alpha_0 u_0 + (1 - \alpha_0)g_0)/\|\alpha_0 u_0 + (1 - \alpha_0)g_0\|'.$$

Furthermore,

$$\begin{aligned} <y_1,b> &= (\alpha_0 <z,b> + (1 - \alpha_0)<y_0,b>)/\|\alpha_0 u_0 + (1 - \alpha_0)y_0\|' \\ &= \beta_1, \text{ by } \textit{Step 6} \text{ of the algorithm.} \end{aligned}$$

Now we have verified the lemma for $k = 1$. For the general index k, the argument is identical and is obtained from above by simply replacing y_0, ξ_0, g_0, u_0, β_0 and x_1 by y_{k-1}, ξ_{k-1}, g_{k-1}, u_{k-1}, β_{k-1} and x_k, respectively. This completes the proof of this lemma.

1.5.3 LEMMA. Let $u,y\epsilon\mathbb{R}^n$, $\gamma,\beta\epsilon\mathbb{R}$ and $Z = \{\alpha\epsilon\mathbb{R} \mid \alpha u + (1 - \alpha)y = 0\}$. Also let $\|\cdot\|$ be a strictly convex norm on $\mathbb{R}^n$. Define the function

$$\varphi(\alpha) = (\alpha\gamma + (1 - \alpha)\beta)/\|\alpha u + (1 - \alpha)y\|'$$

for $\alpha\epsilon\mathbb{R}\backslash Z$.

Then, we have for the derivative of φ

$$\varphi'(\alpha) = \frac{\gamma-\beta}{\|\alpha u+(1-\alpha)y\|'} - \frac{(\alpha\gamma+(1-\alpha)\beta)<(\alpha u+(1-\alpha)y)',u-y>}{\|\alpha u+(1-\alpha)y\|'^2} \tag{1.5.3.1}$$

for $\alpha\epsilon\mathbb{R}\backslash Z$.

Now let $\alpha\epsilon\mathbb{R}\backslash Z$ be such that $\varphi(\alpha) > 0$.

Then we have

(1.5.3.2) If $\varphi'(\alpha) \geq 0$ then $\varphi(\lambda) \leq \varphi(\alpha)$ $\forall\lambda < \alpha$, $\lambda\epsilon\mathbb{R}\backslash Z$.

(1.5.3.3) If $\varphi'(\alpha) \leq 0$ then $\varphi(\lambda) \leq \varphi(\alpha)$ $\forall\lambda > \alpha$, $\lambda\epsilon\mathbb{R}\backslash Z$.

If the norm $\|\cdot\|$ is also smooth, then

(1.5.3.4) $\varphi(\lambda) < \varphi(\alpha)$ in both (1.5.3.2) and (1.5.3.3).

Proof. The proof is a straightforward repetition for the above φ of the proofs of Lemmas 5.3, 5.4, 5.5 and 5.6 in Sreedharan [6]. Note that, as one easily sees, in Lemma 5.5 of [6] we can replace the assumption that y and h are linearly independent by the weaker one that both $y + \alpha h$ and $y + \lambda h$ are nonzero. We have used this weaker assumption for the above Lemma. (1.5.3.1) corresponds to relation (5.4.3) of Lemma 5.4 in [6] and (1.5.3.2), (1.5.3.3) correspond respectively to statements (i) and (ii) of Lemma 5.6 in [6].

1.5.4 LEMMA. Let $u,y\epsilon\mathbb{R}^n$, $\gamma,\beta\epsilon\mathbb{R}$. Also let $\|\cdot\|$ be a strictly convex norm on $\mathbb{R}^n$. Assume that $\alpha u + (1 - \alpha)y \neq 0$, $\forall\alpha\epsilon[0,1]$. Then define

$$\varphi(\alpha) = (\alpha\gamma + (1 - \alpha)\beta)/\|\alpha u + (1 - \alpha)y\|' \quad \text{for} \quad \alpha\epsilon[0,1].$$

Also assume that

$$\varphi(0) > 0 \quad \text{and} \quad \varphi'(0) > 0.$$

Consider the following algorithm:

(i) If $\gamma \leq 0$ then $\overline{\alpha} = \beta/(\beta - \gamma)$ and GO TO (iv).

Else proceed.

(ii) $\overline{\alpha} = 1$

(iii) If

$$\gamma\langle u',y\rangle \geq \beta\|u\|'$$

set $\alpha = 1$ and STOP. Else proceed.

(iv) Find $\alpha\epsilon(0,\overline{\alpha})$ such that

$$(\alpha\gamma + (1 - \alpha)\beta)<(\alpha u + (1 - \alpha)y)',u - y> =$$
$$= (\gamma - \beta) \|\alpha u + (1 - \alpha)y\|'$$

STOP.

Then this algorithm is well formulated and it produces an α which is a global maximizer of φ on the interval [0,1]. Moreover, if the norm $\|\cdot\|$ is also smooth, then φ has only one global maximizer on [0,1]. If we replace the assumption $\varphi(0) > 0$ by $\varphi(0) = 0$ and $\gamma > 0$ then the Lemma is still true.

Proof. Since $\alpha u + (1 - \alpha)y \neq 0, \forall\alpha\epsilon[0,1]$, φ is continuous over the compact interval [0,1]; thus a global maximizer exists. From (1.5.3.1) we see that φ' is also continuous over [0,1].

It is easy to verify that the following are true: from the hypothesis that $\varphi(0) > 0$, which is the same as $\beta > 0$, and the definition of φ we have that

$$\exists\ \overline{\alpha}\epsilon(0,1] \text{ such that } \varphi(\overline{\alpha}) = 0$$
$$\text{iff}$$
$$\gamma \leq 0.$$

If this is true, i.e. if $\gamma \leq 0$, then $\overline{\alpha} = \beta/(\beta - \gamma)$ and it is unique. Using this and (ii), we see that we always have $0 < \overline{\alpha} \leq 1$.

From (1.5.3.1) we get that the relation in (iii) is true iff $\varphi'(1) \geq 0$ and the relation in (iv) is true iff $\varphi'(\alpha) = 0$.

Note that writing (iii) makes sense because the assumption $\alpha u + (1 - \alpha)y \neq 0, \forall\alpha\epsilon[0,1]$ guarantees that $u \neq 0$. Note also that $\varphi'(1)$ is defined iff $u \neq 0$. Also the relation in (iv) makes sense for all $\alpha\epsilon(0,\overline{\alpha})$ because $0 < \overline{\alpha} \leq 1$ and thus $\alpha u + (1 - \alpha)y \neq 0$. This is the same as saying that $\varphi'(\alpha)$ is defined for all $\alpha\epsilon(0,\overline{\alpha})$.

Using the above facts we can rewrite the algorithm of the above Lemma as follows:

(i) If exists, then find $\overline{\alpha}\epsilon(0,1]$ such that $\varphi(\overline{\alpha}) = 0$ and go to (iv). If there's no such $\overline{\alpha}$, proceed.

(ii) Put $\overline{\alpha} = 1$

(iii) If $\varphi'(1) \geq 0$, then $\alpha = 1$ and STOP. If $\varphi'(1) < 0$, proceed.

(iv) Find $\alpha\epsilon(0,\overline{\alpha})$ such that

$$\varphi'(\alpha) = 0$$

and STOP.

We distinguish two cases:

Case 1. Suppose $\exists\overline{\alpha}\epsilon(0,1]$ such that $\varphi(\overline{\alpha}) = 0$. Since this $\overline{\alpha}$ is unique and since $\varphi(0) > 0$, then

(1.5.4.1) $$\varphi(\alpha) > 0, \quad \forall\alpha\epsilon[0,\overline{\alpha}).$$

If it were true that

$$\varphi'(\alpha) \geq 0, \ \forall\alpha\epsilon[0,\overline{\alpha}], \text{ then}$$

φ would be increasing over $[0,\overline{\alpha}]$ and thus

$$0 = \varphi(\overline{\alpha}) \geq \varphi(0) > 0 \quad \text{which is a contradiction.}$$

So, $\exists\tilde{\alpha}\epsilon[0,\overline{\alpha}]$ such that

$$\varphi'(\tilde{\alpha}) < 0.$$

Since $\varphi'(0) > 0$, then

$$\exists\alpha\epsilon(0,\tilde{\alpha}) \text{ such that } \varphi'(\alpha) = 0.$$

Since $\tilde{\alpha} \leq \overline{\alpha}$, $\alpha\epsilon(0,\overline{\alpha})$. By (1.5.4.1) we also have $\varphi(\alpha) > 0$. Now (1.5.3.2) and (1.5.3.3) of Lemma 1.5.3 imply that this α is a global maximizer of φ on $[0,1]$ and it is unique, if the norm $\|\cdot\|$ is smooth, by (1.5.3.4).

Case 2. Suppose there is no $\overline{\alpha}\epsilon(0,1]$ such that $\varphi(\overline{\alpha}) = 0$. Then $\overline{\alpha} = 1$ and since $\varphi(0) > 0$, we have

(1.5.4.2) $$\varphi(\alpha) > 0, \quad \forall\alpha\epsilon[0,1].$$

Now if $\varphi'(1) < 0$, then

$$\exists \alpha \epsilon (0,1) \text{ such that } \varphi'(\alpha) = 0$$

because $\varphi'(0) > 0$. Also $\varphi(\alpha) > 0$ by (1.5.4.2). As in *Case 1*, we now conclude that this α is a global maximizer of φ on [0,1] and it is unique if the norm $\|\cdot\|$ is smooth.

Now suppose $\varphi'(1) \geq 0$. From (1.5.4.2) we have $\varphi(1) > 0$. Now (1.5.3.2) of Lemma 1.5.3 implies that 1 is a global maximizer of φ on [0,1]. If the norm $\|\cdot\|$ is smooth, then this global maximizer is unique by (1.5.3.4).

Now suppose that $\varphi(0) = 0$, i.e. $\beta = 0$, and $\gamma > 0$. Then (i) is not executed and $\bar{\alpha} = 1$.

Also $\varphi(0) = 0$ and $\varphi(\alpha) > 0$, $\forall \alpha \epsilon (0,1]$.

Suppose the criterion of (iii) is satisfied, i.e. $\varphi'(1) \geq 0$. Since $\varphi(1) > 0$, we have that 1 is a global maximizer of φ on [0,1] via (1.5.3.2) of Lemma 1.5.3. If $\varphi'(1) < 0$, and since $\varphi'(0) > 0$, then $\exists \alpha \epsilon (0,1)$ s.t. $\varphi'(\alpha) = 0$ and this α is a global maximizer of φ on [0,1] via Lemma 1.5.3. This α is unique if the norm $\|\cdot\|$ is also smooth via Lemma 1.5.3.

So α is picked to be a global maximizer of φ on [0,1] and it is unique if the norm $\|\cdot\|$ is also smooth, in all cases.

1.5.5 LEMMA. Let $k \geq 0$. Suppose that at the kth iteration of Algorithm 1.4.1 $u_k \neq 0$. Then, by Lemma 1.5.2, the function

$$\varphi_k(\alpha) = (\alpha\gamma_k + (1-\alpha)\beta_k)/\|\alpha u_k + (1-\alpha)g_k\|' \quad \text{where} \quad \gamma_k, \beta_k,$$

u_k, g_k are as in Algorithm 1.4.1, is finite over the interval [0,1] since the denominator does not vanish.

Also assume that $\beta_k > 0$.

Then,

i) the α_k specified by Algorithm 1.4.1 in the Steps from 3 to 5, is a global maximizer of φ_k on [0,1]. α_k is unique if the norm $\|\cdot\|$ is also smooth.

ii) $\beta_{k+1} > \beta_k$.

Proof. We have $\varphi_k(0) = \beta_k > 0$ by assumption.

Also

$$\begin{aligned}
\varphi_k'(0) &= \gamma_k - \beta_k - \beta_k < g_k', u_k - g_k>, \quad \text{by (1.5.3.1)} \\
&= \gamma_k - \beta_k < g_k', u_k> \\
&= <u_k, x_{k+1}> - <a_k, u_k> \\
&= \|u_k\|_2^2 > 0 \quad \text{since } u_k \neq 0 \text{ by assumption.}
\end{aligned}$$

Now (i) follows from Lemma 1.5.4 and *Steps 3,4,5* of Algorithm 1.4.1. Since α_k is a global maximizer of φ_k on [0,1], we have $\varphi_k(\alpha_k) \geq \varphi_k(0) = \beta_k$; but in addition to this, we know that $\varphi_k'(0) > 0$ and thus $\varphi_k(\alpha_k) > \varphi_k(0) = \beta_k$. From *Step 6* of Algorithm 1.4.1 we have that $\beta_{k+1} = \varphi_k(\alpha_k)$ and this completes the proof of (ii).

1.5.6 LEMMA. The sequence (β_k) generated by Algorithm 1.4.1 is positive and strictly increasing.

Proof. From Lemma 1.5.5 we have that for $u_k \neq 0$, if $\beta_k > 0$ then $\beta_{k+1} > \beta_k$ and thus $\beta_{k+1} > 0$, too. In other words: $\beta_k > 0$ implies $\beta_{k+1} > 0$, unless the algorithm is terminated at the kth iteration. From Lemma 1.5.1 we have $\beta_0 > 0$. So by induction, the finite or infinite sequence (β_k) is positive and thus, by Lemma 1.5.5 (ii), it is also strictly increasing. So, if it is not a finite sequence, it converges to some limit which, by Lemma 1.5.2, is less than or equal to the value of problem (P):

$$\min\{\|x\| \mid Ax = b,\ x \geq 0\},$$ or the value of the problem (P') dual to (P).

1.5.7 THEOREM. If u_k at the kth iteration of Algorithm 1.4.1 equals zero, then x_{k+1} solves (P):

$$\min\{\|x\| \mid Ax = b, x \geq 0\}.$$

Proof. If $u_k = 0$, then we have by *Step 2* of the algorithm,

$$x_{k+1} = \beta_k g_k' \quad \text{and} \quad x_{k+1} \text{ is feasible for (P).}$$

The last relation via Lemmas 1.5.2, 1.5.5 and 1.5.6 becomes

$$x_{k+1} = \langle y_k, b\rangle(A^T y_k + \xi_k)', \quad \|A^T y_k + \xi_k\|' = 1.$$

Now apply Theorem 1.3.2 for $a = 0$ to get the result. Take into account that $\beta_k > 0$ by Lemma 1.5.6 and that $b \neq 0$ which was assumed in Section 1.4.

1.6 CONVERGENCE OF THE ALGORITHM

We can now prove convergence of Algorithm 1.4.1.

1.6.1 THEOREM. Assume that the norm $\|\cdot\|$ on $\mathbb{R}^n$ is strictly convex. Then the sequence (x_k) generated by Algorithm 1.4.1 either terminates at or converges to the unique solution of (P) i.e. of the problem

(P): $$Ax = b, \ x \geq 0, \ \|x\| \ (\min).$$

Proof. If Algorithm 1.4.1 terminates at x_k then $u_{k-1} = 0$, in which event this theorem reduces to Theorem 1.5.7. So consider the case when we have a genuine infinite sequence (x_k). We shall first show that the sequence (u_k) constructed in Algorithm 1.4.1 converges to zero.

Denote by d the value of (P). Then $\beta_k = \langle y_k, b\rangle \leq d$, $\forall k$. By Lemma 1.5.6 (β_k) is a strictly increasing sequence, and so $\beta_k \uparrow \beta \leq d$. Fix $\hat{x} \geq 0$, with $A\hat{x} = b$. We have

$$\begin{aligned} \|x_{k+1}\|_2 &\leq \|x_{k+1} - a_k\|_2 + \|a_k\|_2 \\ &\leq \|\hat{x} - a_k\|_2 + \|a_k\|_2 \end{aligned}$$

$$\leq \|\hat{x}\|_2 + 2\|a_k\|_2$$
$$= \|\hat{x}\|_2 + 2\beta_k\|g_k'\|_2$$
$$\leq \|\hat{x}\|_2 + 2\beta M\|g_k'\|$$
$$= \|\hat{x}\|_2 + 2\beta M.$$

This shows that the sequence (x_k) is bounded.

It follows that the sequence $u_k = x_{k+1} - a_k$ is also bounded. So if (u_k) does not converge to zero, there is a subsequence $(u_{k'})$ of (u_k) converging to $u \neq 0$.

Observe that

$$\|u_k\|_2^2 = \langle x_{k+1} - a_k, u_k\rangle$$
$$= \langle x_{k+1}, u_k\rangle - \langle a_k, u_k\rangle$$
$$\text{(1.6.1.1)} \qquad = \gamma_k - \beta_k\langle g_k', u_k\rangle.$$

Due to the boundedness of (γ_k) and (g_k), we can pass to a further subsequence, again denoted by (k') such that $\gamma_{k'} \to \gamma$ and $g_{k'} \to g$. Since $\|g_k\|' = 1$, $\forall k$, $\|g\|' = 1$ and by the continuity of the map $z \mapsto z'$ on $\mathbb{R}^n\backslash\{0\}$, we get from (1.6.1.1), as $k' \to \infty$,

$$\text{(1.6.1.2)} \qquad \|u\|_2^2 = \gamma - \beta \langle g', u\rangle.$$

We now distinguish two cases.

