This is to certify that the

dissertation entitled

LEAST DISTANCE ALGORITHMS FOR
LINEAR SYSTEMS

presented by

Said Bahi

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Mathematics

V. P. Sreedharan

Major professor

Date ____________

LEAST DISTANCE ALGORITHMS FOR LINEAR SYSTEMS

By

Said Bahi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1993

ABSTRACT

LEAST DISTANCE ALGORITHMS FOR LINEAR SYSTEMS

BY

Said Bahi

In this thesis, we investigate the numerical methods for three classes of approximation problems. We begin with the study of methods for approximating a non negative solution of an overdetermined system of linear equations. We define a best approximate solution of the system $Ax = b$, $x \geq 0$, to be the vector $x \geq 0$ which minimizes the norm of the residual $r(x) = b - Ax$, for a strictly convex and smooth norm.

We, then, consider a system of linear equations $Ax = b$ which has a non negative solution. We present a method for computing, amongst all these non negative solutions, the one which is of the least norm when the space $\mathbf{R}^n$ is equipped with a strictly convex norm.

Finally, given a system of linear inequalities, $Ax \geq b$, we suggest an algorithm that solves the question of the feasibility of the system, and if it is feasible, it computes the unique solution which has minimal norm.

For each of these three approximation problems, we study a dual problem whose solution leads to the characterization of the solutions of the original problem. Algorithms for computing the solutions are presented and their convergence proved. Numerical results for different l^p approximation problems are discussed.

ACKNOWLEDGMENT

I would like to thank my advisor, Professor V.P. Sreedharan, for his guidance in the reseach leading to this thesis and for all the discusions we have had on this particular subject and many others. I am especially grateful for his enthusiasm, which has made the last two years so enjoyable. I am also grateful to the members of my guidance committee, Professors D. Dunninger, G. Ludden, W. Sledd and D. Yen. I further thank the faculty and staff of the Department of Mathematics at Michigan State University, in particular Mrs. Barbara Miller, for their support of my effort. Finally, I wish to express my gratitude to Dr. D. Horner, Mrs. Elda Keaton and staff of the International Center at Michigan State for their help and to Mrs. Monique Bidaoui.

Contents

Chapter 0

Introduction

The Least Squares Problems have received a considerable amout of attention from different authors for either the special case of the Euclidean norm or the more general case when the norm considered is the l^p-norms, $1 \leq p \leq \infty$.

We shall study the numerical methods for computing solutions of the problem that we state generally as follows.

Let a system of linear equations (resp. linear inequalities) be given and let $\|.\|$ be a norm. If the system is overdetermined, find the 'best' non-negative solutions which minimize the $\|.\|$-error. If the system is feasible, find the unique solution that has minimal norm.

This problem will be considered for general abstract norms. For more practical use, this work will apply to the interesting case of the l^p-norms. First, we start by restating the original problem, specifying the stages of this thesis, as follows

Problem 1. Given a linear system $Ax = b$ that has no non-negative exact solution, find $x \geq 0$ for which Ax is the closest to b.

Problem 2. Given the system of linear equations

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

that is feasible, find the vector $x \geq 0$, solution to the system and which has the minimal norm.

Problem 3. Consider the system of linear inequalities $Ax \geq b$. Find the solution x with the least norm.

In problem 3, we make no feasibility assumption. Incidently the proposed method to solve this problem will answer the question of whether the system is consistent or not.

Our aim is to propose some implementable convergent algorithms that solve these problems. After the algorithms are specified, their feasibility and convergence are proved for any smooth, strictly convex norm. We will test these algorithms for the usual l^p-norm, $1 < p < \infty$

$$\|x\|_p = (\sum |x_i|^p)^{1/p} . \tag{0.1}$$

In the first chapter we will be concerned with problem 1 :

Let A be an $m \times n$ real matrix and let b be a real m-vector such that the system $Ax = b$, $x \geq 0$ has no exact solution. Let the norm $\|.\|$ on $\mathbf{R}^m$ be smooth and strictly convex. We seek $x \geq 0$ solution of

$$\|b - Ax\| \text{ (min) }, \ x \geq 0 . \tag{0.2}$$

A more general formulation of this problem can be stated as follows: Given a closed convex cone K in a finite dimensional Banach space X with a smooth, strictly convex norm $\|.\|$ and a point $b \in X \setminus K$, find $z \in K$ that minimizes the distance $\|b - x\|$, $x \in K$. We point out also that our results, including our algorithm can be formulated in infinite dimensional spaces as was done in [9].

This problem has been extensively studied for the norms $\|.\|_2$, $\|.\|_1$ and $\|.\|_\infty$. The last one is refered as the Chebyshev problem. Since the l^1 and the l^∞ norms fail to be strictly convex or smooth, differentiable optimization techniques cannot be applied. Special methods have been developed to handle these problems. In each of these cases, thanks to linear programming, the special structure of the norm may be used. See for example Barrodale and Young [1]. The case when the norm considered is the l^2 norm

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

has received the most consideration. For a more complete discussion see Lawson and Hanson [5] where an efficient finite algorithm called the NNLS algorithm as well as a Fortran implementation are proposed.

With no side constraints, the problem

$$\|b - Ax\|_p \text{ (min)}$$

for any p, $1 < p < \infty$, has also been studied by different authors. See Fletcher, Grant and Hebden [3], Owens [8], Owens and Sreedharan [10], Sreedharan [15], Späth [13] and Watson [21]. It presents a lesser difficulty since the set of approximation

$$K = \{ \ Ax \mid x \in \mathbf{R}^n \ \}$$

is a linear subspace.

If the condition $x \geq 0$ is added, the set of approximation is no longer a linear subspace. In this case we are seeking y in the cone

$$K = \{ \ Ax \mid x \geq 0 \ , x \in \mathbf{R}^n \ \}$$

which is the closest to b.

Sreedharan in [20] has developed an algorithm for solving this approximation problem for any smooth, strictly convex norm with applications to the minimization problem

$$\min\{ \ \|b - Ax\|_p \mid x \in \mathbf{R}^n \ , \ x \geq 0 \ \} \ .$$

In [20] a duality theory was developed and a characterization of the solution was given. The algorithm we propose in chapter 1 will use these ideas. We will give a dual problem and review the related duality theory. After stating the algorithm, we will show that all the steps of the algorithm are valid, then the proof of the convergence will be given. It is worth mentioning that our algorithm does not depend explicitly on the special structure of the l^p norm, but only on its strict convexity and smoothness. Numerical results for the norms $\|.\|_p$, for various p, will be presented in chapter 4.