Case 1. There exist an infinity of indices k' for which *Step 5* is executed. Denote this subsequence by (k') again. Furthermore we may assume that the sequence $(\alpha_{k'})$ is such that $\alpha_{k'} \to \alpha\epsilon[0,1]$. Let us also assume that $\alpha u + (1 - \alpha)g \neq 0$. We shall take care of the possibility $\alpha u + (1 - \alpha)g = 0$ shortly. Using *Step 5* of the algorithm and allowing $k' \to \infty$, since $\alpha u + (1 - \alpha)g \neq 0$, $(\alpha u_{k'} + (1 - \alpha)g_{k'})' \to (\alpha u + (1 - \alpha)g)'$. This yields the equation

$$\text{(1.6.1.3)} \qquad (\alpha\gamma + (1 - \alpha)\beta) \langle (\alpha u + (1 - \alpha)g)', u - g\rangle$$
$$= (\gamma - \beta)\ \|\alpha u + (1 - \alpha)g\|'.$$

Also by *Step 6* of the algorithm

$$(1.6.1.4) \qquad \beta\|\alpha u + (1 - \alpha)g\|' = \alpha\gamma + (1 - \alpha)\beta$$

Since $\alpha u + (1 - \alpha)g \neq 0$, combining (1.6.1.3) and (1.6.1.4), we get

$$(1.6.1.5) \qquad \beta \langle(\alpha u + (1 - \alpha)g)', u - g\rangle = \gamma - \beta.$$

Since $\langle(\alpha u + (1 - \alpha)g)', g + \alpha(u - g)\rangle = \|\alpha u + (1 - \alpha)g\|'$, we have

$$(1.6.1.6) \qquad \alpha \langle(\alpha u + (1 - \alpha)g)', u - g\rangle = \|\alpha u + (1 - \alpha)g\|' - \langle(\alpha u + (1 - \alpha)g)', g\rangle.$$

Combining (1.6.1.6), (1.6.1.5) and (1.6.1.4) we get the equation

$$(1.6.1.7) \qquad \beta = \beta \langle(\alpha u + (1 - \alpha)g)', g\rangle,$$

i.e.

$$(1.6.1.8) \qquad \langle(\alpha u + (1 - \alpha)g)', g\rangle = 1, \text{ since } \beta > 0.$$

Since $\|g\|' = 1$, with the norm $\|\cdot\|$ strictly convex, (1.6.1.8) imples that

$$(1.6.1.9) \qquad (\alpha u + (1 - \alpha)g)' = g'.$$

Inserting this in (1.6.1.5) gives us the equation

$$(1.6.1.10) \qquad \beta \langle g', u - g\rangle = \gamma - \beta,$$

i.e.

$$(1.6.1.11) \qquad \beta \langle g', u\rangle = \gamma, \text{ since } \langle g', g\rangle = 1,$$

which in view of (1.6.1.2), implies $u = 0$, a contradiction.

Now suppose that $\alpha u + (1 - \alpha)g = 0$. Since $g \neq 0$, $\alpha \neq 0$. So $u = \alpha^{-1}(\alpha - 1)g$. By *Step 6* of the algorithm

$$\beta_{k'+1}\|\alpha_{k'}u_{k'} + (1 - \alpha_{k'})g_{k'}\|' = \alpha_{k'}\gamma_{k'} + (1 + \alpha_{k'})\beta_{k'}.$$

Allowing $k' \to \infty$, we get

$$(1.6.1.12) \qquad \alpha\gamma + (1 - \alpha)\beta = 0.$$

Writing $u = \alpha^{-1}(\alpha - 1)g$ in equation (1.6.1.2) yields

$$(1.6.1.13) \qquad \|u\|_2^2 = \gamma - \beta\alpha^{-1}(\alpha - 1) = 0, \text{ by (1.6.1.12)},$$

once more a contradiction.

Case 2. There exist an infinity of indices k' for which *Step 4* of the algorithm is answered affirmatively. Passing to a subsequence, again denoted (k'), we assume this to be the case for all k'. Then by *Step 6* of the althorithm,

$$\beta_{k'+1} = \gamma_{k'}/\|u_{k'}\|'.$$

Since (β_k) is an increasing sequence with limit $\beta > 0$, $(\beta_{k'+1}) \to \beta$. Let $\gamma_{k'} \to \gamma$. Then $\gamma > 0$ and

$$\gamma = \beta\|u\|'. \tag{1.6.1.14}$$

Due to *Step 4*

$$\gamma_{k'} < u'_{k'}, g_{k'}> \geq \beta_{k'}\|u_{k'}\|',$$

which in the limit yields

$$\gamma <u',g> \geq \beta\|u\|' = \gamma.$$

Since $\gamma > 0$, this shows that $<u',g> \geq 1$. But since $\|g\|' = 1$ with $\|u'\| = 1$, we conclude that $<u',g> = 1$ and hence $u' = g'$. Using this in (1.6.1.2) we get

$$\|u\|_2^2 = \gamma - \beta<u',u>$$

$$= \gamma - \beta\|u\|' = 0, \text{ by } (1.6.1.14) \tag{1.6.1.15}$$

contradicting the assumption $u \neq 0$. So we have shown that in all cases $u_k \to 0$.

To show that (x_k) converges to the solution of (P), we first show that every cluster point $\bar{x}$ of (x_k) is a solution of (P). Then since solutions of (P) are unique, with (x_k) bounded, we can conclude that the sequence (x_k) converges to the unique solution $\bar{x}$ of problem (P). Let then $\bar{x}$ be any cluster point of (x_k). Then there is a subsequence (k') such that $x_{k'+1} \to \bar{x}$. Recall that by Lemmas 1.5.2, 1.5.5 and 1.5.6 $\exists y_k \in \mathbb{R}^m$, $\xi_k \in \mathbb{R}^n$, $\xi_k \geq 0$ such that

$$g_k = A^T y_k + \xi_k, \quad \|g_k\|' = 1, \tag{1.6.1.16}$$

and

(1.6.1.17) $\qquad \beta_k = \langle y_k, b\rangle, \quad \forall k.$

Furthermore, (β_k) converges to $\beta > 0$.

By passing to a further subsequence of (k'), if necessary, and denoting the new subsequence again by (k'), we can assume that

$g_{k'} \to g$, $\|g\|' = 1$. By *Step 2* of the algorithm $u_k = x_{k+1} - a_k$, and so

(1.6.1.18) $\qquad x_{k'+1} = u_{k'} + a_{k'} = u_{k'} + \beta_{k'} g'_{k'}.$

Allowing $k' \to \infty$, due to the strict convexity of the norm $\|\cdot\|$, we get

(1.6.1.19) $\qquad \bar{x} = \beta g', \quad \|g\|' = 1.$

Since $Ax_{k+1} = b$, with $x_{k+1} \geq 0$, $\forall k$, we get $A\bar{x} = b$, $\bar{x} \geq 0$.

If $y \in \mathbb{R}^m$ and $\xi \in \mathbb{R}^n$, $\xi \geq 0$ is such that $\|A^T y + \xi\|' \leq 1$, by the weak duality principle,

(1.6.1.20) $\qquad \langle y, b\rangle \leq \|\bar{x}\| = \beta,$ by (1.6.1.19).

But as observed above, $\beta > 0$ is the limit of the sequence $(\langle y_k, b\rangle)$, $y_k \in \mathbb{R}^m$, $\xi_k \in \mathbb{R}^n$, $\xi_k \geq 0$, $\|A^T y_k + \xi_k\|' = 1$. So equality holds in (1.6.1.20), showing that $\bar{x}$ is a solution of (P), thus completing the proof of the theorem.

CHAPTER 2

2.1 INTRODUCTION

In this chapter we assume that the system of m real linear equations in n unknowns $Ax = b$ has a non-negative solution. We give another implementable iterative algorithm converging to the least norm $\|\cdot\|$ solution of $Ax = b$, $x \geq 0$ for a strictly convex norm $\|\cdot\|$ on $\mathbb{R}^n$. This algorithm is given as a special case of a more general algorithm and it is analogous to the one of Chapter 1, but it is never used for actually solving the above problem since the algorithm of Chapter 1 is a better algorithm for this purpose. We include it because the theorem about its convergence will be heavily used in Chapter 3 where we prove the convergence of a more general algorithm. Some Lemmas and Definitions, which will be used in Chapter 3, are also given.

2.2 NOTATION AND SOME PRELIMINARIES

Besides the notation in Section 1.2 we also need the following:

A *convex polytope* in $\mathbb{R}^n$ is the convex hull of a finite (at least one) number of points in $\mathbb{R}^n$.

Let C be a convex cone in $\mathbb{R}^n$. Then its *negative polar*, denoted by C^o, is defined by

$$C^o = \{y \in \mathbb{R}^n \mid \langle y,x \rangle \leq 0, \forall x \in C\}.$$

A *finitely generated convex cone* in $\mathbb{R}^n$ is a set of the form

$$\{\lambda_1 x_1 + \ldots + \lambda_m x_m \mid \lambda_i \geq 0, \forall\ i = 1,\ldots,m\},$$

where $x_i \in \mathbb{R}^n$, $\forall\ i = 1,\ldots,m$. A finitely generated convex cone is a convex cone. We also define $\mathbb{R}^n_+$ by $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$.

We shall also use the symbol :=. Thus $x := y$ will mean x is defined by y.

2.3 SOME DUALITY THEORY

Let $K := \{x \in \mathbb{R}^n | Ax = b, x \geq 0\}$. As already stated, we have assumed that K is non-empty. Then, according to Theorem 19.1 in [5], K is the sum of a convex polytope P and a finitely generated convex cone C. We have the following lemma.

2.3.1 LEMMA. Let $K = P + C$ with K, P and C as above. Then $C = (\mathrm{Ker}A) \cap \mathbb{R}^n_+$, where $\mathrm{Ker}A$ is the kernel of A. This shows that in the $P + C$–representation of K, C is unique and independent of b.

Proof. Let $p \in P$ and $c \in C$; then $p + \lambda c \in P + C = K, \forall \lambda \geq 0$ and thus $p \in K$ and $p + c \in K$. By the definition of K we have

(2.3.1.1) $$Ap = A(p+c) = b$$

and

(2.3.1.2) $$p + \lambda c \geq 0, \forall \lambda \geq 0.$$

(2.3.1.1) implies $c \in \mathrm{Ker}A$ and (2.3.1.2) implies $c \geq 0$. So, C is a subset of $(\mathrm{Ker}A) \cap \mathbb{R}^n_+$.

Conversely, let $k \in (\mathrm{Ker}A) \cap \mathbb{R}^n_+$ and $p \in P$. Then $p \in K = P + C$. Let $\lambda > 0$. $p \in K$ implies $Ap = b$, $p \geq 0$. Then $p + \lambda k \geq 0$ and $A(p + \lambda k) = Ap = b$ and thus $p + \lambda k \in K, \forall \lambda > 0$. Since $K = P + C$ we see that $\exists\ p_\lambda \in P, c_\lambda \in C$ such that $p + \lambda k = p_\lambda + c_\lambda$ or $p - p_\lambda = c_\lambda - \lambda k = \lambda(\frac{c_\lambda}{\lambda} - k)$. Letting $\lambda \to +\infty$ we see that $\lambda(\frac{c_\lambda}{\lambda} - k)$ is bounded and thus $\frac{c_\lambda}{\lambda} - k$ has to go to zero. Since

$\dfrac{c_\lambda}{\lambda} \in C$, $\forall\lambda > 0$, we have that $k \in \overline{C} = C$ because a finitely generated convex cone is closed. This completes the proof.

Let us denote by F the negative of the negative polar of the above cone C, i.e. $F: = -C^{o}$. Note that

$$\begin{aligned} F &= -(\operatorname{Ker}A \cap \mathbb{R}^n_+)^{o} \\ &= -(\operatorname{Im}A^{\tau} + \mathbb{R}^n_-) = \operatorname{Im}A^{\tau} + \mathbb{R}^n_+ \\ &\qquad\qquad = \{A^{\tau}y + \zeta \,|\, y \in \mathbb{R}^m, \zeta \in \mathbb{R}^n, \zeta \geq 0\} \end{aligned}$$

and thus $F \neq \{0\}$. We have the following lemma.

2.3.2 LEMMA. Assume that $K: = \{x \in \mathbb{R}^n \,|\, Ax = b, x \geq 0\}$ is non–empty. Then, the $\inf\{<y,x> | x \in K\}$ is finite, i.e. it is not $-\infty$, if and only if $y \in F$, in which case the infimum is attained.

Proof. Let P,C as in Lemma 2.3.1. Let $y \in \mathbb{R}^n$. Let α_1 and α_2 be real numbers such that

$$-\infty < \alpha_2 \leq <y,p> \leq \alpha_1 < +\infty, \forall p \in P.$$

We have,

$$\begin{aligned} \alpha_1 + \inf\{<y,c> | c \in C\} &= \inf\{\alpha_1 + <y,c> | c \in C\} \\ &\geq \inf\{<y,p+c> | p \in P, c \in C\} \\ &= \inf\{<y,x> | x \in K\} \\ &\geq \inf\{\alpha_2 + <y,c> | c \in C\} \\ &= \alpha_2 + \inf\{<y,c> | c \in C\}. \end{aligned}$$

From these relations, it is easy to see that

$\inf\{<y,x> | x \in K\}$ is finite, i.e. it is not $-\infty$, if and only if $\inf\{<y,c> | c \in C\}$ is finite.

Because C is a cone, we have that

$\inf\{\langle y,c\rangle \mid c\in C\}$ is finite if and only if $\langle y,c\rangle \geq 0$, $\forall c\in C$, i.e. if and only if $y\in F = -C^{\circ}$.

Whenever the infimum is finite, then due to the polyhedral convexity of the set K, we see that the infimum is actually attained.

2.3.3. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let $a\in\mathbb{R}^n$. With the "primal" problem (P)

(P) $\min\{\|x-a\| \mid x\in\mathbb{R}^n,\ Ax = b,\ x \geq 0\}$

we associate a "dual" problem (P')

(P') $\max\{-\langle y,a\rangle + \min\{\langle y,x\rangle \mid x\in\mathbb{R}^n, Ax = b,\ x \geq 0\} \mid y\in\mathbb{R}^n,\ \|y\|' = 1,\ y\in F\}$.

Note that the objective function of (P') makes sense for $y\in F$ because of Lemma 2.3.2. Now we have the following basic theorem.

2.3.4 THEOREM. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$ and $a\in\mathbb{R}^n$. Assume that $K := \{x\in\mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ is non-empty and that $a\notin K$. Then the above problems (P) and (P') have the same value.

Proof. The value of (P) equals $\inf\{\|x-a\| \mid x\in K\}$ because K is a non-empty closed set and the objective function has non-negative values: so the inf is min. Now from Nirenberg [4, page 39] we have, since K is convex,

$$\text{value of (P)} = \max_{\|y\|'\leq 1} (\langle y,a\rangle - \sup\{\langle y,x\rangle \mid x\in K\})$$

$$= \max_{\|y\|'\leq 1} (\langle y,a\rangle + \inf\{\langle -y,x\rangle \mid x\in K\})$$

$$= \max_{\|y\|'\leq 1} (\langle -y,a\rangle + \inf\{\langle y,x\rangle \mid x\in K\})$$

$$= \max_{\substack{\|y\|' \leq 1 \\ y \in F}} (\langle -y,a\rangle + \min\{\langle y,x\rangle \mid x \in K\}) \quad \text{by}$$

Lemma 2.3.2,

(2.3.4.1) $$= \langle -\bar{y},a\rangle + \min\{\langle \bar{y},x\rangle \mid x \in K\},$$

for some $\bar{y} \in F$, $\|\bar{y}\|' \leq 1$,

$$\leq \langle -\frac{\bar{y}}{\|\bar{y}\|'}, a\rangle + \min\{\langle \frac{\bar{y}}{\|\bar{y}\|'}, x\rangle \mid x \in K\} \quad \text{because}$$

$a \notin K$ and so the value of (P) is greater than zero,

$$\leq \max_{\substack{\|y\|' = 1 \\ y \in F}} (\langle -y,a\rangle + \min\{\langle y,x\rangle \mid x \in K\})$$

$$\leq \max_{\substack{\|y\|' \leq 1 \\ y \in F}} (\langle -y,a\rangle + \min\{\langle y,x\rangle \mid x \in K\})$$

and now the result clearly follows.

2.3.5. i) The part of Theorem 2.3.4 stating that the value of (P') is not greater than the value of (P) will be referred to as the *weak duality principle.* The proof of this is easy and it does not invoke any duality theorems as the one in Nirenberg [4].

ii) If we omit the condition $a \notin K$ in Theorem 2.3.4, then the theorem is still true if we modify (P') slightly, namely

(P') $\max\{\langle -y,a\rangle + \min\{\langle y,x\rangle \mid x \in \mathbb{R}^n_+,\ Ax = b\} \mid y \in F,\ \|y\|' \leq 1\}$.