General algorithms (such as in Polak [11], Zangwill [22], Zoutendijk [23]) that minimize smooth, strictly convex functions subject to linear constraints could be used to minimize the function $f(x) = \|b - Ax\|_p^p$, subject to $x \geq 0$. Or use may be made of nonsmooth optimization algorithms in Sreedharan ([16], [17], [19], [18]) and others. All of these algorithms for their guarantee of global convergence, however require the use of their built-in "antijaming" precautions, i.e., procedures that circumvent the possibility of the generated sequence clustering at or even converging to non-optimal points. In contrast to these, the algorithm to be presented here for the solution of the problem at hand has no antijaming precautions because none is needed, and it is globally convergent. General algorithms like those in Gill, et. al. [4] deal with the optimization of a twice continuously differentiable objective function subject to linear constraints and as such are not applicable to our problem, since in general our objective function, the norm, need not be twice differentiable througout $\mathbf{R}^m \setminus \{0\}$. In particular, for the l^p norms, $1 < p < 2$, there are non-zero points in $\mathbf{R}^m$, $m \geq 2$, where the norms fail to be twice differentiable. Despite this lack of differentiability, one may try to apply second order methods, the validity of which is predicated on the continuity of the Hessian of the objective function. Numerical answers obtained on specific examples are then checked a posteriori to see whether they are the "right" answers. Such tests lend credance to the applicability of second order methods, even when the relevent hypotheses are violated, but such findings do not prove the convergence of the algorithm. We, however, in the present work, give a proof of convergence of our algorithm, which includes the l^p case, $1 < p < 2$.

In chapter 2, we will be considering the following minimization problem

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

The norm is assumed to be smooth, strictly convex and the system $Ax = b$ to have a non negative exact solution. This problem has been studied for the l^2-norm with linear inequalities constraints in [5] and for general norms in [6]. Lawson and Hanson in [5] gave an algorithm that solves the problem $\|x\|_2$ (min), $Ax \geq b$, called the LDP (Least Distance Programming) algorithm. It uses the transformation of the present problem into a NNLS (Non Negative Least Squares) problem. A Fortran code was also presented.

As we pointed out in the discussion of problem 1, general purpose algorithms may also be applied to solve problem 2, when the norm considered is the l^p-norm. However, the objective function $f(x) = \|x\|_p^p$ poses the same differentiability problems for $1 < p < \infty$. Moreover, as it was outlined in [6], during machine implementation, some precautions would have to be taken to circumvent the possibility that the generated sequence of approximations be clustering or converging to a non optimal solution, though we prove in theory the convergence of the sequence. On the other hand, our statements including the proposed algorithm do not depend explicitly on the norm, need no extra differentiability of the objective function and the global convergence of the algorithm is proved. This work was motivated by similar algorithms in [6] and [20].

A major step in this algorithm involves the solution, at each iteration cycle k, of an l^2 problem

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

where a_k is defined at the begining of each iteration. The feasibilty of the algorithm will be studied and its convergence proved.

Next, we examine in chapter 3 the problem

$$\|x\|_p(\min) \; , \; Ax \geq b$$

for any p, $1 < p < \infty$. Lawson and Hanson studied a similar problem when $p = 2$. Our work is an extension of the so called LDP algorithm in [5]. This problem is solved via the algorithm in chapter 1. A dual problem will be presented and its correspondance with the original problem will be studied. A related least squares problem will be given and used, as mentioned above, to find the solution to our problem. After presenting the algorithm, we will prove its feasibility and that it computes the solution if it exists. The fact that no assumption regarding the feasibility of the system

$$Ax \geq b$$

is involved will explain the usefulness of this algorithm. Moreover its implementation is particularly simple.

In chapter 4, some numerical results are given and a subalgorithm is presented. These results show that the coded algorithm does well compared to other algorithms in the literature, for different p-norms, when p is in the range $(2, \infty)$ and not far away from 2.

Chapter 1

A Least Distance Algorithm For A Smooth Strictly Convex Norm

We study in this chapter the system of linear equations

$$Ax = b \tag{1.1}$$

where x is a non-negative n-vector in $\mathbf{R}^n$, A is an $m \times n$ and b is in $\mathbf{R}^m$. Let $\|.\|$ be a smooth, strictly convex norm on $\mathbf{R}^n$. In many applications the system above is overdetermined. In this case, one seeks to minimize the error $r(x) = b - Ax$.

We shall be concerned with the following problem

(P) $\qquad \min \{\|b - Ax\| \mid x \in \mathbf{R}^n, x \geq 0\}.$

Our objective is to give an implemetable algorithm to solve (P), and prove its convergence. In practice the proposed algorithm computes simultanuously a solution of (P), a solution of its dual (P') and their common value. At each iteration cycle, an approximation to the solution of (P) is computed and then used to find an approximation of (P'). The later enables us to compute a new, improved, approximation of (P).

In [20], Sreedharan suggested a dual problem (P') , and the relation between problems (P) and (P') was studied. We will review briefly problem (P') .

We begin by establishing some notations and definitions. Given two vectors x and y in $\mathbf{R}^n$, the usual inner product is denoted by $\langle .,. \rangle$

$$\langle x, y \rangle = \sum x_i . y_i.$$

A partial order in $\mathbf{R}^n$ is defined as follows. For any $u = (u_1, ... u_n) \in \mathbf{R}^n, u \geq 0$ iff $u_i \geq 0, i = 1, ..., n$. If A is a matrix, its transpose is denoted by A^T. Associated with the above inner product is the usual Euclidean norm $\|.\|_2$. In formulating the duality theorems, we will use the well known notion of dual norm. Let $\|.\|$ be a norm on $\mathbf{R}^m$, $m \geq 1$, *the dual norm* $\|.\|'$ on $\mathbf{R}^m$ is defined by

$$\|y\|' = max\{\langle x, y \rangle | \quad \|x\| = 1, \quad x \in \mathbf{R}^m\}.$$

For any y in $\mathbf{R}^m$, $y \neq 0$, a $\|.\| -$ *dual vector* y' is then defined by

$$\|y'\| = 1, \ and \ \langle y', y \rangle = \|y\|'. \tag{1.2}$$

Similarly, a $\|.\|' -$ *dual vector* y^* is given by

$$\|y^*\|' = 1 \ , \ and \ \langle y^*, y \rangle = \|y\|. \tag{1.3}$$

Recall that a norm $\|.\|$ is *strictly convex* iff

$$\|x\| = \|y\| = \|x + y\|/2 \quad implies \quad x = y,$$

and the norm $\|.\|$ is *smooth* if and only if through each point of the unit norm there passes a unique hyperplane supporting the closed unit ball $B = \{y \in \mathbf{R}^m | \quad \|y\| \leq 1\}$. We note that $\|.\|$ is smooth if and only if $\|.\|'$ is strictly convex. If the norm $\|.\|$ is strictly convex, then the $\|.\|$-duals are unique, the map $y \mapsto y'$ is odd, continuous on $\mathbf{R}^m \setminus \{0\}$, and positively homogenous of degree zero. If the norm $\|.\|$ is smooth, then the $\|.\|'$-duals are unique, and the map $y \mapsto y^*$ is odd, continuous, positively homogenous of degree zero on $\mathbf{R}^m \setminus \{0\}$. It is easily seen that for $y \neq 0$,

$$y'^* = y/\|y\|' \ and \quad y^{*'} = y/\|y\|, \tag{1.4}$$

when the norm $\|.\|$ is smooth and strictly convex, respectively.