This is clear from (2.3.4.1).

2.3.6 THEOREM. Let the norm $\|\cdot\|$ on $\mathbb{R}^n$ be strictly convex and let $a \in \mathbb{R}^n$. Assume that $K := \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ is non-empty and that $a \in \mathbb{R}^n \setminus K$. Then $\bar{x} \in \mathbb{R}^n$ solves (P) if and only if $A\bar{x} = b$, $\bar{x} \geq 0$

and $\exists\ y \in \mathbb{R}^n$, $\|y\|' = 1$ such that the following are true:

(2.3.6.1) $$- \langle y,a \rangle + \min\{\langle y,x \rangle \mid x \in K\} > 0$$

and

(2.3.6.2) $$\bar{x} - a = (-\langle y,a \rangle + \min\{\langle y,x \rangle \mid x \in K\})y'.$$

Proof. "If" part: Since (2.3.6.1) is true, then, by Lemma 2.3.2, $y \in F$. By (2.3.6.2) and (2.3.6.1) we get

$$\|\bar{x} - a\| = - \langle y,a \rangle + \min\{\langle y,x \rangle \mid x \in K\}.$$

Now the weak duality principle implies that $\bar{x}$ solves (P) and y solves (P').

"Only if" part: Let $\bar{x}$ solve (P) and y solve (P'). By Theorem 2.3.4, $\|\bar{x} - a\| = \langle -y,a \rangle + \min\{\langle y,x \rangle \mid x \in K\}$. Note that

$$\begin{aligned} \langle \bar{x} - a, y \rangle &\geq - \langle y,a \rangle + \min\{\langle y,x \rangle \mid x \in K\} \quad \text{because } \bar{x} \in K \\ &= \|\bar{x} - a\| > 0. \end{aligned}$$

So,

$$\langle \frac{\bar{x}-a}{\|\bar{x}-a\|}, y \rangle \geq 1 = \|y\|'.$$

Since the reverse inequality is also satisfied and the norm $\|\cdot\|$ is strictly convex, we get

$$\frac{\bar{x}-a}{\|\bar{x}-a\|} = y' \quad \text{and then (2.3.6.2) clearly follows.}$$

This completes the proof.

2.3.7. In proving Theorem 2.3.8 we will use (1.3.8.1), which may be rewritten as:

If K is a non-empty closed convex subset of $\mathbb{R}^n$ and $a \in \mathbb{R}^n$, then $\bar{x} \in \mathbb{R}^n$ is the unique point in K nearest to a for the Euclidean norm if and only if $\bar{x} \in K$ and

(2.3.7.1) $$\langle \bar{x} - a, \bar{x} \rangle \leq \langle \bar{x} - a, x \rangle, \ \forall x \in K$$

or

(2.3.7.2) $\qquad <\bar{x} - a,\bar{x}> \; = \; \min\{<\bar{x} - a,x> \mid x \epsilon K\}.$

2.3.8 THEOREM. Let $a \epsilon \mathbb{R}^n$. Assume that the set $K: = \{x \epsilon \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ is non-empty. Then $\bar{x} \epsilon \mathbb{R}^n$ solves (P) for the Euclidean norm $\|\cdot\|_2$ if and only if $A\bar{x} = b,\ \bar{x} \geq 0$ and the following relation is true

(2.3.8.1) $\qquad \|\bar{x} - a\|_2^2 = -<a,\bar{x} - a> + \min\{<\bar{x} - a,x> \mid Ax = b,\ x \geq 0\}.$

In the case when this is true, we have $\bar{x} - a \epsilon F$. Note that we can replace (2.3.8.1) by

(2.3.8.2) $\qquad <\bar{x} - a,\bar{x}> \; = \; \min\{<\bar{x} - a,x> \mid Ax = b,\ x \geq 0\}.$

Proof. We can see that (2.3.8.1) and (2.3.8.2) are identical by writing $\|\bar{x} - a\|_2^2 = <\bar{x} - a,\bar{x} - a>$. Now the theorem immediately follows from 2.3.7. Finally, if $\bar{x}$ solves (P) for the Euclidean norm, then $\bar{x} - a \epsilon F$ because of Lemma 2.3.2 and the fact that $\min\{<\bar{x} - a,x> \mid Ax = b,\ x \geq 0\}$ is finite.

2.4 ALGORITHMS AND FEASIBILITY

Let $\|\cdot\|$ be a strictly convex norm on $\mathbb{R}^n$.

Let $b_k \epsilon \{Ax \mid x \epsilon \mathbb{R}^n_+\}$, $\forall k \geq 0$, with $\lim b_k = b$. Then $b \epsilon \{Ax \mid x \epsilon \mathbb{R}^n_+\}$ too, because any finitely generated convex cone is closed (see Theorem 19.1 in [5]). Now consider the following Algorithm, which is an infinite one, since there are no stopping criteria:

2.4.1 ALGORITHM.

Step 0. Let $g_0 \epsilon \mathbb{R}^n$ be such that $g_0 \epsilon F$ and $\|g_0\|' = 1$. Recalling the definition of F, which is given following Lemma 2.3.1, we see that such

g_0 exists since F is a cone not equal to $\{0\}$. Put $k = 0$.

Step 1. Calculate

$$\rho_k := \min\{<g_k,x> \mid Ax = b_k,\ x \geq 0\}.$$

If $\rho_k \geq 0$, put

$$y_k := g_k \text{ and } \beta_k := \rho_k \text{ and GO TO STEP 3.}$$

If $\rho_k < 0$, proceed.

Step 2. Pick any y_k such that

$$y_k \epsilon F,\ \|y_k\|' = 1 \text{ and}$$

$$\beta_k := \min\{<y_k,x> \mid Ax = b_k,\ x \geq 0\} \geq 0.$$

(A way of doing this is the following:

Let z_k be the solution of the problem

$$Ax = b_k,\ x \geq 0,\ \|x\|_2 \text{ (min)}.$$

By Theorem 2.3.8 we have $z_k \epsilon F$ and

$$\|z_k\|_2^2 = \min\{<z_k,x> \mid Ax = b_k,\ x \geq 0\}. \tag{2.4.1.1}$$

Let $y_k := z_k/\|z_k\|'$ and

$$\beta_k := \|z_k\|_2^2/\|z_k\|').$$

Steps 3 through *10* calculate g_{k+1} from y_k. Note that for the already computed y_k and β_k we have

$$\beta_k = \min\{<y_k,x> \mid Ax = b_k,\ x \geq 0\} \geq 0.$$

Step 3. Let $a_k := \beta_k y_k'$

and let x_k be the solution of the problem

$$Ax = b_k,\ x \geq 0,\ \|x - a_k\|_2 \text{ (min)}.$$

Let $u_k := x_k - a_k$.

Step 4. If $u_k = 0$, then put $\gamma_k := 0$, $\alpha_k := 0$, $g_{k+1} := y_k$, increment k by 1 and RETURN TO *Step 1*, *else* proceed.

Step 5. Let $\gamma_k := <u_k,x_k>$. Note that $\gamma_k = \|u_k\|_2^2 + <u_k,a_k>$. Note also that by Theorem 2.3.8 we have $u_k \epsilon F$ and

(2.4.1.2) $\gamma_k = \min\{<u_k,x> | Ax = b_k,\ x \geq 0\}$.

Step 6. If $\gamma_k \leq 0$, then $\bar{\alpha}_k := \beta_k/(\beta_k - \gamma_k)$ and GO TO *Step 9*, *else* proceed.

Step 7. $\bar{\alpha}_k := 1$.

Step 8. If

$$\gamma_k <u_k', y_k> \geq \beta_k \|u_k\|',$$

then $\alpha_k := 1$ and GO TO *Step 10*, *else* proceed.

Step 9. Find α_k in the interval $(0, \bar{\alpha}_k)$ such that

$$(\alpha_k \gamma_k + (1 - \alpha_k)\beta_k) < (\alpha_k u_k + (1 - \alpha_k) y_k)', u_k - y_k>$$
$$= (\gamma_k - \beta_k) \|\alpha_k u_k + (1 - \alpha_k) y_k\|'.$$

It will be shown that such an α_k exists. α_k is unique if the norm $\|\cdot\|$ is also smooth.

Step 10. Let

$$g_{k+1} := (\alpha_k u_k + (1 - \alpha_k) y_k)/\|\alpha_k u_k + (1 - \alpha_k) y_k\|'.$$

Increment k by 1 and RETURN TO *Step 1.*

REMARK. Note that *Step 2* makes sense: suppose that $b_k = 0$. Then

$$\rho_k = \min\{<x,g_k> | Ax = 0,\ x \geq 0\} \leq 0 \text{ by taking } x = 0.$$

If $\exists$ x, $Ax = 0$, $x \geq 0$ such that $<x,g_k> < 0$ then

$\inf\{<x,g_k> | Ax = 0,\ x \geq 0\} = -\infty$, by taking

λx, $\lambda \to +\infty$. But since $g_k \epsilon F$, $\rho_k \geq 0$. So finally $\rho_k = 0$ and after this we see that *Step 2* is skipped and

$x_k = 0$, $g_{k+1} = y_k = g_k$. So whenever *Step 2* is executed, we have $b_k \neq 0$ and thus $z_k \neq 0$.

2.4.2. We consider the following algorithm which is just Algorithm 2.4.1 for the case when $b_k = b$, $\forall k = 0,1,\ldots.$ We are assuming that $b \in \{Ax \mid x \in \mathbb{R}^n, x \geq 0\}$. Also, the norm $\|\cdot\|$ on $\mathbb{R}^n$ is assumed strictly convex. Everything that is valid for Algorithm 2.4.1 is valid for the following algorithm too, but not vice versa.

2.4.3 ALGORITHM.

In Algorithm 2.4.1 replace b_k by b, $\forall k = 0,1,\ldots$ and z_k by z, $\forall k = 0,1,\ldots$ where z is the solution of the problem

$$Ax = b,\ x \geq 0,\ \|x\|_2 \text{ (min)}.$$

Now we deal with the feasibility of Algorithm 2.4.1.

2.4.4 LEMMA. Suppose that Algorithm 2.4.1 has been able to reach its kth iteration cycle, $k \geq 0$, having produced a g_k such that $g_k \in F$ and $\|g_k\|' = 1$. Then all *Steps* from *1* through *10* are executable and g_{k+1} will be calculated such that $g_{k+1} \in F$, $\|g_{k+1}\|' = 1$. If the u_k of *Step 3* is nonzero, then u_k and y_k are linearly independent and we can define the function

$$\varphi_k(\alpha) = (\alpha\gamma_k + (1-\alpha)\beta_k)/\|\alpha u_k + (1-\alpha)y_k\|' \quad \text{for} \quad \alpha \in [0,1],$$

and the α_k which is calculated in *Steps 6* through *9* is a global maximizer of φ_k on $[0,1]$. α_k is unique if the norm $\|\cdot\|$ is also smooth.

Proof. From the Algorithm, we can see that *Steps 1* and *2* are executable—note that ρ_k makes sense because $g_k \in F$ by assumption—and a y_k will be produced such that $\|y_k\|' = 1$. Also, we will have $y_k \in F$ because $g_k \in F$ by assumption, and $z_k \in F$ by *Step 2.* Note that we always have $\beta_k = \min\{<y_k,x> | Ax = b_k,\ x \geq 0\} \geq 0$ and $\beta_k \geq \rho_k$.

Now assume $u_k = \lambda y_k$ for some real number λ. By *Step 5*, we have

(2.4.4.1) $$\|u_k\|_2^2 + <u_k,a_k> = \min\{<u_k,x> \ |Ax = b_k,\ x \geq 0\}.$$

Using *Step 3* and the assumption $u_k = \lambda y_k$, the last relation becomes

$$\lambda^2\|y_k\|_2^2 + \lambda\beta_k = \begin{cases} \lambda\beta_k, & \text{for } \lambda \geq 0 \\ \lambda \max\{<y_k,x> | Ax = b_k,\ x \geq 0\}, & \text{for } \lambda < 0 \end{cases}$$
$$\leq \lambda\ \beta_k.$$

So, $\lambda^2\|y_k\|_2^2 \leq 0$ which implies that $\lambda = 0$ i.e. $u_k = 0$. So, if $u_k \neq 0$ then y_k and u_k are linearly independent and so φ_k can be defined. Note that $\varphi_k(0) = \beta_k \geq 0$ and, by (1.5.3.1),

(2.4.4.2) $$\varphi_k'(0) = \gamma_k - \beta_k <u_k,y_k'> = \gamma_k - <u_k,a_k> = \|u_k\|_2^2 > 0.$$

Now apply Lemma 1.5.4 to see that *Steps 6* through *9* are properly formulated and α_k is a global maximizer of φ_k on $[0,1]$. α_k is unique if the norm $\|\cdot\|$ is also smooth. Note that if $\beta_k = 0$, then $\alpha_k = 0$ and thus $\gamma_k = \|u_k\|_2^2 > 0$ which makes Lemma 1.5.4 applicable.

Now it is clear that the Algorithm will produce a g_{k+1} such that $\|g_{k+1}\|' = 1$, since all Steps are well-formulated. Note also that $g_{k+1} \epsilon F$, since $y_k \epsilon F$ and $u_k \epsilon F$.

Also, we have for $u_k \neq 0$

$$\min\{<g_{k+1},x> | Ax = b_k,\ x \geq 0\} \geq \varphi_k(\alpha_k)$$

(2.4.4.3) $$> \varphi_k(0) = \beta_k \geq 0,$$ since

$\varphi_k'(0) > 0$. Also $\min\{<g_{k+1},x> | Ax = b_k,\ x \geq 0\} \geq \varphi_k(1) = \gamma_k/\|u_k\|'$.

If $u_k = 0$,

(2.4.4.4) $$\min\{<g_{k+1},x> | Ax = b_k,\ x \geq 0\} = \beta_k.$$

So, we always have

$$\text{(2.4.4.5)} \qquad \min\{\langle g_{k+1},x\rangle \mid Ax = b_k,\ x \geq 0\} \geq \beta_k \geq 0$$

and

$$\text{(2.4.4.6)} \qquad \|u_k\|'\min\{\langle g_{k+1},x\rangle \mid Ax = b_k, x \geq 0\} \geq \gamma_k (\text{if } u_k=0, \gamma_k = 0).$$

2.4.5 LEMMA. Algorithm 2.4.1 is feasible and it is generating infinite sequences (g_k), (y_k) such that

$$g_k \epsilon F,\ \|g_k\|' = 1,\ y_k \epsilon F,\ \|y_k\|' = 1,\ \forall k = 0,1,2,\ldots.$$

Proof. Follows by a simple induction argument using Lemma 2.4.4 and *Step 0* of the Algorithm.

2.4.6. The sequence (β_k), $k \geq 0$, generated by Algorithm 2.4.1 need not be increasing.

2.4.7. In practice, in order to initialize, we can replace *Step 0* of Algorithm 2.4.1 with the following *Steps 0*, *01*, *02*:

Step 0. $k = 0$.

Step 01. If $b_k \neq 0$, let $g_0 := \dfrac{x_0}{\|x_0\|'}$, where x_0 is the solution of the problem

$$Ax = b_k,\ x \geq 0,\ \|x\|_2 \ (\min),$$

and GO TO *Step 1*, *else* proceed.

Step 02. Let $x_k := 0$, increment k by 1 and RETURN TO *Step 01.*

2.4.8. From *Steps 1 and 2* of Algorithm 2.4.1, we see that if at some stage $\rho_k < 0$, then g_k is thrown away and the Algorithm restarts at this point, in the sense that no previous information is used except for the current b_k. However, the previous information is used in the following variation of Algorithm 2.4.1.

2.4.9 ALGORITHM.

All the hypotheses of Algorithm 2.4.1 are assumed here also.

Steps 1 through *1e* calculate y_k from g_k and b_k.

Step 0. as in 2.4.1.

Step 1. as in 2.4.1.

Step 1a. Let z_k be the solution of

$$Ax = b_k,\ x \geq 0,\ \|x\|_2 \ (\min).$$

Step 1b. If $\rho_k \leq \|z_k\|_2^2 < z_k', g_k>/\|z_k\|'$,

then put

$$\hat{\alpha}_k := 0,\ y_k := z_k/\|z_k\|',\ \beta_k := \|z_k\|_2^2/\|z_k\|'$$

and GO TO *Step 3*, *else* proceed.

Step 1c. $\tilde{\alpha}_k := \|z_k\|_2^2/(\|z_k\|_2^2 - \|z_k\|'\rho_k)$.