In the special case when $\|.\|$ is the l^p-norm, with $1 < p < \infty$, it follows easily from the Holder inequality that $\|.\|_p' = \|.\|_q$, where $p + q = pq$. In this case, the dual vectors take the particularly simple forms

$$y_i' = (|y_i|/\|y\|_q)^{q-1} sgny_i, \qquad i = 1, ..., m,$$

$$y_i^* = (|y_i|/\|y\|_p)^{p-1} sgny_i, \qquad i = 1, ..., m. \tag{1.5}$$

for any nonzero vector $y \in \mathbf{R}^m$.

A *convex cone* K in $\mathbf{R}^m$ is a convex subset for which $\lambda y \in K, \forall y \in K$ and $\forall \lambda \geq 0$. The *negative polar* of K is defined by $K^\circ = \{z \in \mathbf{R}^m \mid \langle z, y \rangle \leq 0 , \ \forall y \in K\}$. In particular, if

$$K = \{Ax \in \mathbf{R}^m \mid x \in \mathbf{R}^n , x \geq 0\},$$

then its negative polar is given by

$$K^\circ = \{z \in \mathbf{R}^m \mid A^T z \leq 0\}.$$

Following Lawson and Hanson [5], if the system $Ax = b$ is overdetermined, then the minimizers of (P) are called the non-negative least squares (NNLS), when the norm is the l^2-norm. The corresponding residual is called the NNLS-residual. In general the NNLS solutions are not unique. A finite algorithm for solving the NNLS problem is given in [5]. At the computational stage of this chapter, we will use this algorithm.

The following dual problem (P') can be now introduced [20]

$$(P') \qquad max\{\langle b, z\rangle \mid z \in \mathbf{R}^m , A^T z \leq 0 , \|z\|' = 1\}.$$

Theorem 1.1 *(Nirenberg [7]) Let K be a convex cone in $\mathbf{R}^m$. Let $b \in \mathbf{R}^m$, then the dual problems*

$$min\{\|b - y\| \mid y \in K\}, \tag{1.6}$$

and

$$max\{\langle b, z\rangle \mid z \in K^{\circ}, \|z\|' = 1\}, \tag{1.7}$$

have the same value for any norm $\|.\|$ *on* $\mathbf{R}^m$.

From this, the following corollary is immediate

Corollary 1.2 *Let* $\mathbf{R}^m$ *be equipped with a norm* $\|.\|$.*Then the dual pair (P) and* (P') *have the same value.*

An important result in [20] will be used to solve (P). It characterizes the solutions of (P) in the case when the norm $\|.\|$ is strictly convex and the system $Ax = b$ has no non-negative exact solution.

Theorem 1.3 *Let the norm* $\|.\|$ *be strictly convex. The following are equivalent*

(i) $\bar{x} \geq 0$ *is a non-negative least error* $\|.\|$ *solution of the system* $Ax = b$.

(ii) There exists $y \in \mathbf{R}^m$, $\|y\|' = 1$, $A^T y \leq 0$, *such that*

$$A\bar{x} = b \; - \langle b, y\rangle y'. \tag{1.8}$$

Moreover, $\langle y, A\bar{x}\rangle = 0$ *and* y *solves* (P'). *If in addition the norm* $\|.\|$ *is smooth, then (i) and (ii) are equivalent to*

(iii) Let $r = b - A\bar{x}$, *then*

$$A^T r^* \leq 0, \tag{1.9}$$

and

$$\langle r^*, A\bar{x}\rangle = 0. \tag{1.10}$$

An immediate consequence of the theorem follows easily when $\|.\| = \|.\|_2$. In this case clearly $r^* = r/\|r\|_2$, for any $r \neq 0$. The vector $\bar{x} \geq 0$ is a solution of (P) in the

l^2-sense iff the residual $r = b - A\bar{x}$ is such that

$$A^T r \leq 0 \quad \textit{and} \quad \langle r, A\bar{x} \rangle = 0. \tag{1.11}$$

Now, we give our algorithm for solving (P). It is assumed that the norm $\|.\|$ is smooth, strictly convex and that the system $Ax = b$, $x \geq 0$ has no exact solution. The main results are the feasability of the algorithm and the theorem concerning its convergence. Proofs that all the steps of the algorithm are valid will also be given.

Algorithm 1.1

Step 0. Find x_0 a NNLS solution of $Ax = b$. Let $r_0 = b - Ax_0$, $y_0 = r_0/\|r_0\|'$, $k = 0$.

Step 1. Find x_k a NNLS solution of

$$Ax = b \; - \langle b, y_k \rangle y_k' \; .$$

Let r_k be the residual

$$r_k = b \; - \langle b, y_k \rangle y_k' - Ax_k \; .$$

Step 2. If $r_k = 0$, GO TO step 8 ; else continue.

Step 3. If

$$\langle b, r_k \rangle \;\; \geq \;\; \langle b, y_k \rangle \|r_k\|' + \|r_k\|_2^2/4 \; .$$

Set $y_{k+1} = r_k/\|r_k\|'$ and GO TO step 7.

Step 4. Let

$$\mu_k = (\langle b, r_k \rangle - \|r_k\|_2^2/2)/\langle b, y_k \rangle \; .$$

Step 5. Find $\alpha_k > 0$ such that

$$\|y_k + \alpha_k r_k\|' = 1 + \alpha_k \mu_k \; .$$

Step 6. Set

$$y_{k+1} = (y_k + \frac{1}{2}\alpha_k r_k)/\|y_k + \frac{1}{2}\alpha_k r_k\|' \ .$$

Step 7. Increase k by 1, and RETURN to step 1.

Step 8. x_k is a non-negative $\|.\|$ minimal solution of $Ax = b$. The computation is complete.

We make a few observations. A crucial statement in the algorithm is step 5 defining the number α_k. Any efficient algorithm for finding the zeros of a function can be used in the current situation . Later, we will suggest a procedure to compute the α_k occuring in step 5. The stopping criterion given in step 2 is a very convenient one in the proof of the convergence. From a computational point of view, a more practical stopping rule will be to require that the so called relative duality gap goes to zero. So, in writing a routine to implement the algorithm, step 2. will be replaced by

Step 2'. If

$$(\|b - Ax_k\| - \langle b, y_k\rangle)/\|b - Ax_k\| \quad \leq \epsilon \ ,$$

where ϵ is a fixed stopping rule parameter, GO TO step 8 (the computation is complete); else proceed.

All the other steps of the algorithm remain unchanged.

Remark. Later in proposition 1.5, we shall show that the sequence $(\langle b, y_k\rangle)$ generated by the algorithm is strictly increasing. We claim that

$$\langle b, y_k\rangle \ > 0 \ , \qquad for \ \ all \ \ k \ .$$

To see this, let r_0 be the NNLS residual of the system $Ax = b$, $x \geq 0$, as defined in step 0. Recall that the system is assumed to have no non-negative exact solution, so

$r_0 \neq 0$. From equation (1.11) we get

$$\langle b, r_0 \rangle = \|r_0\|_2^2 > 0 \ .$$

Once more, by step 0 of the algorithm, $y_0 = r_0/\|r_0\|'$, so that $\langle b, y_0 \rangle > 0$, which proves the claim.

We now turn our focus to the convergnce theory. We begin by proving the validity of the various steps of the algorithm. In the next theorem, we prove that under the assumption that the primal norm $\|.\|$ is smooth, the number $\alpha_k > 0$ defined by step 5 always exists.

Theorem 1.4 *Let the norm $\|.\|$ on $\mathbf{R}^m$ be smooth. Then if the algorithm is at the stage of entering step 5, $\exists \alpha_k > 0$ such that*

$$\|y_k + \alpha_k r_k\|' = 1 + \alpha_k \mu_k \tag{1.12}$$

where μ_k is defined by

$$\mu_k = (\langle b, r_k \rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle \ .$$

Proof. Since the norm $\|.\|$ is smooth, the dual norm is strictly convex. Consider the strictly convex function $f : \mathbf{R} \longrightarrow \mathbf{R}$ defined by

$$f(\lambda) = \|y_k + \lambda r_k\|' \tag{1.13}$$

and let

$$l(\lambda) = 1 + \lambda \mu_k \ .$$

Then $f(0) = l(0) = 1$, $l'(0) = \mu_k$ and $f'(0) = \langle y_k', r_k \rangle$. (For details about f', see [14]). We note that the algorithm enters step 5 only if $r_k \neq 0$. By step 1 and (1.11), we obtain

$$\|r_k\|_2^2 = \langle b, r_k \rangle - \langle b, y_k \rangle \langle y_k', r_k \rangle \ . \tag{1.14}$$

Using(1.14), we have

$$
\begin{aligned}
f'(0) - l'(0) &= \langle y_k', r_k \rangle - \mu_k \\
&= \langle y_k', r_k \rangle - (\langle b, r_k \rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle \\
&= (\langle y_k', r_k \rangle \langle b, y_k \rangle - \langle b, r_k \rangle + \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle \\
&= -\frac{1}{2}\|r_k\|_2^2/\langle b, y_k \rangle \\
&< 0\,.
\end{aligned}
$$

This implies that $f(\alpha) - l(\alpha) < 0$ for some $\alpha > 0$. Note that step 5 is executed only if step 3 is answered negatively, i.e,

$$
\langle b, r_k \rangle < \langle b, y_k \rangle \|r_k\|' + \frac{1}{4}\|r_k\|_2^2\,. \tag{1.15}
$$

In this case, since $\|y_k\|' = 1$, for any $\lambda > 0$ we have

$$
\begin{aligned}
f(\lambda) - l(\lambda) &= \|\lambda r_k + y_k\|' - \lambda\mu_k - 1 \\
&= \|\lambda r_k + y_k\|' - \lambda(\langle b, r_k \rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle - 1 \\
&\geq \lambda\|r_k\|' - \|y_k\|' - \lambda(\langle b, r_k \rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle - 1 \\
&= \lambda(\|r_k\|'\langle b, y_k \rangle - \langle b, r_k \rangle + \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle - 2 \\
&> \lambda(-\frac{1}{4}\|r_k\|_2^2 + \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle - 2\,,
\end{aligned}
$$

due to (1.15). Hence

$$
f(\lambda) - l(\lambda) \;>\; \frac{\lambda}{4}(\|r_k\|_2^2/\langle b, y_k \rangle) - 2
$$

goes to ∞ as $\lambda \longrightarrow \infty$.

Thus, there exists $\lambda > 0$ such that $f(\lambda) - l(\lambda) > 0$. Because the function $f(\lambda) - l(\lambda)$ is continuous, this implies that there exists an $\alpha_k > 0$ such that (1.12) holds as claimed, and the proof is complete.

Proposition 1.5 *Let r_k and y_k be as defined in the algorithm. Then*

(a) $y_k + \alpha r_k \neq 0, \quad \forall\alpha,$ *if* $r_k \neq 0.$

(b) $A^T y_k \leq 0$ *and* $\|y_k\|' = 1$, $\forall k.$

Proof. To prove (a), suppose $r_k = \alpha y_k$ for some α. The algorithm defines r_k by

$$r_k = b - \langle b, y_k \rangle y_k' - Ax_k \; .$$

Now, because of the second half of (1.11), we get

$$\|r_k\|_2^2 = \langle b - \langle b, y_k \rangle y_k', r_k \rangle \; . \tag{1.16}$$

Hence

$$\begin{aligned} \|r_k\|_2^2 &= \alpha\langle b, y_k \rangle - \alpha\langle b, y_k \rangle\langle y_k', y_k \rangle \\ &= 0 \, , \end{aligned}$$

since $\langle y_k', y_k \rangle = \|y_k\|' = 1$. This implies that $r_k = 0$, a contradiction.

To prove (b), once more by (1.11) we have $A^T r_0 \leq 0$, so that by step 0 of the algorithm $A^T y_0 \leq 0$. A simple induction shows that

$$A^T y_k \leq 0, \; \forall k \textit{ and } \|y_k\|' = 1 \; .$$

The proof is complete.

The next result shows that the sequence (y_k) built by the algorithm leads to an improved approximation of the value of (P'), after each iteration cycle.