Step 1d. Find $\hat{\alpha}_k$ in $(0,\tilde{\alpha}_k)$ such that

$$(\hat{\alpha}_k\rho_k + (1 - \hat{\alpha}_k)\frac{\|z_k\|_2^2}{\|z_k\|'}) < (\hat{\alpha}_k g_k + (1-\hat{\alpha}_k)\frac{z_k}{\|z_k\|'})',g_k - \frac{z_k}{\|z_k\|'} >$$

$$= (\rho_k - \frac{\|z_k\|_2^2}{\|z_k\|'})\ \|\hat{\alpha}_k g_k + (1 - \hat{\alpha}_k)\frac{z_k}{\|z_k\|'}\|'.$$

It will be shown that such an $\hat{\alpha}_k$ exists and $\hat{\alpha}_k$ is unique, if the norm $\|\cdot\|$ is also smooth.

Step 1e. Let

$$y_k := ((1 - \hat{\alpha}_k) \frac{z_k}{\|z_k\|'} + \hat{\alpha}_k g_k)/\|(1 - \hat{\alpha}_k) \frac{z_k}{\|z_k\|'} + \hat{\alpha}_k g_k\|'.$$

Step 2. Calculate

$$\beta_k := \min\{<y_k,x> | Ax = b_k, x \geq 0\}.$$

Steps 3 through *10* are the same as in 2.4.1.

Now we deal with the feasibility of Algorithm 2.4.9.

2.4.10 LEMMA. Suppose that Algorithm 2.4.9 has been able to reach its kth iteration cycle, $k \geq 0$, and that g_k has thus been computed so that $g_k \epsilon F$. Suppose $\rho_k < 0$; if z_k and g_k are linearly dependent , then the criterion of *Step 1b* is answered affirmatively.

Proof. First note that ρ_k makes sense, since $g_k \epsilon F$ by assumption. Suppose $g_k = \lambda z_k$ for some real number λ. Using *Step 1a* and relation (2.4.1.1) which is still valid, we see that the criterion of *Step 1b* checks if

$$\min\{\lambda < z_k,x> | Ax = b_k, x \geq 0\} \leq \lambda \|z_k\|_2^2$$
$$= \lambda \min \{<z_k,x> | Ax = b_k, x \geq 0\}.$$

This is clearly true for $\lambda \geq 0$ as equality. If $\lambda < 0$, the left hand side becomes

$\lambda \max \{<z_k,x> | Ax = b_k, x \geq 0\}$ and the above relation is again found to be true.

2.4.11 LEMMA. Suppose that Algorithm 2.4.9 has been able to reach its kth iteration cycle, $k \geq 0$, and that g_k has thus been computed so that $g_k \epsilon F$ and $\|g_k\|' = 1$. Also assume that $\rho_k < 0$ and g_k, z_k are linearly independent. Then, defining

$$\hat{\varphi}_k(\alpha) = ((1 - \alpha) \frac{\|z_k\|_2^2}{\|z_k\|'} + \alpha \rho_k)/\|(1 - \alpha) \frac{z_k}{\|z_k\|'} + \alpha g_k\|'$$

for $\alpha\epsilon[0,1]$, we have that

Steps 1b through *1d* of Algorithm 2.4.9 are properly formulated and the resulting $\hat{\alpha}_k$ is a global maximizer of $\hat{\varphi}_k$ on $[0,1]$. $\hat{\alpha}_k$ is unique if the norm $\|\cdot\|$ is also smooth.

Proof. $\hat{\varphi}_k$ is clearly defined because the denominator never vanishes as a result of the linear independence assumption, and also because ρ_k makes sense as a result of the $g_k \epsilon F$ assumption. Note that $\hat{\varphi}_k(0) > 0$. We can easily check that the following is true by referring to relation (1.5.3.1): The Criterion of *Step 1b* is satisfied if and only if $(\hat{\varphi}_k)'(0) \leq 0$. If this criterion is satisfied, then by (1.5.3.3) we get that $\hat{\alpha}_k = 0$ is a global maximizer of $\hat{\varphi}_k$ on $[0,1]$. $\hat{\alpha}_k$ is unique if the norm $\|\cdot\|$ is also smooth by (1.5.3.4). When the criterion of *Step 1b* is true, we put $y_k = z_k/\|z_k\|'$ which is the same as *Step 1e* for $\hat{\alpha}_k = 0$.

If the Criterion of *Step 1b* is not satisfied, then $(\hat{\varphi}_k)'(0) > 0$ and it is clear that *Steps 1c* and *1d* are properly formulated and they produce an $\hat{\alpha}_k$ with the properties stated, by referring to Lemma 1.5.4 and noting that $\rho_k < 0$.

From the above, it is clear that a y_k is produced such that $y_k \epsilon F$—because z_k and $g_k \epsilon F$—and $\|y_k\|' = 1$. Thus the β_k of *Step 2* makes sense, since $y_k \epsilon F$. From *Step 1e*, which is valid even if $\hat{\alpha}_k = 0$ and *Step 2* we get

$$\|(1 - \hat{\alpha}_k) \frac{z_k}{\|z_k\|'} + \hat{\alpha}_k g_k\|' \beta_k \geq$$

$$\geq \frac{(1-\hat{\alpha}_k)}{\|z_k\|'} \min\{<z_k,x> | Ax = b_k, x \geq 0\} +$$

$$+ \hat{\alpha}_k \min \{<g_k,x> | Ax = b_k, x \geq 0\}.$$

But this becomes

$$\beta_k \geq \hat{\varphi}_k(\hat{\alpha}_k)$$

(2.4.11.1) $$\geq \hat{\varphi}_k(0) = \frac{\|z_k\|_2^2}{\|z_k\|'} > 0.$$

Also, $$\beta_k \geq \hat{\varphi}_k(\hat{\alpha}_k)$$

(2.4.11.2) $$\geq \hat{\varphi}_k\,(1) = \rho_k.$$

2.4.12. At the kth iteration cycle of Algorithm 2.4.9, assuming $\rho_k < 0$, we have the relations

$$\beta_k \geq \|z_k\|_2^2 / \|z_k\|' > 0 \quad \text{and} \quad \beta_k \geq \rho_k.$$

If z_k, g_k are linearly independent, these are just (2.4.11.1), (2.4.11.2). If they are linearly dependent, then

$$\beta_k = \|z_k\|_2^2 / \|z_k\|' \quad \text{and by } \textit{Step 1b},$$

$$\rho_k \leq \beta_k < z_k', g_k> \leq \beta_k \|z_k'\| \; \|g_k\|' = \beta_k.$$

Note that the relations $\beta_k \geq \rho_k$ and $\beta_k \geq 0$ are true even if $\rho_k \geq 0$.

2.4.13. Lemmas 2.4.10, 2.4.11 and 2.4.4 show that Algorithm 2.4.9 is feasible.

2.4.14. Now that we know that the Algorithms are feasible, we see that the assumption of Lemmas 2.4.4, 2.4.10 and 2.4.11 that "the Algorithm has reached its kth iteration cycle with a g_k computed such that $\|g_k\|' = 1$, $g_k \epsilon F$ ", is always true.

2.5 CONVERGENCE OF ALGORITHM 2.4.3

In this section we are still assuming that $\|\cdot\|$ is a strictly convex norm on $\mathbb{R}^n$ and that $b \in \{Ax \mid x \in \mathbb{R}^n, x \geq 0\}$. We prove that Algorithm 2.4.3 converges to the solution $\bar{x}$ of the problem

(P_0) $\quad Ax = b,\ x \geq 0,\ \|x\|$ (min).

Note that Algorithm 2.4.3 is infinite because Algorithm 2.4.1 is. The assumptions regarding Algorithm 2.4.1 are the same as those in Section 2.4.

2.5.1 LEMMA. Consider the functions

$$f_1(c,d) := \min\{\langle c,x\rangle \mid Ax = d,\ x \geq 0\}$$

over $\quad F \times \{Ax \mid x \in \mathbb{R}^n,\ x \geq 0\}$ and

$$f_2(a,d) := \min\{\|x - a\|_2 \mid Ax = d,\ x \geq 0\}$$

over $\quad \mathbb{R}^n \times \{Ax \mid x \in \mathbb{R}^n,\ x \geq 0\}$.

Then f_1, f_2 are continuous over their respective domains.

Proof. For the continuity of f_1 we refer to [1], [9]. Now we prove that f_2 is continuous. Let the sequences (a_k) and (d_k) be such that $a_k \to a$ and $d_k \to d$, where $a_k, a \in \mathbb{R}^n$ and $d_k, d \in \{Ax \mid x \in \mathbb{R}^n,\ x \geq 0\}$. Let x_k be the solution of the problem

$$Ax = d_k,\ x \geq 0,\ \|x - a_k\|_2 \ \text{(min)}$$

and thus $\|x_k - a_k\|_2 = f_2(a_k,d_k)$.

From 2.3.5 ii) we have that $\exists\, y_k \in F$, $\|y_k\|_2 \leq 1$ such that

(2.5.1.1) $\qquad \|x_k - a_k\|_2 = \langle -y_k,a_k\rangle + \min\{\langle y_k,x\rangle \mid Ax = d_k,\ x \geq 0\}.$

Since (a_k), (y_k), (d_k) are bounded and the functions f_1 and $\langle\cdot,\cdot\rangle$ are continuous, we get that the right hand side of (2.5.1.1) is bounded. Since $\|x_k\|_2 \leq \|x_k - a_k\|_2 + \|a_k\|_2$, the sequence (x_k) is bounded.

Now let $x_{k_j} \to \bar{x}$. By taking further subsequences, we can assume that $y_{k_j} \to \bar{y}$.

Taking limits in (2.5.1.1) we get

$$\|\bar{x} - a\|_2 = \langle -\bar{y},a\rangle + \min\{\langle \bar{y},x\rangle \mid Ax = d,\ x \geq 0\}$$

with $\bar{y} \in F$, $\|\bar{y}\|_2 \leq 1$ and also $A\bar{x} = d$, $\bar{x} \geq 0$. In view of 2.3.5 ii), these relations imply that $\bar{x}$ solves the problem

$$Ax = d,\ x \geq 0,\ \|x - a\|_2 \ (\min)$$

and thus $\|\bar{x} - a\|_2 = f_2(a,d)$. Since such a solution is unique, we have

(2.5.1.2) $\qquad \lim x_k = \bar{x}$

and thus for $k \to \infty$,

$$\lim f_2(a_k,d_k) = \lim \|x_k - a_k\|_2 = \|\bar{x} - a\|_2 = f_2(a,d)$$

which completes the proof that the function f_2 is continuous.

2.5.2. Let $f_3(a,d)$ be the minimizer of the problem

$$Ax = d,\ x \geq 0,\ \|x - a\|_2 \ (\min)$$

where $a \in \mathbb{R}^n$ and $d \in \{Ax \mid x \geq 0\}$. $f_3(a,d)$ is, of course, a vector in $\mathbb{R}^n$. Then, due to (2.5.1.2), f_3 is a continuous function of $\mathbb{R}^n \times \{Ax \mid x \geq 0\}$ into $\mathbb{R}^n$.

2.5.3 REMARK. Referring to Algorithm 2.4.1, the following are true because of Lemma 2.5.1:

(β_k) is bounded, since (y_k) and (b_k) are.

(a_k) is bounded, since (β_k) is.

(u_k) is bounded, since (b_k) and (a_k) are.

(x_k) is bounded, since (u_k) and (a_k) are.

(γ_k) is bounded, since (u_k) and (x_k) are.

(ρ_k) is bounded, since (b_k) and (g_k) are.

In the same fashion, all sequences appearing in Algorithm 2.4.9 are bounded, too.

2.5.4 LEMMA. Consider Algorithm 2.4.1 for a strictly convex norm $\|\cdot\|$ on $\mathbb{R}^n$ and

$b_k \epsilon \{Ax \mid x \epsilon \mathbb{R}^n_+\}$, $k \geq 0$, with $\lim b_k = b$, $b \neq 0$. Assume that the algorithm is initiated with a given $g_0 \epsilon F$ with $\|g_0\|' = 1$. Then,

(i) Every cluster point of (ρ_k), $k \geq 0$, is greater than zero.

(ii) $\exists\, k_0$ such that

$$\rho_k > 0, \quad \forall\, k \geq k_0.$$

k_0 depends on (b_k), $k \geq 0$, and g_0.

Proof. Suppose (i) is not true. Then there is a subsequence such that

(2.5.4.1) $\lim \rho_{k_j} = \lim \min\{<g_{k_j},x> \mid Ax = b_{k_j},\ x \geq 0\} \leq 0.$

From (2.4.4.5) and (2.4.4.6) we get

(2.5.4.2) $$\min\{<g_{k_j},x> \mid Ax = b_{(k_j)-1},\ x \geq 0\} \geq \beta_{(k_j)-1} \geq 0$$

for all $k_j \geq 1$, and

(2.5.4.3) $$\|u_{(k_j)-1}\|' \min\{<g_{k_j},x> \mid Ax = b_{(k_j)-1},\ x \geq 0\} \geq \gamma_{(k_j)-1} = \\ = \|u_{(k_j)-1}\|_2^2 + <u_{(k_j)-1}, a_{(k_j)-1}> \text{ for all } k_j \geq 1.$$

By taking further subsequences, we can assume that

$g_{k_j} \to g$. Note that $g \epsilon F$, since F is a closed set.

By allowing $j \to \infty$ in (2.5.4.2), we get that

$$\lim \min\{<g_{k_j},x> \mid Ax = b_{(k_j)-1},\ x \geq 0\} \geq 0$$

$$= \min\{<g,x> \mid Ax = b,\ x \geq 0\}$$

$$= \lim \rho_{k_j} \text{ by (2.5.4.1)} \leq 0$$

and so

(2.5.4.4) $\lim \rho_{k_j} = 0.$

By (2.5.4.4) and (2.5.4.2),

$$\lim \beta_{(k_j)-1} = 0, \text{ as } j \to \infty.$$

But then

(2.5.4.5) $a_{(k_j)-1} = \beta_{(k_j)-1}\, y'_{(k_j)-1} \to 0, \text{ as } j \to \infty.$

Moving $\langle u_{(k_j)-1}, a_{(k_j)-1}\rangle$ to the left hand side of (2.5.4.3) and then taking the lim as $j \to \infty$, we get

$$\lim u_{(k_j)-1} = 0, \text{ as } j \to \infty.$$

For this conclusion we have used (2.5.4.4), (2.5.4.5) and the fact that $(u_{(k_j)-1})$ is a bounded sequence.

But

(2.5.4.6) $\|u_{(k_j)-1}\|_2 = \min\{\|x - a_{(k_j)-1}\|_2 \mid Ax = b_{(k_j)-1},\ x \geq 0\}.$

Passing to the limit in (2.5.4.6) we get

$$0 = \min\{\|x - 0\|_2 \mid Ax = b,\ x \geq 0\}$$

which implies that $b = 0$. But one of the hypotheses was that b is nonzero. So (2.5.4.1) cannot be valid and, thus, (i) follows. Now suppose (ii) is not true. Then there is a subsequence such that

(2.5.4.7) $\rho_{k_j} \leq 0, \forall\ j.$

Since (ρ_k) is a bounded sequence, (2.5.4.7) contradicts (i).

2.5.5 REMARK. Lemma 2.5.4 implies that, in Algorithm 2.4.1 with $b \neq 0$, there exists k_0 such that

(2.5.5.1) $\beta_k = \rho_k > 0, \forall\ k \geq k_0,$

(2.5.5.2) $y_k = g_k, \quad \forall\ k \geq k_0$ and

(2.5.5.3) *Step 2* is not executed $\forall\ k \geq k_0$, where k_0 is the same as the one of Lemma 2.5.4 (ii).

It is easy to see that the proof of Lemma 2.5.4 applies to Algorithm 2.4.9 as well. So, there is a k_0' such that *Steps 1a* through (and including) *Step 2* of Algorithm 2.4.9 are not executed for all $k \geq k_0'$ because $\rho_k > 0$, $\forall\ k \geq k_0'$.

2.5.6 THEOREM. Assume that the norm $\|\cdot\|$ on $\mathbb{R}^n$ is strictly convex and that $b \in \{Ax \mid x \in \mathbb{R}^n_+\}$. Then the sequence (x_k) generated by Algorithm 2.4.3 converges to the solution $\bar{x}$ of problem (P_0):

(P_0) $\qquad Ax = b,\ x \geq 0,\ \|x\|$ (min).

Proof. Assume $b \neq 0$. Note that for all $k \geq 0$

(2.5.6.1) $$\rho_{k+1} = \min\{\langle g_{k+1},x\rangle \mid Ax = b,\ x \geq 0\} \geq \beta_k \geq 0$$

by (2.4.4.5) and the fact that $b_{k+1} = b_k = b$. Now (2.5.6.1) implies that for all $k \geq 1$ we have:

(2.5.6.2) $\beta_k = \rho_k$ and $y_k = g_k$ and *Step 2* is not executed.

For $k = 0$, we have $\beta_0 \geq 0$ and $-\infty < \rho_0 < +\infty$, as in Algorithm 2.4.1. However, if $\rho_0 \geq 0$, then *Step 2* is never executed.