Proposition 1.6 *Let (y_k) be the sequence generated by the algorithm. Then, the sequence $(\langle b, y_k \rangle)$ is strictly increasing and convergent.*

Proof. Here two cases have to be considered. Either (a) the question in step 3 is answered affirmatively, in which case,

$$\langle b, r_k \rangle \geq \langle b, y_k \rangle \|r_k\|' + \frac{1}{4}\|r_k\|_2^2 \; ; \tag{1.17}$$

or (b) the question in step 3 is answered negatively, so that

$$\langle b, r_k \rangle < \langle b, y_k \rangle \|r_k\|' + \frac{1}{4}\|r_k\|_2^2 . \tag{1.18}$$

It is easy to see that case (a) yields a strictly increasing sequence. By step 4 of the algorithm

$$y_{k+1} = r_k / \|r_k\|' .$$

Inequality (1.17) implies

$$\begin{aligned} \langle b, y_{k+1} \rangle &= \langle b, r_k \rangle / \|r_k\|' \\ &\geq \langle b, y_k \rangle + \frac{1}{4}\|r_k\|_2^2 / \|r_k\|' . \end{aligned}$$

Thus

$$\langle b, y_{k+1} \rangle > \langle b, y_k \rangle ,$$

(since $r_k \neq 0$), as desired.

We now take up the second case (b). Because the norm $\|.\|$ is smooth, the dual norm $\|.\|'$ is strictly convex. By proposition 1.5-(a) y_k and r_k are linearly independent. We have

$$\begin{aligned} \|y_k + \frac{1}{2}\alpha_k r_k\|' &= \|\frac{1}{2}(y_k + y_k + \alpha_k r_k\|' \\ &< \frac{1}{2}\|y_k + \alpha_k r_k\|' + \frac{1}{2}\|y_k\|' \\ &= 1 + \frac{1}{2}\alpha_k \mu_k \quad (\textit{due to step } 5) \\ &= 1 + \frac{1}{2}\alpha_k(\langle b, r_k \rangle - \frac{1}{2}\|r_k\|_2^2) / \langle b, y_k \rangle . \end{aligned}$$

We conclude that

$$\langle b, y_k \rangle \|y_k + \frac{1}{2}\alpha_k r_k\|' < \langle b, y_k \rangle + \frac{1}{2}\alpha_k \langle b, r_k \rangle - \frac{1}{4}\alpha_k \|r_k\|_2^2 , \tag{1.19}$$

which, in view of step 6, implies that

$$\begin{aligned} \langle b, y_k \rangle &< (\langle b, y_k + \frac{1}{2}\alpha_k r_k \rangle - \frac{1}{4}\alpha_k \|r_k\|_2^2) / \|y_k + \frac{1}{2}\alpha_k r_k\|' \\ &< \langle b, y_{k+1} \rangle , \end{aligned}$$

as claimed.

To prove the convergence of the sequence $(\langle b, y_k\rangle)$, we simply observe that because $\|y_k\|' = 1$, we have

$$\begin{aligned}\langle b, y_k\rangle &\leq \|b\|.\|y_k\|' \\ &= \|b\| .\end{aligned}$$

The proof is complete.

The next result will be needed in the proof of the convergence theorem. It also gives a further insight into the existence and the uniqueness of the number $\alpha_k > 0$, defined in step 5.

Lemma 1.7 *Let $\alpha_k > 0$ be as defined by step 5 of the algorithm. Then α_k satisfies the following inequality*

$$\langle (y_k + \alpha_k r_k)', r_k\rangle\langle b, y_k\rangle + \frac{1}{2}\|r_k\|_2^2 \geq \langle b, r_k\rangle . \tag{1.20}$$

Proof. By the definition of the dual norm, we have

$$\|y_k + \alpha_k r_k\|' = \langle (y_k + \alpha_k r_k)', y_k\rangle + \alpha_k\langle (y_k + \alpha_k r_k)', r_k\rangle . \tag{1.21}$$

Using the inequality

$$\begin{aligned}\langle (y_k + \alpha_k r_k)', y_k\rangle &\leq \|(y_k + \alpha_k r_k)'\|.\|y_k\|' \\ &= 1 ,\end{aligned}$$

and (1.12), we obtain

$$\begin{aligned}1 + \alpha_k(\langle b, r_k\rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k\rangle &= \|y_k + \alpha_k r_k\|' \\ &\leq 1 + \alpha_k\langle (y_k + \alpha_k r_k)', r_k\rangle , \end{aligned} \tag{1.22}$$

which proves the desired inequality (1.20).

We are now able to prove the convergence theorem.

Theorem 1.8 *Either*

(a) the algorithm solves (P) and (P') *in a finite number of iterations; or*

(b) the algorithm generates infinite sequences (x_k) *and* (y_k) *such that the sequence* (y_k) *converges to the solution of* (P') *and every cluster point of* (x_k) *is a solution of (P).*

Proof. To prove (a), we note that the sequences (x_k) and (y_k) are finite sequences (i.e, the algorithm terminates in a finite number of iterations) if and only if $r_k = 0$, for some fixed integer k. In this case by theorem 1.3, x_k and y_k are solutions of (P) and (P') respectively.

To prove (b), let (x_k), (y_k), (r_k) and (α_k) be the infinite sequences genarated by the algorithm.

Let

$$d = \min \{\|b - Ax\| \mid x \geq 0, \; x \in \mathbf{R}^n\} \; . \tag{1.23}$$

Since the system $Ax = b$, $x \geq 0$ is assumed to have no non-negative exact solution, $d > 0$. By (1.16)

$$\begin{aligned} \|r_k\|_2^2 &= \langle b - \langle b, y_k \rangle y_k', r_k \rangle \\ &\leq (\|b\|_2 + d\|y_k'\|_2)\|r_k\|_2 \end{aligned}$$

Let $M > 0$ be such that $\|z\|_2 \leq M\|z\|$, for any $z \in \mathbf{R}^m$. Then, because $\|y_k'\| = 1$, we obtain

$$\|r_k\|_2 \leq \|b\|_2 + Md \; . \tag{1.24}$$

Hence, the sequence (r_k) is bounded. Our goal is to prove that (r_k) converges to zero. Suppose that this not true, i.e, the sequence (r_k) does not converge to zero. then, by (1.24), there exists a subsequence, denoted again (r_k) such that

$$r_k \longrightarrow r \neq 0$$

There are two cases to investigate.

Case 1. Assume that step 5 of the algorithm is executed for an infinite number of indices (k), so that

$$\langle b, r_k \rangle \ < \langle b, y_k \rangle \|r_k\|' + \frac{1}{4}\|r_k\|_2^2 \ . \tag{1.25}$$

By taking a subsequence if necessary, we may assume that

$$\|r_k\|_2^2 \ \geq \ \delta \qquad \forall k \geq 0 \ , \ for\ some\ \delta > 0 \ . \tag{1.26}$$

Now,

$$\begin{aligned} \alpha_k \|r_k\|' - 1 \ &\leq \ \|y_k + \alpha_k r_k\|' \\ &= \ 1 + \alpha_k(\langle b, r_k \rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle \end{aligned}$$

(because of step 5), from which one obtains

$$\begin{aligned} 2 \ &\geq \ \alpha_k(\|r_k\|'\langle b, y_k \rangle - \langle b, r_k \rangle + \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle \\ &> \ \alpha_k(-\frac{1}{4}\|r_k\|_2^2 + \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k \rangle \ , \end{aligned}$$

(due to (1.25)). By the weak duality and (1.23) this implies

$$\begin{aligned} \alpha_k \|r_k\|_2^2 \ &< \ 8\langle b, y_k \rangle \\ &< \ 8d \ . \end{aligned}$$

Using (1.26), we see that

$$0 < \alpha_k < 8d/\delta \ . \tag{1.27}$$

We have ,thus, shown that the positive sequence (α_k) is bounded from above. Since $\|y_k\|' = 1$, by passing to a further subsequence of (k), which we call again for simplicity (k), we may assume that there exists a non-negative number α and a vector y such that

$$\alpha_k \longrightarrow \alpha \quad and \ \ y_k \longrightarrow y \ .$$

Letting $k \longrightarrow \infty$ in (1.20), we get

$$\langle (y+\alpha r)', r\rangle \;\geq\; (\langle b, r\rangle - \frac{1}{2}\|r\|_2^2)/\langle b, y\rangle \tag{1.28}$$

(this is possible because the map $z \mapsto z'$ is continuous on $\mathbf{R}^m \setminus \{0\}$). We will prove that both possibilities, $\alpha = 0$ and $\alpha > 0$, lead to contradictions.

If $\alpha = 0$, then (1.28) becomes

$$\langle y', r\rangle \, \langle b, y\rangle \;\geq\; \langle b, r\rangle - \frac{1}{2}\|r\|_2^2 \,. \tag{1.29}$$

If one takes the limit as $k \longrightarrow \infty$ in (1.16), one has

$$\|r\|_2^2 = \langle b, r\rangle - \langle b, y\rangle \langle y', r\rangle \,. \tag{1.30}$$

Inequality (1.29) and equation (1.30) together imply

$$\|r\|_2^2 \;\leq\; \frac{1}{2}\|r\|_2^2$$

which implies $r = 0$, leading to a contradiction, as claimed.

If $\alpha > 0$, then by step 6 of the algorithm

$$y_{k+1} \longrightarrow (y + \frac{1}{2}\alpha r)/\|y + \frac{1}{2}\alpha r\|' = \hat{y} \,, \tag{1.31}$$

as $k \longrightarrow \infty$. Both subsequences $(\langle b, y_k\rangle)$ and $(\langle b, y_{k+1}\rangle)$ have the same limit. Let ρ be this limit. Then

$$\langle b, y\rangle = \langle b, \hat{y}\rangle = \rho \,. \tag{1.32}$$

Using equations (1.31) and (1.32) simultaneously gives

$$\langle b, y\rangle \|y + \frac{1}{2}\alpha r\|' = \langle b, y\rangle + \frac{1}{2}\alpha\langle b, r\rangle \,.$$

Rewriting this equation yields

$$\|y + \frac{1}{2}\alpha r\|' = 1 + \frac{1}{2}\alpha\langle b, r\rangle / \langle b, y\rangle \,. \tag{1.33}$$

Allowing $k \longrightarrow \infty$ in the equation (1.12) defining α_k, we obtain

$$\langle b,y\rangle \|y+\alpha r\|' = \langle b,y\rangle + \alpha(\langle b,r\rangle - \frac{1}{2}\|r\|_2^2) .$$

Because the norm $\|.\|$ is smooth, we apply the same argument as in the proof of proposition 1.5 using the strict convexity of the dual norm $\|.\|'$ (equation (1.19)) to conclude that

$$\langle b,y\rangle \|y+\frac{1}{2}\alpha r\|' < \langle b,y\rangle + \frac{1}{2}\alpha\langle b,r\rangle - \frac{1}{4}\alpha\|r\|_2^2 . \tag{1.34}$$

It is clear now that equations (1.33) and (1.34) combined imply

$$0 < -\frac{1}{4}\alpha\|r\|_2^2 ,$$

leading to the desired contradiction. The proof of case 1 is complete.

Case 2. Suppose that step 3 is executed for all but a finite number of iterations. Passing to a subsequence denoted again (k), we may assume

$$y_k \longrightarrow y \qquad \textit{and} \;\; r_k \longrightarrow r ,$$

as $k \longrightarrow \infty$. In view of arriving to a contradiction, assume that $r \neq 0$. By step 3 of the algorithm

$$y_{k+1} = r_k/\|r_k\|' . \tag{1.35}$$

The subsequence (y_{k+1}) converges to

$$\hat{y} = r/\|r\|' .$$

Thus $\hat{y}$, besides y, is a cluster point of the sequence (y_k). Because of the fact that the original numerical sequence $(\langle b,y_k\rangle)$ is convergent (proposition 1.5), one has

$$\langle b,y\rangle = \langle b,\hat{y}\rangle . \tag{1.36}$$

Letting $k \longrightarrow \infty$ in both sides of the inequality defining step 3, we find

$$\langle b,r\rangle \;\geq\; \langle b,y\rangle\|r\|' + \frac{1}{4}\|r\|_2^2 ,$$

i.e,

$$\langle b, \hat{y}\rangle \|r\|' \ \geq \ \langle b, y\rangle \|r\|' + \frac{1}{4}\|r\|_2^2 \ ,$$

(since $\langle b, r\rangle = \langle b, \hat{y}\rangle \|r\|'$) . In view of equation (1.36), this implies $r = 0$, yielding a contradiction. So, in both cases, we have proved that the sequence (r_k) converges to zero, as claimed.

It remains to prove that the sequence (y_k) converges to the solution of problem (P'). Let y be a cluster point of the sequence (y_k). Then, there exists a subsequence, denoted again (y_k) converging to y. From step 1 of the algorithm, we have

$$r_k = b \ - \langle b, y_k\rangle y_k' - Ax_k \ . \tag{1.37}$$

Since $r_k \longrightarrow 0$, the sequence (Ax_k) converges to $b \ - \langle b, y\rangle y'$. From the fact that the cone $\{Ax \mid x \geq 0, x \in \mathbf{R}^n\}$ is closed, we get

$$b \ - \langle b, y\rangle y' = Ax \ , \quad \ \mathit{for\ some} \quad x \geq 0 \ .$$

Now applying theorem 1.3, we see that y solves problem (P'). The same argument applies to any cluster point x of the sequence (x_k). Let $(x_{k'})$ be a subsequence converging to x. By taking a further subsequence, if necessary, we may assume that $y_{k'} \longrightarrow y$. Considering equation (1.37) and letting $k' \longrightarrow \infty$ in (1.37), we obtain $Ax = b \ - \langle b, y\rangle y'$, (since $r_k \longrightarrow 0$). Once more applying theorem 1.3, we conclude that x is a solution of problem (P), as claimed. This completes the proof.