(2.5.6.1) also implies that (β_k), $k \geq 0$, is an increasing sequence of non–negative numbers:

$$\beta_{k+1} = \rho_{k+1} \geq \beta_k \geq 0, \quad \forall\ k \geq 0.$$

From (2.4.4.3) and (2.4.4.4) it is clear that for $k \geq 0$,

(2.5.6.3) $\beta_{k+1} = \beta_k$ if and only if $u_k = 0$.

Now suppose $\exists\ \bar{k}$ such that $u_{\bar{k}} = 0$. Then

$$x_{\bar{k}} = \beta_{\bar{k}}\, y_{\bar{k}}' \text{ and thus}$$

$$\|x_{\bar{k}}\| = \beta_{\bar{k}}, \text{ since } \beta_{\bar{k}} \geq 0 \text{ (in fact } \beta_{\bar{k}} > 0 \text{ since } b_{\bar{k}} \neq 0)$$

(2.5.6.4) $$= \min\{\langle y_{\bar{k}},x\rangle \mid Ax = b,\ x \geq 0\}.$$

Since $y_{\bar{k}} \in F$, $\|y_{\bar{k}}\|' = 1$ and $Ax_{\bar{k}} = b$, $x_{\bar{k}} \geq 0$, then (2.5.6.4) together with 2.3.5(i), i.e. the weak duality principle, imply that $x_{\bar{k}}$ solves (P_0) and $y_{\bar{k}}$ solves the dual problem of (P_0) defined in 2.3.3 and $\beta_{\bar{k}}$ is the value of the two problems. So, $\beta_{\bar{k}}$ cannot be increased further and we must have

(2.5.6.5) $$\beta_k = \beta_{\bar{k}}, \quad \forall\ k \geq \bar{k}$$

because (β_k), $k \geq 0$, is an increasing sequence. (2.5.6.5) and (2.5.6.3) imply that

$$u_k = 0, \quad \forall\ k \geq \bar{k}.$$

Let $k \geq \bar{k}$. Since $u_k = 0$, the above argument, for k instead of $\bar{k}$, shows that x_k solves (P_0) and y_k solves its dual. The solution to (P_0) is unique, so $x_k = x_{\bar{k}}$, $\forall\ k \geq \bar{k}$ and thus the sequence (x_k) converges to the solution of (P_0) as an eventually constant sequence. Also, y_k solves the dual of (P_0), $\forall\ k \geq \bar{k}$. So any cluster point of (y_k) also solves (P_0), because of the continuity of the map $\min\{<\cdot,x> | Ax = b,\ x \geq 0\}$ over F.

Now we can assume that

$$u_k \neq 0, \quad \forall\ k \geq 0.$$

This implies that (β_k), $k \geq 1$, is a strictly increasing positive sequence. Let $\lim \beta_k = \beta$. Clearly $\beta > 0$. Also $\beta < +\infty$ because (β_k) is a bounded sequence by Remark 2.5.3.

We will prove that $\lim u_k = 0$ as $k \to \infty$.

Suppose this is not true. Then, there is a subsequence of (u_k), $k \geq 0$, denoted by $(u_{k'})$ such that $u_{k'} \to u \neq 0$.

By considering further subsequences, and by Remark 2.5.3 we can assume that $y_{k'} \to y$, $\gamma_{k'} \to \gamma$ and $\alpha_{k'} \to \alpha \in [0,1]$.

From *Step 5* of the Algorithm we have

(2.5.6.6) $\qquad \gamma = \|u\|_2^2 + \beta < u,y'>.$

Note that the map $x \mapsto x'$ is continuous on $\mathbb{R}^n \setminus \{0\}$ due to the strict convexity of the norm $\|\cdot\|$.

In Lemma 2.4.4 we showed that if $u_k \neq 0$, then u_k and y_k are linearly independent. By repeating that argument we can show that u and y are linearly independent because $u \neq 0$. Of course we have to use that

$$\beta = \min\{<y,x> \mid Ax = b,\ x \geq 0\}$$

and

$$\gamma = \min\{<u,x> \mid Ax = b,\ x \geq 0\},$$ which result from the corresponding relations for the indices k'.

Case 1. Assume that

$$\alpha_{k'} \neq 1, \text{ for all } k'.$$

Then, by the equation for α_k of *Step 9* of the Algorithm we get

$$(\alpha\gamma + (1 - \alpha)\beta) <(\alpha u + (1 - \alpha)y)', u - y> =$$

(2.5.6.7) $$= (\gamma - \beta)\ \|\alpha u + (1 - \alpha)y\|'.$$

From the relations

$$\begin{aligned} \beta_{k+1} &= \min\{<g_{k+1},x> \mid Ax = b,\ x \geq 0\} \text{ since } b_{k+1} = b_k = b, \\ &\geq \varphi_k(\alpha_k) = (\alpha_k\gamma_k + (1 - \alpha_k)\beta_k)/\|\alpha_k u_k + (1 - \alpha_k)y_k\|' \quad \text{by } 2.4.4.3, \\ &> \beta_k \geq 0 \quad \text{by } 2.4.4.3, \end{aligned}$$

we get

$$\beta \geq (\alpha\gamma + (1 - \alpha)\beta)/\|\alpha u + (1 - \alpha)y\|' \geq \beta$$

and thus (2.5.6.7) becomes

(2.5.6.8) $\qquad \beta < (\alpha u + (1 - \alpha)y)', u - y> = \gamma - \beta.$

Since

$$\alpha < (\alpha u + (1 - \alpha)y)', u - y> =$$
$$= \|\alpha u + (1 - \alpha)y\|' - < (\alpha u + (1-\alpha)y)', y>,$$

(2.5.6.8) becomes

$$\alpha\gamma - \alpha\beta = \beta\ \|\alpha u + (1-\alpha)y\|' - \beta < (\alpha u + (1-\alpha)y)', y>$$

(2.5.6.9) $$= \alpha\gamma + (1-\alpha)\beta \qquad - \beta < (\alpha u + (1-\alpha)y)', y>.$$

Since $\beta \neq 0$, we get from (2.5.6.9)

(2.5.6.10) $\quad 1 = < (\alpha u + (1-\alpha)y)', y>$. This implies

(2.5.6.11) $\quad (\alpha u + (1-\alpha)y)' = y'$, because the norm $\|\cdot\|$ is strictly convex and $\|y\|' = 1$.

Using (2.5.6.11) in (2.5.6.8), we get

(2.5.6.12) $\quad \beta <y',u> = \gamma$ which, together with (2.5.6.6), implies that $u = 0$ which is a contradiction.

Case 2. Assume that

$$\alpha_{k'} = 1, \text{ for all } k'.$$

From *Step 8* of the algorithm we get

(2.5.6.13) $\quad \gamma <u',y> \geq \beta\ \|u\|'$.

As in *Case 1*, we still have

$$\beta\ \|\alpha u + (1-\alpha)y\|' = \alpha\gamma + (1-\alpha)\beta,$$

which for our case becomes

(2.5.6.14) $\quad \beta\ \|u\|' = \gamma$.

In particular, this implies that $\gamma > 0$. (2.5.6.13) and (2.5.6.14) imply that

$$<u',y> \geq 1, \text{ since } \gamma > 0.$$

But

$$<u',y> \leq \|u'\|\ \|y\|' = 1.$$

So,

$<u',y> = 1$ which implies that $u' = y'$ because $\|y\|' = 1$ and the norm $\|\cdot\|$ is strictly convex. Now (2.5.6.6) becomes

$$\gamma = \|u\|_2^2 + \beta <u,u'>$$

(2.5.6.15) $$= \|u\|_2^2 + \beta\ \|u\|'.$$

(2.5.6.15) and (2.5.6.14) imply that $u = 0$ which is a contradiction.

It is clear that by taking further subsequences we can assume that either *Case 1* or *Case 2* is true for $(\alpha_{k'})$ and thus we always have a contradiction. So, u has to be zero and thus $\lim u_k = 0$.

Now let $x_{k'} \to \tilde{x}$. By considering further subsequences we can assume that $y_{k'} \to \tilde{y}$.

Taking limits in the relation

$$u_{k'} = x_{k'} - a_{k'}$$
$$= x_{k'} - \beta_{k'} y'_{k'}$$

we get

$$\tilde{x} = (\min\{<\tilde{y},x> | Ax = b, x \geq 0\})\ \tilde{y}'.$$

Then $\|\tilde{x}\| = \beta = \min\{<\tilde{y},x> | Ax = b, x \geq 0\}$. We also have $\tilde{y} \epsilon F$, $\|\tilde{y}\|' = 1$ and $A\tilde{x} = b$, $\tilde{x} \geq 0$, since the corresponding relations for the indices k' hold. Now the weak duality principle 2.3.5(i) implies that $\tilde{x}$ solves (P_0), $\tilde{y}$ solves its dual and that β is the common value of the two problems. Since the solution of (P_0) is unique, the sequence (x_k) converges to the solution of (P_0). Also β, the limit of the sequence (β_k), is the value of (P_0). Now, since
$\lim \beta_k = \lim (\min\{<y_k,x> | Ax = b, x \geq 0\}) =$ value of the dual of (P_0), we have that any cluster point of (y_k) is a solution of the dual of (P_0), by continuity of the function $\min\{<\cdot,x> | Ax = b, x \geq 0\}$ over F.

Now assume $b = 0$; then $\bar{x} = 0$ and

$$\rho_0 = \min\ \{<g_0,x> | Ax = 0, x \geq 0\} = 0 \quad \text{since} \quad g_0 \epsilon F.$$

Thus $x_0 = 0$, $g_1 = g_0$.

Repeating, we get that

$$g_k = g_0, \forall\ k \quad \text{and} \quad x_k = 0, \forall\ k.$$

So, indeed

$$x_k \to \bar{x} = 0$$

and $\quad g_k \to g_0$ which solves problem

(P_0') $\quad \max\{\min\{\langle y,x\rangle \mid Ax = 0,\ x \geq 0\} \mid \|y\|' = 1\}$

because

$$\min\{\langle y,x\rangle \mid Ax = 0,\ x \geq 0\} = -\infty, \quad \forall\, y \notin F$$

$$\min\{\langle y,x\rangle \mid Ax = 0,\ x \geq 0\} = 0, \quad \forall\, y \in F.$$

2.5.7 DEFINITION. Consider Algorithm 2.4.3 with the norm $\|\cdot\|$ on $\mathbb{R}^n$ strictly convex and $b \in \{Ax \mid x \in \mathbb{R}^n_+\}$. Let $g \in F$ such that $\|g\|' = 1$. Denote the sequence (β_k), $k \geq 0$, generated by Algorithm 2.4.3 initiating from $g_0 := g$, by (β_k^g), $k \geq 0$. By Theorem 2.5.6, we see that (β_k^g), $k \geq 1$, is an increasing positive sequence converging to $\|\bar{x}\|$ for any $g \in F$, $\|g\|' = 1$, where $\bar{x}$ is the solution of the problem

(P_0) $\quad Ax = b,\ x \geq 0,\ \|x\|$ (min).

Now consider the following algorithm which is an infinite one.

2.5.8 ALGORITHM.

Assume that the norm $\|\cdot\|$ on $\mathbb{R}^n$ is strictly convex. Let $b_k \to b$ such that $b_k \in \{Ax \mid x \in \mathbb{R}^n_+\}$, $\forall\, k \geq 0$.

Step 0. Same as in Algorithm 2.4.1.

Step 1. Calculate

$$\rho_k := \min\{\langle g_k,x\rangle \mid Ax = b_k,\ x \geq 0\}.$$

Step 2. Put

$$y_k := g_k \quad \text{and} \quad \beta_k := \rho_k.$$

Step 3 – 10. Same as in Algorithm 2.4.1.

We will not worry about the existence of α_k or about the feasibility of this algorithm because all we need is the following remark.

2.5.9 REMARK. Assume $b \neq 0$. Apply Algorithm 2.4.1 to (b_k), $k \geq 0$, initiating from a $g_0 \epsilon F$, $\|g_0\|' = 1$. Let k_0 be as in Remark 2.5.5. Then it is clear that from the k_0–th iteration onwards, Algorithm 2.4.1 is identical to Algorithm 2.5.8 applied to (b_k), $k \geq k_0$, and initiated from g_{k_0}. In particular this is true for Algorithm 2.4.3 which is identical to the corresponding Algorithm 2.5.8 after the first iteration.

The same is true for Algorithm 2.4.9 from the k_0'–th iteration onwards. So, Algorithms 2.4.1 and 2.4.9 can just be viewed as Algorithm 2.5.8 applied to a sequence (b_k), $k \geq 0$, and initiating from a g_0 such that $\rho_k \geq 0$, $\forall$ $k \geq 0$. It is clear that if this happens, i.e. if $\rho_k \geq 0$, $\forall$ $k \geq 0$, then Algorithm 2.5.8 is feasible because it coincides with 2.4.1 and 2.4.9 which are feasible.

Note that Algorithms 2.4.1 and 2.4.9 do not necessarily generate sequences that eventually coincide.

2.5.10 LEMMA. For Algorithm 2.4.1 under the same set of hypotheses as in Section 2.4, with $b \neq 0$ we have that there is an $i_0 \geq k_0$, k_0 as in Lemma 2.5.4(ii), such that, $\forall$ $k \geq i_0$,

(2.5.10.1) $$\min\{\langle y_k,x\rangle \mid Ax=b, x \geq 0\}=\min\{\langle g_k,x\rangle \mid Ax = b, x \geq 0\} > 0.$$

i_0 depends on (b_k), $k \geq 0$, and g_0.

Moreover,

(2.5.10.2) All cluster points of $(\min\{\langle y_k,x\rangle \mid Ax = b, x \geq 0\})$, $k \geq 0$, $(\min\{\langle g_k,x\rangle \mid Ax = b, x \geq 0\})$, $k \geq 0$, are greater than zero. The same Lemma is true for Algorithm 2.4.9 for an i_0'.

Proof. Via (2.5.5.2), it is enough to consider only g_k 's.

Suppose there is a subsequence such that $\lim (\min\{\langle g_{k_j},x\rangle \mid Ax = b,\ x \geq 0\}) \leq 0$. By taking a further subsequence, we can assume $\lim g_{k_j} = \bar{g}$. Then,

$0 \geq \min\{\langle \bar{g},x\rangle \mid Ax = b,\ x \geq 0\} = \lim \rho_{k_j}$ contrary to Lemma (2.5.4)(i). The rest follows easily.

CHAPTER 3

3.1 INTRODUCTION

We keep the notation introduced in Chapters 1 and 2.

In this chapter we consider Algorithm 2.4.1 with $b_k \to b$, $g_0 \epsilon F$, $\|g_0\|' = 1$, and we prove that it converges to the solution $\bar{x}$ of the problem

(P_0) $\qquad Ax = b,\ x \geq 0,\ \|x\|\ (\min).$

For this we assume smoothness of the norm $\|\cdot\|$ on $\mathbb{R}^n$ in addition to the already assumed strict convexity. As an immediate application of this, an algorithm for nonnegative least error minimal norm solutions is given. Some numerical results follow.

3.2 CONTINUITY OF THE β_ℓ FUNCTIONS

In this section certain functions are defined and then their continuity proved.

3.2.1 DEFINITIONS. Let $\overline{B}$ be a closed ball in $\mathbb{R}^m$ not containing 0 such that it contains at least one point of $\{Ax \mid x \epsilon \mathbb{R}^n_+\}$. $\overline{B}$ will be considered fixed. Let $\|\cdot\|$ be a strictly convex and smooth norm on $\mathbb{R}^n$ and let $T_0 := \{(g,b) \epsilon (F \cap \{z \epsilon \mathbb{R}^n \mid \|z\|' = 1\}) \times (\{Ax \mid x \epsilon \mathbb{R}^n_+\} \cap \overline{B})$ such that $f_1(g,b) \geq 0\}$, where $f_1(g,b) := \min\{<g,x> \mid Ax = b,\ x \geq 0\}$. In Section 2.5.1 we saw that f_1 is continuous on the set where it is finite. It is easily seen that T_0 is a compact subset of $\mathbb{R}^n \times \mathbb{R}^m$. The set $\{Ax \mid x \epsilon \mathbb{R}^n,\ x \geq 0\}$ is closed because it is a finitely generated convex cone.