The algorithm studied in this chapter can be generalized as follows. We keep the previous notations. At each iteration cycle of algorithm 1.1, a prototype approximation is given by

$$y_{k+1} = (y_k + \frac{1}{2}\alpha_k r_k)/\|y_k + \frac{1}{2}\alpha_k r_k\|' \ . \tag{1.38}$$

As it was done in [14] a natural extension results from choosing an arbitrary coefficient λ_k belonging to a set to be determined as a substitute to the coefficient $1/2$ in equation (1.38). We set down these changes in the following generalized algorithm.

Algorithm 1.2

Step 0 to *step* 5. Same as in algorithm 1.1

Step 6. Let

$$y_{k+1} = (y_k + \lambda_k \alpha_k r_k)/\|y_k + \lambda_k \alpha_k r_k\|' \ ,$$

where $\lambda_k \in \Lambda$. Λ is a fixed compact subset of the open interval $(\frac{1}{4}, 1)$.

Step 7. Increase k by 1 and return to step 2.

Step 8. x_k is a non-negative $\|.\|$ minimal norm solution of $Ax = b$. The computation is complete.

A carefull inspection shows that most of the results concerning algorithm 1.2 carry over to the present situation. Only a few changes are needed in the proof of the convergence.

Theorem 1.9 *Let Λ be a non-empty compact subset of the open interval $(\frac{1}{4}, 1)$. Assume the norm $\|.\|$ to be smooth and strictly convex. Let the sequence (y_k) be defined as in step 6 of the algorithm by*

$$y_{k+1} = (y_k + \lambda_k \alpha_k r_k)/\|y_k + \lambda_k \alpha_k r_k\|' \ ;$$

where λ_k is arbitrary, $\lambda_k \in \Lambda$. Then

(a) the sequence $(\langle b, y_k \rangle)$ generated by the algorithm 1.3 is strictly increasing and convergent.

(b) the sequence (y_k) converges to the solution of problem (P').

Proof. Recall that the smoothness of the norm $\|.\|$ implies the strict convexity of the dual norm $\|.\|'$. From this observation we obtain

$$\|y_k + \lambda_k \alpha_k r_k\|' \ < \ \lambda_k \|y_k + \alpha_k r_k\|' + (1 - \lambda_k)\|y_k\|'$$

$$= \lambda_k(1 + \alpha_k\mu_k) + (1 - \lambda_k) \quad \textit{(due to step 5)}$$

$$= 1 + \lambda_k\alpha_k(\langle b, r_k\rangle - \frac{1}{2}\|r_k\|_2^2)/\langle b, y_k\rangle \; . \tag{1.39}$$

Rewriting (1.39), we get

$$\begin{aligned}\langle b, y_k\rangle &< (\langle b, y_k + \lambda_k\alpha_k r_k\rangle - \frac{1}{2}\lambda_k\alpha_k\|r_k\|_2^2)/\|y_k + \lambda_k\alpha_k r_k\|' \\ &< \langle b, y_{k+1}\rangle \; ,\end{aligned}$$

since $\lambda_k > \frac{1}{4}$, $\alpha_k > 0$ and $r_k \neq 0$. Thus, the sequence $(\langle b, y_k\rangle)$ is strictly increasing. The rest of the proof of (a) is similar to the proof of proposition 1.5.

To prove the convergence of (y_k) claimed in (b), we show that

$$\lim r_k = 0 \; . \tag{1.40}$$

As in theorem 1.8, we assume $r_k \longrightarrow r \neq 0$ and show that this leads to a contradiction. The difference is that we have to pass to an additional subsequence of (λ_k), that we shall denote (λ_k), such that

$$\lambda_k \longrightarrow \lambda \in \Lambda \; .$$

The two cases $\alpha = 0$ and $\alpha > 0$ are studied. The case $\alpha = 0$ carries over exactly as in the proof of theorem 1.8, while in the case $\alpha > 0$ by letting $k \longrightarrow \infty$, we have

$$y_{k+1} \longrightarrow (y + \lambda\alpha r)/\|y + \lambda\alpha r\|' = \hat{y} \; . \tag{1.41}$$

Since the subsequences $(\langle b, y_k\rangle)$ and $(\langle b, y_{k+1}\rangle)$ have the same limit, equation (1.41) shows that

$$\langle b, y\rangle\|y + \lambda\alpha r\|' = \langle b, y\rangle + \lambda\alpha\langle b, r\rangle \; . \tag{1.42}$$

Now, let $k \longrightarrow \infty$ in (1.39). We get

$$\langle b, y\rangle\|y + \lambda\alpha r\|' < \langle b, y\rangle + \lambda\alpha\langle b, r\rangle - \frac{1}{2}\lambda\alpha\|r\|_2^2 \; . \tag{1.43}$$

Combinig eq.(1.42) and inequality (1.43), leads to

$$0 < -\frac{1}{2}\lambda\alpha\|r\|_2^2 \ .$$

which contradicts the assumption that $r \neq 0$. the rest of the proof is completed exactly as in theorem 1.8.

The reader will note that this theorem was motivated by a similar work in [14].

Chapter 2

An Algorithm For Minimal Norm Solution Of A Linear System

In this chapter, we will be considering a system of real linear equations

$$Ax = b \; . \tag{2.1}$$

where A is an $m \times n$ real matrix, b an m-real vector, and x an n-real vector. Under the assumption that the system (2.1) has a non-negative solution, we shall be concerned with the following problem: amongst all the non-negative solutions of (2.1), select that solution which has the least norm $\|.\|$, when $\mathbf{R}^n$ is equipped with a strictly convex norm $\|.\|$ (e.g, the l^p-norm, $1 < p < \infty$) . This problem will be referred to as problem (P)

$$(P) \qquad min\{\|x\| \mid Ax = b \, , x \in \mathbf{R}^n \, , x \geq 0\} \; .$$

We assume that $b \neq 0$, for otherwise the problem is trivial. The notations and definitions are those of chapter 1.

We present an algorithm for computing the solution of (P). Its feasibility and the convergence will then be studied . All the steps in this algorithm will be shown to be feasible. Its global convergence will then be proved.