Given $(g,b) \epsilon T_0$, let us perform one iteration of Algorithm 2.4.1 for $k = 0$, replacing g_0 by g and b_0 by b. This is the same as doing Algorithm

2.5.8 because $f_1(g,b) \geq 0$. Then the algorithm will evaluate the following: ρ_0, y_0, β_0, a_0, x_0, u_0, γ_0, α_0, g_1. These can be viewed as the values of corresponding functions at (g,b). Let us denote these functions by

(3.2.1.1) $$\rho_0(\cdot,\cdot),\ y_0(\cdot,\cdot),\ \beta_0(\cdot,\cdot),\ a_0(\cdot,\cdot),\ x_0(\cdot,\cdot),\ u_0(\cdot,\cdot),\ \gamma_0(\cdot,\cdot),\ \alpha_0(\cdot,\cdot),\ g_1(\cdot,\cdot) \text{ respectively.}$$

Their domain is T_0. Note that $\alpha_0(\cdot,\cdot)$ is a function because the α_0 of *Step 9* is unique, by the smoothness and strict convexity of the norm $\|\cdot\|$. So $g_1(\cdot,\cdot)$ is a function, too. Also for $(g,b)\epsilon T_0$ and $u_0(g,b) \neq 0$ define $\varphi_0(g,b):[0,1] \to \mathbb{R}$ such that for $\alpha\epsilon[0,1]$, $\varphi_0(g,b)(\alpha) = \dfrac{\alpha\gamma_0(g,b)+(1-\alpha)\beta_0(g,b)}{\|\alpha u_0(g,b)+(1-\alpha)y_0(g,b)\|'}$ which is defined because, as we know, $u_0(g,b)$ and $y_0(g,b)$ are linearly independent.

3.2.2 THEOREM. Assume that $\|\cdot\|$ is a strictly convex and smooth norm on $\mathbb{R}^n$. Then the function $g_1(\cdot,\cdot)$ is continuous over T_0.

Proof. Let $(g_k,b_k)\epsilon T_0$, $\forall k$

and

$$(g_k,b_k) \to (g,b).\ \text{Then, of course, } (g,b)\epsilon T_0.$$

Note that we have

$$\rho_0(\cdot,\cdot) = f_1(\cdot,\cdot), \text{ over } T_0,$$
$$\beta_0(\cdot,\cdot) = f_1(\cdot,\cdot), \text{ over } T_0,$$

and

$$y_0(g,b) = g,\ \forall(g,b)\epsilon T_0.$$

So, $\rho_0(\cdot,\cdot)$, $\beta_0(\cdot,\cdot)$, $y_0(\cdot,\cdot)$ are all clearly continuous over T_0.

Also,

$$a_0(\cdot,\cdot) = \beta_0(\cdot,\cdot)\,(y_0(\cdot,\cdot))',$$

so $a_0(\cdot,\cdot)$ is continuous over T_0.

For $(g,b)\epsilon T_0$, we have

$$x_0(g,b) = f_3(a_0(g,b),b)$$

and thus $x_0(\cdot,\cdot)$ is continuous over T_0. $u_0(\cdot,\cdot)$ is also continuous over T_0, since $u_0(\cdot,\cdot) = x_0(\cdot,\cdot) - a_0(\cdot,\cdot)$. The same is true for $\gamma_0(\cdot,\cdot)$, since $\gamma_0(\cdot,\cdot) = \langle u_0(\cdot,\cdot), x_0(\cdot,)\rangle$.

Case 1. Suppose $u_0(g,b) \neq 0$. Then for all large enough k we have $u_0(g_k,b_k) \neq 0$ by the continuity of $u_0(\cdot,\cdot)$.

Subcase 1. Suppose $(\alpha_0(g_k,b_k))$ were determined using *Step 9* for infinitely many k's in an index set I. Extract a further subsequence $k_j \epsilon I$ such that

$$\alpha_0(g_{k_j},b_{k_j}) \to \alpha, \quad 0 \le \alpha \le 1.$$

By taking limits in the relation of *Step 9* and using continuity of $\gamma_0(\cdot,\cdot)$, $\beta_0(\cdot,\cdot)$, $u_0(\cdot,\cdot)$, $y_0(\cdot,\cdot)$ over T_0, and the fact that $u_0(g,b)$ and $y_0(g,b)$ are linearly independent as $u_0(g,b)$ is non–zero, we have

$$(\alpha\gamma_0(g,b)+(1-\alpha)\beta_0(g,b)) \langle(\alpha u_0(g,b)+(1-\alpha)y_0(g,b))',u_0(g,b) - y_0(g,b)\rangle =$$

$$= \|\alpha u_0(g,b) + (1-\alpha)y_0(g,b)\|' \,(\gamma_0(g,b) - \beta_0(g,b)). \tag{3.2.2.1}$$

But (3.2.2.1) says that

$$(\varphi_0(g,b))' \,(\alpha) = 0.$$

We know that $\alpha_0(g_k,b_k)$ was chosen to be the global maximizer of $\varphi_0(g_k,b_k)$ on $[0,1]$ and thus

$$(\varphi_0(g_k,b_k))(\alpha_0(g_k,b_k)) \ge (\varphi_0(g_k,b_k))(0) = \beta_0(g_k,b_k) \ge 0.$$

So, $\alpha_0(g_k,b_k)\gamma_0(g_k,b_k) + (1-\alpha_0(g_k,b_k))\beta_0(g_k,b_k) \ge 0$.

Allowing $j \to \infty$ in the subsequence (k_j) we get

$$\alpha\gamma_0(g,b) + (1-\alpha)\beta_0(g,b) \ge 0,$$

i.e. $\varphi_0(g,b)(\alpha) \geq 0.$

Let us assume

(3.2.2.2) $$\varphi_0(g,b)(\alpha) = 0.$$

Then by (3.2.2.1),

(3.2.2.3) $$\gamma_0(g,b) = \beta_0(g,b).$$

By (3.2.2.2) and (3.2.2.3),

$$\gamma_0(g,b) = \beta_0(g,b) = 0.$$

But $\beta_0(g,b) = 0$ implies $a_0(g,b) = 0$.

Since $\gamma_0(g,b) = \|u_0(g,b)\|_2^2 + \langle u_0(g,b), a_0(g,b)\rangle$,

$u_0(g,b) = 0$ which is a contradiction.

So, $\varphi_0(g,b)(\alpha) > 0$. Since also $(\varphi_0(g,b))'(\alpha) = 0$, then, by Lemma 1.5.3, α is the global maximizer of $\varphi_0(g,b)$ on $[0,1]$. Such an α is unique because the norm $\|\cdot\|$ is smooth and strictly convex. But we know that $\alpha_0(g,b)$ is the global maximizer, since $u_0(g,b) \neq 0$. So, $\alpha = \alpha_0(g,b)$. It follows that $\lim_{k \in I} \alpha_0(g_k,b_k) = \alpha = \alpha_0(g,b)$ as the only cluster point.

Subcase 2. Suppose $(\alpha_0(g_k,b_k))$ were determined using *Step 8* for infinitely many k's in an index set J, so that $\alpha_0(g_k,b_k) = 1$ for these k's. Taking limits in the Criterion of *Step 8* and by the continuity of $\gamma_0(\cdot,\cdot)$, $\beta_0(\cdot,\cdot)$, $u_0(\cdot,\cdot)$, $y_0(\cdot,\cdot)$, we get

(3.2.2.4) $$\gamma_0(g,b) \langle (u_0(g,b))', y_0(g,b)\rangle \geq \beta_0(g,b) \|u_0(g,b)\|'.$$

(3.2.2.4) means that

(3.2.2.5) $$(\varphi_0(g,b))'(1) \geq 0.$$

By *Step 6*, $\gamma_0(g_k,b_k) > 0$ and thus $\gamma_0(g,b) \geq 0$. If $\gamma_0(g,b) = 0$, then by (3.2.2.4), $0 \geq \beta_0(g,b)$. But $\beta_0(g,b) = f_1(g,b) \geq 0$ because $(g,b) \epsilon T_0$. So, $\beta_0(g,b) = 0$. As in *Subcase 1*, we get a contradiction from $\gamma_0(g,b) = 0 = \beta_0(g,b)$. So, $\gamma_0(g,b) > 0$ which means that

(3.2.2.6) $\qquad \varphi_0(g,b)(1) > 0.$

(3.2.2.6) and (3.2.2.5) imply that 1 is the global maximizer of $\varphi_0(g,b)$ on $[0,1]$ and thus $\alpha_0(g,b) = 1$, since $u_0(g,b) \neq 0$. So,

$\lim\limits_{k\epsilon J} \alpha_0(g_k,b_k) = \lim\limits_{k\epsilon J} 1 = 1 = \alpha_0(g,b).$

Subcases 1 and *2* together imply that

$$\lim \alpha_0(g_k,b_k) = \alpha_0(g,b), \text{ as } k \to \infty$$

and thus from *Step 10*

$$\lim g_1(g_k,b_k) = g_1(g,b), \text{ as } k \to \infty,$$

since $u_0(g,b)$ and $y_0(g,b)$ are linearly independent. Also $u_0(g_k,b_k)$, $y_0(g_k,b_k)$ are linearly independent eventually for all k.

Case 2. Suppose $u_0(g,b) = 0$. Then, $g_1(g,b) = y_0(g,b)$ and

$$\begin{aligned} x_0(g,b) = a_0(g,b) &= \beta_0(g,b)(y_0(g,b))' \\ &= (\min\{<y_0(g,b),x> \mid Ax = b, x \geq 0\})\ (y_0(g,b))'. \end{aligned}$$

Also, $\beta_0(g,b) = f_1(g,b) \geq 0$ because $(g,b)\epsilon T_0$.

Also, $Ax_0(g,b) = b$ and $x_0(g,b) \geq 0$.

Suppose $\beta_0(g,b) = 0$; then $x_0(g,b) = a_0(g,b) = 0$. But $x_0(g,b)$ is the solution of the problem $\min\{\|x\|_2 \mid Ax = b, x \geq 0\}$ since $a_0(g,b) = 0$. This implies that $b = 0$, a contradiction to $(g,b)\epsilon T_0$.

So, $\beta_0(g,b) > 0$. Since $\|x_0(g,b)\| = \beta_0(g,b) = \min\{<y_0(g,b),x> \mid Ax = b, x \geq 0\}$, $\|y_0(g,b)\|' = 1$, then due to the weak duality principle 2.3.5, $y_0(g,b)$ solves the problem

(P_0') $$\max_{\|y\|'=1} (\min\{<y,x> \mid Ax = b, x \geq 0\}).$$

Now let a subsequence (k_j) be chosen such that

$$g_1(g_{k_j},b_{k_j}) \to \bar{g},\ \bar{g}\epsilon F,\ \|\bar{g}\|' = 1.$$

As we know from (2.4.4.5),

$$\min\{<g_1(g_{k_j},b_{k_j}),x> \mid Ax = b_{k_j},\ x \geq 0\} \geq$$
$$\geq \min\{<y_0(g_{k_j},b_{k_j}),x> \mid Ax = b_{k_j},\ x \geq 0\} =$$
$$= \beta_0(g_{k_j},b_{k_j}) \geq 0 \text{ because } (g_{k_j},b_{k_j}) \epsilon T_0.$$

By taking the limit as $j \to \infty$, we get

$$\min\{<\bar{g},x> \mid Ax = b,\ x \geq 0\} \geq \tag{3.2.2.7}$$
$$\geq \min\{<y_0(g,b),x> \mid Ax = b,\ x \geq 0\} = \beta_0(g,b) \geq 0.$$

(3.2.2.7) can be true only if equality holds and $\bar{g}$ solves (P_0'). Then we have

$$x_0(g,b) = \beta_0(g,b)\ (y_0(g,b))'$$
$$= \beta_0(g,b)\ \bar{g}' \text{ as in Theorem 2.3.6.} \tag{3.2.2.8}$$

Since the norm $\|\cdot\|$ is smooth and $\beta_0(g,b) \neq 0$, (3.2.2.8) implies $\bar{g} = y_0(g,b)$. So, $\lim g_1(g_k b_k) = y_0(g,b) = g_1(g,b)$, as $k \to \infty$. This completes the proof that $g_1(\cdot,\cdot)$ is continuous over T_0.

3.2.3 DEFINITIONS. Let the norm $\|\cdot\|$ on $\mathbb{R}^n$ be strictly convex and smooth. We defined earlier the compact set T_0 and the functions

$$g_1(\cdot,\cdot)\colon T_0 \to F \cap \{z \epsilon \mathbb{R}^n \mid \|z\|' = 1\}$$

and

$$\beta_0(\cdot,\cdot)\colon T_0 \to [0, +\infty).$$

We have verified earlier that these functions are continuous on T_0. Let ℓ be an integer, $\ell \geq 1$. Suppose we have defined a compact set $T_{\ell-1}$ subset of $\mathbb{R}^n \times (\mathbb{R}^m)^\ell$ and a function

$$g_\ell(\underbrace{\cdot,\ldots,\cdot}_{\ell+1})\colon T_{\ell-1} \to F \cap \{z \epsilon \mathbb{R}^n \mid \|z\|' = 1\}$$

which is continuous over $T_{\ell-1}$.

Then define the set

$$T_\ell = \{(g,a_1,\ldots,a_\ell,b) \mid (g,a_1,\ldots,a_\ell) \in T_{\ell-1} \text{ and } b \in \{Ax \mid x \in \mathbb{R}^n_+\} \cap \overline{B} \text{ such that } f_1(g_\ell(g,a_1,\ldots,a_\ell),b) \geq 0\}.$$

T_ℓ is a subset of $\mathbb{R}^n \times (\mathbb{R}^m)^{\ell+1}$.

T_ℓ is compact, i.e. closed and bounded, $f_1(g_\ell(g,a_1,\ldots,a_\ell),b)$ being continuous as a function of $(g,a_1,\ldots,a_\ell,b)$ over the set

$$\{(g,a_1,\ldots,a_\ell,b) \mid (g,a_1,\ldots,a_\ell) \in T_{\ell-1},\ b \in \{Ax \mid x \in \mathbb{R}^n_+\}\}.$$

Noting that for $(g,a_1,\ldots,a_\ell,b) \in T_\ell$, $(g_\ell(g,a_1,\ldots,a_\ell),b) \in T_0$, we can make the following definitions:

Define the functions

$$\beta_\ell(\underbrace{\cdot,\ldots,\cdot}_{\ell+2}) : T_\ell \to [0,+\infty),$$

$$\beta_\ell(g,a_1,\ldots,a_\ell,b) = \beta_0(g_\ell(g,a_1,\ldots,a_\ell),b) = f_1(g_\ell(g,a_1,\ldots,a_\ell),b)$$

and

$$g_{\ell+1}(\underbrace{\cdot,\ldots,\cdot}_{\ell+2}) : T_\ell \to F \cap \{z \in \mathbb{R}^n \mid \|z\|' = 1\}.$$

$$g_{\ell+1}(g,a_1,\ldots,a_\ell,b) = g_1(g_\ell(g,a_1,\ldots,a_\ell),b),$$

for $(g,a_1,\ldots,a_\ell,b) \in T_\ell$

Clearly, $g_{\ell+1}(\cdot,\ldots,\cdot)$ is continuous over T_ℓ because of the continuity of $g_1(\cdot,\cdot)$ and $g_\ell(\cdot,\ldots,\cdot)$ over T_0 and $T_{\ell-1}$ respectively. Similarly, $\beta_\ell(\cdot,\ldots,\cdot)$ is continuous over T_ℓ.

So now we have defined compact sets T_ℓ and continuous functions β_ℓ, $g_{\ell+1}$ for all $\ell = 0,1,2,\ldots$.

3.2.4. Apply Algorithm 2.4.1 to the sequence (b_k), $k \geq 0$, $b_k \in \{Ax \mid x \in \mathbb{R}^n_+\} \cap \overline{B}$, $\forall k \geq 0$, $b_k \to b$, starting with $g_0 \in F$, $\|g_0\|' = 1$, and let k_0 be as in Lemma 2.5.4, $\rho_k \geq 0$, $\forall k \geq k_0$. Then, $\forall k \geq k_0$ and $\forall\, \ell \geq 0$ we have

(3.2.4.1) $$(g_k, b_k, b_{k+1},\dots, b_{k+\ell}) \in T_\ell$$

because $\rho_j \geq 0$ $\forall j=k,\dots,k+\ell$,

and moreover

(3.2.4.2) $$\beta_\ell(g_k, b_k. b_{k+1},\dots, b_{k+\ell}) = \beta_{k+\ell}$$

(3.2.4.3) $$g_{\ell+1}(g_k, b_k, b_{k+1},\dots, b_{k+\ell}) = g_{k+\ell+1}.$$

(g_k), $k \geq 0$, (β_k), $k \geq 0$, (ρ_k), $k \geq 0$ are the sequences generated by the Algorithm.

T_ℓ $\beta_\ell(\cdot,\dots,\cdot)$, $g_{\ell+1}(\cdot,\dots,\cdot)$ were defined in 3.2.3.

3.3 CONVERGENCE OF ALTHORITHM 2.4.1.

In this section we consider Algorithm 2.4.1 for a given sequence $b_k \in \{Ax|x\in\mathbb{R}^n_+\}$, $\forall k \geq 0$, with, $\lim b_k = b$. Suppose the algorithm starts from a given $g_0\in F$, $\|g_0\|' = 1$. We keep the assumption that the norm $\|\cdot\|$ on $\mathbb{R}^n$ is strictly convex and smooth.

We prove that the algorithm converges to the solution $\bar{x}$ of the problem

(P_0) $$Ax = b, x \geq 0, \|x\| \text{ (min)}.$$

From this and Remark 2.5.9, it is clear that Algorithm 2.4.9, applied to (b_k), $k \geq 0$, and g_0, also converges to $\bar{x}$.

3.3.1 THEOREM. Let $\|\cdot\|$ be a strictly convex and smooth norm on $\mathbb{R}^n$. Let $b\in\mathbb{R}^m$, $b \neq 0$ and assume that $Ax = b$, $x \geq 0$ is feasible. Let $b_k \in \{Ax|x\in\mathbb{R}^n_+\}$, $\forall k \geq 0$, such that $\lim b_k = b$. Let $g_0\in F$, $\|g_0\|' = 1$. Then, the sequence (x_k), $k \geq 0$, generated by Algorithm 2.4.1 converges to the solution $\bar{x}$ of the problem

(P_0) $$Ax = b, x \geq 0, \|x\| \text{ (min)}.$$

Proof. Let $0 < M < \|b\|_2$ and $\bar{B} = \{x\in\mathbb{R}^m | \|x - b\|_2 \leq M\}$.

Let i_0 as in Lemma 2.5.10 and let k_1 such that $b_k \epsilon \overline{B}, \forall k \geq k_1$. Let $k_2 = \max\{i_0, k_1\}$. It suffices to prove that Algorithm 2.4.1 applied to the data g_{k_2} and (b_k), $k \geq k_2$, converges. So, it constitutes no loss of generality in assuming that

$b_k \epsilon \overline{B}$, $\rho_k > 0$, $\rho_k = \beta_k$, $y_k = g_k$, $\min\{<y_k,x> \mid Ax = b, x \geq 0\} > 0, \forall k \geq 0$.

In accordance to this, the k_0 of 3.2.4 will be taken to be zero. Clearly,

$\forall \delta > 0$, $\exists\, i(\delta)$ depending on δ such that

(3.3.1.1) $\qquad \|b_k - b\|_2 < \delta, \ \forall k \geq i(\delta).$

For each $k \geq 0$, we consider Algorithm 2.4.3 for b starting from g_k.

Since $\min\{<g_k,x> \mid Ax = b, x \geq 0\} > 0$, we have that

(3.3.1.2) $\qquad (\beta_i^{g_k})$, $i \geq 0$, is a positive sequence increasing to $\|\overline{x}\|$.

(See 2.5.7 for the meaning of $(\beta_i^{g_k})$).

Also, $\forall \ell \geq 0, \forall k \geq 0$,

$$(g_k, \underbrace{b, \ldots, b}_{\ell+1 \text{ times}}) \ \epsilon T_\ell$$

From 3.2.4, we also have that $\forall \ell \geq 0, \forall k \geq 0$,

$$(g_k, b_k, b_{k+1}, \ldots, b_{k+\ell}) \ \epsilon T_\ell$$

Let $w = (y, c_0, c_1, \ldots, c_\ell) \epsilon T_\ell$, $y \epsilon \mathbb{R}^n$, $c_i \epsilon \mathbb{R}^m$, $\forall i$, $0 \leq i \leq \ell$.

Then define

$$|\|(y, c_0, c_1, \ldots, c_\ell)\|| := \|y\| + \sum_{i=0}^{\ell} \|c_i\|_2.$$

It is easy to verify that $|\|\cdot\||$ is a norm on $\mathbb{R}^n \times (\mathbb{R}^m)^{\ell+1}$.

Note that $T_\ell \subseteq \mathbb{R}^n \times (\mathbb{R}^m)^{\ell+1}$.

Now let $\ell \geq 0$ be a given integer and let $\epsilon > 0$ be given.

$\beta_\ell(\cdot, \ldots, \cdot)$ is uniformly continuous over the compact T_ℓ and thus

$\exists\, \eta(\epsilon, \ell) > 0$ depending on ϵ and ℓ, such that

$\forall\, w_1, w_2 \epsilon T_\ell$, $|\|w_1 - w_2\|| < \eta(\epsilon, \ell)$ we have that

(3.3.1.3) $$|\beta_\ell(w_1) - \beta_\ell(w_2)| < \epsilon/2.$$

In (3.3.1.1) pick $\delta = \eta(\epsilon,\ell)/(\ell+1)$. Then $\exists\ i(\delta) = i(\epsilon,\ell)$ i.e. depending on ϵ and ℓ, such that

$\forall k \geq i(\epsilon,\ell)$ we have that

$$\|b_k - b\|_2 < \eta(\epsilon,\ell)/(\ell+1) \text{ and also}$$
$$\|b_{k+1} - b\|_2 < \eta(\epsilon,\ell)/(\ell+1),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\|b_{k+\ell} - b\|_2 < \eta(\epsilon,\ell)/(\ell+1).$$

Summing up, we get that

$$|\,\|(g_k, b_k, b_{k+1},\ldots,b_{k+\ell}) - (g_k,\underbrace{b,\ldots,b}_{\ell+1 \text{ times}})\,\|\,| < \eta(\epsilon,\ell)$$

and thus, by (3.3.1.3), we have that

$$|\beta_\ell(g_k,b_k,\ldots,b_{k+\ell}) - \beta_\ell(g_k,\underbrace{b,\ldots,b}_{\ell+1 \text{ times}})| < \epsilon/2.$$

This last relation, via (3.2.4.2) and Definition 2.5.7, is the same as

$$|\beta_{k+\ell} - \beta_\ell^{g_k}| < \epsilon/2.$$

Summarizing, we have that

$\forall\ \ell \geq 0$, ℓ interger, $\forall \epsilon > 0$,

$\exists\ i(\epsilon,\ell)$ depending on ϵ and ℓ such that

(3.3.1.4) $$|\beta_{k+\ell} - \beta_\ell^{g_k}| < \epsilon/2, \forall k \geq i(\epsilon,\ell).$$

Now define

$$K_b := \{g \epsilon F \text{ such that } \|g\|' = 1 \text{ and } f_1(g,b) \geq 0\}.$$

K_b is compact as closed and bounded, since $f_1(\cdot,b)$ is continuous over F.

Note that from 3.2.4 and Algorithm 2.4.3 we have that if $g \epsilon K_b$, then

$$(g,\underbrace{b,\ldots,b}_{\ell+1 \text{ times}})\epsilon T_\ell,\ \forall \ell \geq 0.$$

Also note that, in our case, $g_k \epsilon K_b$, $\forall k \geq 0$, since

$$\min\{\langle g_k,x\rangle \mid Ax = b, x \geq 0\} = f_1(g_k,b) > 0.$$

Now let $\epsilon > 0$ and $g \epsilon K_b$ both be given.

In view of 2.5.7,

$\exists$ k (ϵ,g) depending on ϵ and g, such that

(3.3.1.5) $$\|\bar{x}\| - \beta^g_{k(\epsilon,g)} < \epsilon/4.$$

By uniform continuity of $\beta_{k(\epsilon,g)}(\cdot,\ldots,\cdot)$ over $T_{k(\epsilon,g)}$ we have that

$\exists\ \eta(\epsilon,k(\epsilon,g)) = \eta(\epsilon,g) > 0$ depending on ϵ and g, such that

$\forall\ w_1, w_2 \epsilon T_{k(\epsilon,g)}$, $|\|w_1 - w_2\|| < \eta(\epsilon,g)$ we have that

(3.3.1.6) $$|\beta_{k(\epsilon,g)}(w_1) - \beta_{k(\epsilon,g)}(w_2)| < \epsilon/4.$$

Now let

$$y \epsilon K_b,\ \|y - g\| < \eta(\epsilon,g).$$

Then,

$$|\|(\underbrace{y,b,\ldots,b}_{k(\epsilon,g)+1}) - (\underbrace{g,b,\ldots,b}_{k(\epsilon,g)+1\ \text{times}})\|| < \eta(\epsilon,g).$$

By (3.3.1.6),

$$|\beta_{k(\epsilon,g)}(g,b,\ldots,b) - \beta_{k(\epsilon,g)}(y,b,\ldots,b)| < \epsilon/4$$

which is the same as

(3.3.1.7) $$|\beta^g_{k(\epsilon,g)} - \beta^y_{k(\epsilon,g)}| < \epsilon/4.$$

(3.3.1.7) is true $\forall y \epsilon K_b \cap B(\epsilon,g)$ where

$B(\epsilon,g) := \{x \epsilon \mathbb{R}^n \mid \|x - g\| < \eta(\epsilon,g)\}$ is an open ball in the $\|\cdot\|$ norm with center g and radius $\eta(\epsilon,g)$. The notation $B(\epsilon,g)$ denotes that it depends on ϵ and g.

From (3.3.1.5) and (3.3.1.7) we get that

$$\forall\ \epsilon > 0,\ \forall\ g \epsilon K_b,$$

$\exists$ an integer $k(\epsilon,g)$ and a neighborhood of $g, B(\epsilon,g)$ such that

(3.3.1.8) $$\|\bar{x}\| - \beta^y_k < \epsilon/2,\ \forall\ k \geq k(\epsilon,g),\ \forall\ y \epsilon K_b \cap B(\epsilon,g).$$

By summing (3.3.1.5) and (3.3.1.7) we get (3.3.1.8) for $k = k(\epsilon,g)$;

after this the result is clear because for each y, (β_k^y), $k \geq 0$, increases to $\|\bar{x}\|$.

Let $\epsilon > 0$. Since K_b is compact, $\exists\, \ell_0$ and $g_1,\dots,g_{\ell_0} \epsilon K_b$ such that

$$K_b = \bigcup_{i=1}^{\ell_0} (K_b \cap B(\epsilon,g_i)). \tag{3.3.1.9}$$

Writing (3.3.1.8) for each g_i, $i = 1,\dots,\ell_0$, we get

$$\|\bar{x}\| - \beta_k^y < \epsilon/2,\ \ \forall k \geq k(\epsilon,g_i),\ \ \forall y \epsilon K_b \cap B(\epsilon,g_i). \tag{3.3.1.10}$$

Let $k(\epsilon) = \max\{k(\epsilon,g_i) \mid i = 1,\dots,\ell_0\}$; $k(\epsilon)$ depends on ϵ.

Combining (3.3.1.9) and (3.3.1.10) we get that

$\forall\, \epsilon > 0$, $\exists$ an integer $k(\epsilon)$ depending on ϵ, such that

$$\|\bar{x}\| - \beta_k^y < \epsilon/2,\ \ \forall k \geq k(\epsilon),\ \ \forall\, y \epsilon K_b.$$

In particular,

$$\|\bar{x}\| - \beta_i^{g_k} < \epsilon/2,\ \ \forall\, i \geq k(\epsilon),\ \ \forall\, k \geq 0. \tag{3.3.1.11}$$

Now let $\epsilon > 0$ be given. By choosing $\ell = k(\epsilon)$ in (3.3.1.4), we get that

$\exists\, i(\epsilon,k(\epsilon)) = i(\epsilon)$, i.e. depending on ϵ, such that

$$|\beta_{k+k(\epsilon)} - \beta_{k(\epsilon)}^{g_k}| < \epsilon/2,\ \ \forall\, k \geq i(\epsilon). \tag{3.3.1.12}$$

From (3.3.1.11) we get

$$\|\bar{x}\| - \beta_{k(\epsilon)}^{g_k} < \epsilon/2,\ \ \forall\, k \geq 0. \tag{3.3.1.13}$$

Adding (3.3.1.12) and (3.3.1.13), we conclude that

$\forall\, \epsilon > 0$, $\exists$ non–negative integers $k(\epsilon)$ and $i(\epsilon)$ such that

$$|\ \|\bar{x}\| - \beta_{k+k(\epsilon)}| < \epsilon,\ \ \forall\, k \geq i(\epsilon).$$

Letting $j(\epsilon) := i(\epsilon) + k(\epsilon)$, the last statement is the same as

$\forall\, \epsilon > 0$, $\exists\, j(\epsilon)$ such that

$$|\ \|\bar{x}\| - \beta_j| < \epsilon,\ \ \forall\, j \geq j(\epsilon),$$

i.e.

$$\lim \beta_k = \|\bar{x}\|. \tag{3.3.1.14}$$

Now assume $\lim a_{k_j} = a$. By taking further subsequences we can assume that $\lim y_{k_j} = \hat{y}$.

Then,

$$\lim \beta_{k_j} \qquad = \min\{<x,\hat{y}> \mid Ax = b,\ x \geq 0\}$$

$$(3.3.1.15) \qquad = \|\bar{x}\| \text{ by } (3.3.1.14).$$

Via the "Only if" part of Theorem 2.3.6 and since $b \neq 0$, (3.3.1.15) implies

$$(3.3.1.16) \qquad \bar{x} = \|\bar{x}\|\ \hat{y}'.$$

So,

$$\lim a_{k_j} = \lim \beta_{k_j} y'_{k_j} = \|\bar{x}\|\hat{y}' = \bar{x}$$

and thus

$$(3.3.1.17) \qquad \lim a_k = \bar{x}.$$

From (3.3.1.17) and the continuity of the function f_3 defined in 2.5.2, we get

$$\lim x_k = \lim f_3(a_k,b_k) = f_3(\bar{x},b).$$

$f_3(\bar{x},b)$ is the solution of the problem

$$Ax = b,\ x \geq 0,\ \|x - \bar{x}\|_2 \ (\min).$$

But the solution is clearly $\bar{x}$, since $A\bar{x} = b$, $\bar{x} \geq 0$.

So, $\lim x_k = \bar{x}$.

Now assume $\lim y_{k_j} = \hat{y}$. By (3.3.1.15) and the weak duality principle 2.3.5(i), we have that $\hat{y}$ is a solution of the problem

$$(P_0') \qquad \max\{\min\{<y,x> \mid Ax = b,\ x \geq 0\} \mid \|y\|' = 1\}.$$

By (3.3.1.16) and since the norm $\|\cdot\|$ is smooth, we get that any two cluster points of (y_k) are equal since they both satisfy (3.3.1.16).

Also, any solution of (P_0') satisfies (3.3.1.16) and thus it is unique.

Consequently, (y_k) converges to the unique solution of (P_0') which is dual to (P_0) as defined in 2.3.3.

The proof of the theorem is now complete.

3.3.2 THEOREM. Let $\|\cdot\|$ be a strictly convex norm on $\mathbb{R}^n$. Let $b_k \in \{Ax \mid x \in \mathbb{R}^n_+\}$, $\forall\, k \geq 0$, and $\lim b_k = 0$. Then the sequence (x_k) generated by Algorithm 2.4.1 converges to zero, which is the solution of the problem

$$Ax = 0,\ x \geq 0,\ \|x\|\ (\min).$$

Proof. Let us take a subsequence $\beta_{k_j} \to \beta$. Then $\beta \geq 0$.

By taking further subsequences we can assume $y_{k_j} \to \hat{y}$.

Then,

$$\beta_{k_j} = \min\{<y_{k_j},x> \mid Ax = b_{k_j},\ x \geq 0\}$$

implies

$$\beta = \min\{<\hat{y},x> \mid Ax = 0,\ x \geq 0\} \leq 0 \text{ by taking } x = 0.$$

So finally, $\beta = 0$ and thus $\lim \beta_k = 0$.

Then $\lim a_k = \lim \beta_k y'_k = 0$.

Consequently,

$$\lim x_k = \lim f_3(a_k,b_k) = f_3(0,0).$$

But $f_3(0,0)$ is the solution of

$$Ax = 0,\ x \geq 0,\ \|x\|_2\ (\min)$$

which is, of course, 0.

So $\lim x_k = 0$.

3.3.3. Theorem 3.3.2 says that Theorem 3.3.1 is still true if we remove the assumption b non–zero.

Theorems 3.3.1, 3.3.2 are also true for Algorithm 2.4.9 due to Remark 2.5.9. Note that in these two theorems we also have $\lim u_k = 0$.

3.4 AN ALGORITHM FOR NON–NEGATIVE LEAST ERROR MINIMAL NORM SOLUTIONS

3.4.1. Let $\|\cdot\|_0$ be a strictly convex norm on $\mathbb{R}^m$ and $\|\cdot\|$ be a strictly convex and smooth norm on $\mathbb{R}^n$. Suppose that the system $Ax = b, x \geq 0$ has no exact solution. We are interested in estimating the solution $\bar{x}$ of the following problem

(R) $Ax = A\tilde{x}, x \geq 0, \|x\|$ (min),

where $\tilde{x}$ is a solution of

(S) $x \geq 0, \|b - Ax\|_0$ (min).

In other words, $\bar{x}$ is the minimum norm solution of (S).

One way to do this is the following:

First we use Algorithm 4.1 in Sreedharan [6] to get an estimate of $A\tilde{x}$ and then apply Algorithm 1.4.1 to problem (R). This does not require smoothness of $\|\cdot\|$. Algorithm 4.1 in [6], if we remove its stopping criterion, is generating a sequence $(A\tilde{x}_k)$, $k \geq 0$, converging to $A\tilde{x}$. Applying Algorithm 2.4.1 to this sequence, we get a sequence (x_k), $k \geq 0$, converging to $\bar{x}$ via Theorems 3.3.1, 3.3.2. An algorithm which implements this idea is given below and it can be used for estimating $\bar{x}$. The norm $\|\cdot\|$ is also required to be smooth. The algorithm calculates $A\tilde{x}_k$ and if the duality gap for problem (S) is small enough, then it calculates x_k. After this, $A\tilde{x}_{k+1}$ is calculated and so on.

3.4.2 ALGORITHM.

$\eta > 0$, $0 < \epsilon < \epsilon_1$ and $\epsilon_2 > 0$ are given. $\|\cdot\|_0$ and $\|\cdot\|$–dual vectors will not be distinguished notationally.

Step 0. Let $\tilde{x}_0$ be a solution of the problem

$$x \geq 0, \|b - Ax\|_2 \text{ (min)}.$$

Let $r_0 := b - A\tilde{x}_0$ and $\tilde{y}_0 := r_0/\|r_0\|_0'$.

A more general choice of $\tilde{y}_0$ requires that $A^T\tilde{y}_0 \leq 0$ and $<b,\tilde{y}_0> > 0$, $\|\tilde{y}_0\|_0' = 1$. Let $\ell = 0$, $\tilde{k} = 0$.

Step 1. Let $g_0 \epsilon F$, $\|g_0\|' = 1$. Put $k = 0$.

Step 2. Let $b_k := b - <b,\tilde{y}_k> \tilde{y}_k'$

and $\tilde{x}_k$ be a solution of the problem

$$x \geq 0, \|b_k - Ax\|_2 \text{ (min)}.$$

Let $r_k := b_k - A\tilde{x}_k$.

Step 3. If $D_k := 1 - (<b,\tilde{y}_k>/ \|b - A\tilde{x}_k\|_0) \leq \epsilon$,

put $\ell = 1$ and GO TO *STEP 8*, *else*, proceed.

If $D_k \leq \epsilon_1$, GO TO *STEP 8*, *else*, put $g_{k+1} := g_k$,

increment k by 1 and proceed.

Step 4. If $<b,r_k> \leq 0$, then

$$\bar{\alpha}_k := <b,\tilde{y}_k> / <b,\tilde{y}_k - r_k>$$

and GO TO *STEP 6*, *else*, $\bar{\alpha}_k := 1$ and proceed.

Step 5. If $<b,r_k> <r_k',\tilde{y}_k> \geq <b,\tilde{y}_k> \|r_k\|_0'$,

set $\alpha_k := 1$ and GO TO *STEP 7*, *else*, proceed.

Step 6. Find α_k, $0 < \alpha_k < \bar{\alpha}_k$ such that

$$<b,\alpha_k r_k + (1 - \alpha_k) \tilde{y}_k> <(\alpha_k r_k + (1 - \alpha_k)\tilde{y}_k)', r_k - \tilde{y}_k>$$
$$= \|\alpha_k r_k + (1 - \alpha_k)\tilde{y}_k\|_0' < b,r_k - \tilde{y}_k>.$$

Step 7. Let

$$\tilde{y}_{k+1} = \frac{\alpha_k r_k + (1 - \alpha_k)\tilde{y}_k}{\|\alpha_k r_k + (1 - \alpha_k)\tilde{y}_k\|_0'}.$$

Increment $\tilde{k}$ by 1 and RETURN TO *STEP 2*.

Step 8. If

$$\|A\tilde{x}_k\|_1 := \sum_{i=1}^{m} |(A\tilde{x}_k)_i| > \epsilon_2,$$

GO TO *STEP 9*, *else*, proceed.

If $\ell = 1$, STOP; $x_k := 0$ solves (R). *Else*, proceed.

Put $x_k := 0$, $g_{k+1} := g_k$, increment k by 1 and GO TO *STEP 4*.

Step 9. Calculate $\rho_k := \min\{<g_k,x> \mid Ax = A\tilde{x}_k, x \geq 0\}$.

If $\rho_k \geq 0$,

$\beta_k := \rho_k$, $y_k := g_k$ and GO TO *STEP 11*, *else*, proceed.

Step 10. Let, z_k

be the solution of the problem

$$Ax = A\tilde{x}_k, \ x \geq 0, \ \|x\|_2 \ (\min).$$

Let $y_k := \dfrac{z_k}{\|z_k\|'}$, $\beta_k := \dfrac{\|z_k\|_2^2}{\|z_k\|'}$.

Step 11. Let $a_k := \beta_k y_k'$.

Let x_k be the solution of

$$Ax = A\tilde{x}_k, \ x \geq 0, \ \|x - a_k\|_2 \ (\min).$$

Let $u_k := x_k - a_k$.

Step 12. If $\dfrac{\|x_k\| - \beta_k}{\|x_k\|} > \eta$, GO TO *STEP 13*, *else*, proceed.

If $\ell = 1$, x_k is taken as the solution of (R) and STOP. *Else*, proceed.

Put $g_{k+1} := y_k$, increment k by 1 and GO TO *STEP 4*.

Step 13. Let $\gamma_k := <u_k,x_k>$.

Step 14. If $\gamma_k \leq 0$, let $\bar{\alpha}_k := \beta_k | (\beta_k - \gamma_k)$ and GO TO *STEP 16*,

else, $\bar{\alpha}_k := 1$ and proceed.

Step 15. If $\gamma_k <u_k',y_k> \geq \beta_k \|u_k\|'$, set $\alpha_k := 1$ and GO TO *STEP 17*, *else*,

proceed.

Step 16. Find $\alpha_k \epsilon (0, \bar{\alpha}_k)$ such that

$$(\alpha_k \gamma_k + (1-\alpha_k)\beta_k) <(\alpha_k u_k + (1-\alpha_k)y_k)', u_k - y_k>$$
$$= \|\alpha_k u_k + (1-\alpha_k)y_k\|' \ (\gamma_k - \beta_k).$$

Step 17. Let

$$g_{k+1}: = \frac{\alpha_k u_k + (1-\alpha_k)y_k}{\|\alpha_k u_k + (1-\alpha_k)y_k\|'} \text{ and increment } k \text{ by } 1.$$

If $\ell = 0$, RETURN TO *STEP 4*, *else*, proceed. $A\tilde{x}_{k+1}: = A\tilde{x}_k$ and RETURN TO *STEP 9*.

This combined algorithm essentially amounts to:

1st) An application of algorithm 4.1 of [6] to problem (S), from which we get a finite sequence $\{A\tilde{x}_0, A\tilde{x}_1, ..., A\tilde{x}_{k_0}\}$ with $A\tilde{x}_{k_0}$ regarded as $A\tilde{x}$, and

2nd) An application of Algorithm 2.4.1 to some finite sequence

$$\{A\tilde{x}_{j_1}, A\tilde{x}_{j_2}, ..., A\tilde{x}_{j_\ell}, A\tilde{x}_{k_0}, A\tilde{x}_{k_0}, ..., A\tilde{x}_{k_0}\}$$

with $0 \le j_1 < j_2 < ... < j_\ell \le k_0 - 1$.

It is not apparent which of the two alternative methods, suggested in this thesis for solving problem (R), is better.

If a "good" g_0 for doing *Step 1* is not at hand, we do the following:

1st) change *Step 1* as follows:

Step 1. Put $g_0: = 0, k = 0$.

2nd) change *Step 9* as follows:

Step 9. If $g_k = 0$, GO TO *STEP 10*, *else*, proceed.

Calculate $\rho_k: = \min\{<g_k, x> \mid Ax = A\tilde{x}_k, x \ge 0\}$ and the rest of *Step 9* is as before.

Instead of this, if $A\tilde{x}_0$ satisfies the Criterion of *Step 8*, we can replace *Step 1* with

Step 1. Put $k = 0$

and *Step 9* with

Step 9. If $k = 0$, GO TO *STEP 10*, *else*, proceed.

Calculate $\rho_k := \min\{\langle g_k, x\rangle \mid Ax = A\tilde{x}_k, x \geq 0\}$ and the rest of *Step 9* is as before.

3.5 NUMERICAL RESULTS

Algorithm 3.4.2 was coded in FORTRAN 77 in double precision for a SUN computer. The norm $\|\cdot\|_0$ was taken to be the $\|\cdot\|_p$ norm for various values of p, i.e.

$$\|x\|_p = (\Sigma_{i=1}^n |x_i|^p)^{1/p},$$

for $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$ and $1 < p < \infty$. The dual norm of $\|\cdot\|_p$ is the $\|\cdot\|_q$ norm for $q = p/(p-1)$. The $\|\cdot\|_p$–dual of $x \neq 0$ has components

$$x_i' = (|x_i|/\|x\|_q)^{q-1} \operatorname{sgn} x_i,$$

for $i = 1,\ldots,n$. (See also [8]).

The other norm $\|\cdot\|$ was taken to be either equal to $\|\cdot\|_0$ or dual to it. As in [6], the problems were done starting from $p = 2$ and then eventually increasing or decreasing p. For each new value of p, the $\tilde{y}_0$ and g_0 of *Steps 0* and *1* were taken to be the properly scaled terminal values of $\tilde{y}_k$ and g_k of the problem for the previous p. The ϵ and ϵ_1 of *Step 3* were taken to be 10^{-6} and 10^{-3} respectively. The η of *Step 12* was 10^{-6}. The α_k's of *Steps 6, 16* were calculated using subalgorithms 4.5 and 4.6 of [6] with tolerance $\eta = 10^{-9}$. *Steps 2* and *11* were executed using the NNLS and LDP programs of [2] respectively. The linear programs were done using the simplex algorithm.

In all the tested examples the algorithm was indeed capable of approximating the solutions, although convergence was at times very slow. On the same examples, the sequential alternative proposed in 3.4.1 was never found to be inferior to Algorithm 3.4.2 in rapidity of convergence. In some cases, Algorithm 3.4.2 did much worse than the sequential alternative.

Below we tabulate the results of Algorithm 3.4.5 applied to the following data:

$$A = \begin{matrix} 1 & 0 & 0.1 & 0.9 \\ 0 & 1 & 0.1 & 0.9 \\ 1 & 1 & 0.2 & 1.8 \\ 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & 0 & 0.2 & 1.8 \end{matrix} \qquad \text{and } b = \begin{matrix} 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 3 \end{matrix} \quad .$$

Note that A is not $1-1$ nor onto, as its rank is 2. The system $Ax = b$ is seen to be inconsistent. The NNLS solution of $Ax = b$ which was calculated using the NNLS program of [2], is

$$(0.064516, 0., 0., 1.505376)$$

and the least $\|\cdot\|_2$ NNLS solution is

$$(0.557673, 0.493157, 0.105083, 0.945747).$$

In this example no $A\tilde{x}_k$ was zero and no ρ_k was less than zero. That also was the case with most of the other examples that were tested.

In the table that follows, $x = (x_1, x_2, x_3, x_4)$ denotes the terminal value of x_k for each p. Also,

$s_1 := \|b - Ax\|_0$,

$s_2 := \|x\|$,

k_1 and k_2 denote the number of times that *Steps 6* and *16* were executed respectively.

k_3 is the number of times that *Step 16* was executed after ℓ was set equal to 1, and

$k_4 := k' - 1$, where k' is the number of times that the duality gap of *Step 12* was $\leq \eta$.

Case 1. $\|\cdot\|_0 = \|\cdot\|_p$, $\|\cdot\| = \|\cdot\|_p$

p	k_1	k_2	k_3	k_4	s_1	s_2	(x_1,x_2,x_3,x_4)
6	14	14	0	0	1.145958	.812068	(.640144, .640047, .463928, .719947)
5.5	12	11	0	1	1.165781	.826833	(.637454, .636924, .445596, .726099)
5	6	6	0	0	1.191253	.844885	(.634068, .632974, .423662, .733806)
4.8	6	6	0	0	1.203523	.853339	(.632562, .631062, .413666, .73751
4.5	7	6	0	1	1.224784	.867860	(.630434, .627651, .397171, .744081)
4	4	3	0	1	1.270246	.897045	(.624941, .62045, .364158, .757479)
3.8	4	4	0	0	1.293109	.911349	(.62249, .616665, .348717, .764323)
3.5	4	4	0	0	1.334227	.936461	(.618263, .609619, .32252, .776705)
3	3	3	0	0	1.428797	.991904	(.608964, .592084, .268588, .805761)
2.5	2	2	0	0	1.578439	1.07432	(.592872, .560547, .197272, .853557)
1.9	1	1	0	0	1.91694	1.244917	(.5447, .470442, .084868, .975027)
1.7	1	1	0	0	2.109877	1.332154	(.504821, .407001, .045666, 1.053866)
1.5	3	2	0	1	2.381531	1.437116	(.427267, .303775, .014517, 1.175417)
1.4	4	3	0	1	2.562909	1.491591	(.360989, .228195, .005188, 1.260717)
1.3	5	4	1	2	2.789927	1.541485	(.269809, .132402, .000901, 1.36654)
1.2	8	3	0	4	3.08179	1.580765	(.170894, .031528, .000035, 1.476788)
1.15	8	6371	6363	0	3.260797	1.595663	(.144062, .003817, 0., 1.506606)
1.1	10	1433	1424	0	3.468509	1.610757	(.141133, .000018, 0., 1.509853)
1.095	7	6	0	1	3.491111	1.612414	(.141225, .000023, 0., 1.509753)
1.09	7	6	0	1	3.514074	1.614101	(.141309, .000017, 0., 1.509661)

TABLE 1

Case 2. $\|\cdot\|_0 = \|\cdot\|_p, \|\cdot\| = \|\cdot\|_q$

p	k_1	k_2	k_3	k_4	s_1	s_2	(x_1,x_2,x_3,x_4)
6	*	12	0	2	*	1.468659	(.074163, .074066, .000023, 1.400359)
5.5		9	0	3		1.463255	(.0979, .097371, .000069, 1.375106)
5		5	0	1		1.455210	(.128449, .127354, .000204, 1.342657)
4.8		5	0	1		1.451075	(.142974, .141475, .000313, 1.327424)
4.5		5	0	2		1.443894	(.167815, .165032, .000595, 1.302166)
4		3	0	1		1.426226	(.216799, .212307, .001716, 1.251242)
3.8		3	0	1		1.417131	(.239757, .233932, .002614, 1.228037)
3.5		3	0	1		1.400453	(.278088, .269444, .004896, 1.189969)
3		3	0	0		1.361760	(.353518, .336639, .013801, 1.1179)
2.5		2	0	0		1.301961	(.445046, .412721, .038351, 1.035466)
1.9		1	0	0		1.183322	(.583329, .509071, .128348, .927275)
1.7		1	0	0		1.124865	(.637786, .539965, .191158, .889962)
1.5		3	1	1		1.049213	(.692718, .569226, .283527, .850581)
1.4		3	0	1		1.001097	(.716131, .583337, .34399, .82847)
1.3		4	0	1		.944481	(.734526, .597119, .415901, .804077)
1.2		4	0	3		.880758	(.748566, .609199, .502064, .77915)
1.15		4	0	4		.847235	(.753729, .613484, .55213, .76785)
1.1		4	0	5		.756889	(.756889, .615774, .608495, .758069)
1.095		3	0	4		.810186	(.757026, .615824, .61457, .757244)
1.09		3	0	4		.806876	(.757068, .615776, .620744, .756514)

* Identical to the corresponding column for Case 1.

TABLE 2

REFERENCES

REFERENCES

1. Gal, T., "Linear parametric programming, a brief survey", Sensitivity, Stability and Parametric Analysis, Mathematical Programming Study 21, a publication of the Mathematical Programming Society, A.V. Fiacco, Editor, North–Holland—Amsterdam, June 1984, 43–68.

2. Lawson, C.L., and Hanson, R.J., Solving Least Squares Problems, Prentice Hall, Englewood Cliffs, N.J. (1974).

3. Luenberger, D.G., Introduction to Linear and Nonlinear Programming, Addison Wesley: Reading, Mass. (1984)

4. Nirenberg, L., Functional Analysis, lectures given in 1960–61, notes by Lesley Sibner, New York University, (1961).

5. Rockafellar, R.T., Convex Analysis, Princeton University Press, Princeton, N.J., 1969.

6. Sreedharan, V.P., An Algorithm for non–negative norm minimal solutions, Numer. Funct. Anal. and Optimiz., 9 (1 & 2), 193–232 (1987).

7. Sreedharan, V.P., Least squares Algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm, Num. Math., 13, (1969), 146–151.

8. Sreedharan, V.P., Least squares algorithms for finding solutions of overdetermined systems of linear equations which minimize error in a smooth strictly convex norm, J. Approximation Theory, 8, (1973), 46–61.

9. Wets, J. B., "On the continuity of the value of a linear program and of related polyhedral–valued multifunctions", Mathematical Programming Essays in Honor of G.B. Dantzig, Part I, Mathematical Programming Study 24, a publication of the Mathematical Programming Society, R.W. Cottle, Editor, North Holland—Amsterdam, October 1985, 14–29.