To solve the given problem, a dual problem, to be denoted by (P'), will be associated with (P). An outline of the correspondence between (P) and (P') will be given. The main application of this work is the l^p-norm case. Namely, Find $x \in \mathbf{R}^n$, that solves the l^p-problem

$$minimize \ \{\|x\|_p \mid Ax = b \ , \ x \in \mathbf{R}^n \ , \ x \geq 0\} \ . \tag{2.2}$$

A worthwhile remark to make here is that the objective function in problem (2.2) need not be twice differentiable in $\mathbf{R}^n \setminus \{0\}$. This is particulary the case when $1 < p < 2$.

For $p = 2$, problem (2.2) was studied in Lawson and Hanson [5]. They refer to (2.2) as the Least Distance Programming (LDP) problem. A finite algorithm for solving LDP was given in [5]. We will use this algorithm as follows. At each iteration cycle of our algorithm, the LDP problem

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

where (a_k) is a sequence defined by the algorithm, is solved using the LDP algorithm of Lawson and Hanson. Briefly, we recall some definitions and notations from chapter 1. A norm $\|.\|$ is said to be *strictly convex* if the unit sphere contains no line segment on its surface. In other words

$$\|x\| = \|y\| = \|(x + y)/2\| = 1 \implies x = y \ .$$

The norm is *smooth* if through each point of the unit sphere in $\mathbf{R}^n$, there passes exactly one hyperplane supporting the closed unit sphere. If the norm $\|.\|$ is strictly convex (resp. smooth), then the $\|.\|$ (resp. $\|.\|'$) dual vector is unique (see chapter 1 for the definition of the dual norm and dual vectors). Moreover, the correspondence $x \mapsto x'$ (resp. $x \mapsto x^*$) from $\mathbf{R}^n$ into the $\|.\|$ ($\|.\|'$)-unit sphere is odd, continuous and positively homogenuous of degree zero on $\mathbf{R}^n \setminus \{0\}$. In the important case of the l^p-norm, defined as usual by

$$\|x\|_p = (\sum |x_i|^p)^{1/p} \ ,$$

with $1 < p < \infty$, we have $\|.\|_p' = \|.\|_q$, where $p + q = pq$. We close this review by recalling that for any smooth, strictly convex norm $\|.\|$

$$v'^{*} = v/\|v\|' \ \ and \ \ v^{*'} = v/\|v\| \ ,$$

if $v \neq 0$. Let

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

Given problem (P), we associate a dual problem ([6])

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

where A^T is the transpose of the matrix A. The relation between problem (P) and problem (P') is investigated in the next two results.

Lemma 2.1 *(weak duality) Suppose K non-empty. Let the norm $\|.\|$ on $\mathbf{R}^n$ be arbitrary, then*

$$value \ of \ (P') \ \leq \ value \ of \ (P) \ . \tag{2.3}$$

Proof. Let $x \in K$, and let ξ and y be as defined in problem (P'). Then

$$\begin{aligned} \langle b, y\rangle &= \langle Ax, y\rangle \\ &= \langle x, A^T y\rangle \\ &\leq \langle x, A^T y + \xi\rangle \qquad (since \ x \geq 0 \ , \ \xi \geq 0) \\ &\leq \|x\|.\|A^T y + \xi\|' \\ &\leq \|x\| \ , \end{aligned}$$

which implies (2.3), as claimed.

Theorem 2.2 *(Duality [6])If $\|.\|$ is an arbitrary norm on $\mathbf{R}^n$, and K is non-empty, then the problems (P) and (P') have the same value.*

The next theorem, due to Sreedharan and Nikolopoulos, states a useful characterization of the solution of (P).

Theorem 2.3 *([6]) Let the norm $\|.\|$ on $\mathbf{R}^n$ be strictly convex. Suppose K non-empty. The following are equivalent*

(i) $\bar{x}$ is a non-negative minimal norm solution of problem (P).

(ii) $A\bar{x} = b$, $\bar{x} \geq 0$, and there exist $\xi \in \mathbf{R}^n$, $\xi \geq 0$ and $y \in \mathbf{R}^m$ such that

$$\langle b, y\rangle > 0, \ \|A^T y + \xi\|' = 1 \ , \tag{2.4}$$

and

$$\bar{x} = \langle b, y\rangle (A^T y + \xi)' \ . \tag{2.5}$$

Furthermore,

$$\langle \xi, (A^T y + \xi)' \rangle = 0 \ . \tag{2.6}$$

If in addition the norm $\|.\|$ is smooth, then

$$\bar{x}^* = \xi + A^T y \ , \text{and } \langle \xi, \bar{x}\rangle \ . \tag{2.7}$$

Finally, (y, ξ) solves (P'), and

$$\langle b, y\rangle = \langle \bar{x}^*, x\rangle \ .$$

A characterization of the solution of problem (P) in the l^2-norm case follows immediately from theorem 2.3. This is an important ingredient in the subsequent developement of this chapter. We record it as

Corollary 2.4 *If $\|.\| = \|.\|_2$, then $\bar{x}$ is the solution of problem (P) iff $\bar{x} \in K$,*

$$\bar{x} = A^T y + \xi \ , \tag{2.8}$$

and

$$\langle \xi, \bar{x} \rangle = 0 \; , \tag{2.9}$$

for some $y \in \mathbf{R}^m$, $\xi \in \mathbf{R}^n$, $\xi \geq 0$.

It should be noted that when $\|.\| = \|.\|_2$, then $\bar{x}^* = \bar{x}/\|\bar{x}\|_2$, so that (2.8) and (2.9) are simply an easy consequences of equations (2.6) and (2.7).

We are now prepared to state our algorithm for computing the solution of problem (P). We assume that the system of linear equations $Ax = b$, $x \geq 0$, is feasible and that the norm $\|.\|$ is strictly convex and smooth.

Algorithm 2.1

Step 0. Find x_0 the solution of $Ax = b$, $x \geq 0$, $\|x\|_2(\min)$. Let

$$g_0 = x_0/\|x_0\|' \; , \beta_0 = \langle g_0, x_0 \rangle \;\; and \;\; k = 0 \; .$$

Step 1. Set $a_k = \beta_k g_k'$. Find x_{k+1} solution of

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

Let $u_k = x_{k+1} - a_k$.

Step 2. If $u_k = 0$, stop. x_{k+1} is the solution of (P); else continue.

Step 3. Set $\gamma_k = \langle u_k, x_k \rangle$.

Step 4. If

$$\gamma_k \;\geq\; \beta_k \|u_k\|' + \frac{1}{4}\|u_k\|_2^2 \; ,$$

let

$$g_{k+1} = u_k/\|u_k\|' \; and \; \beta_{k+1} = \gamma_k/\|u_k\|' \; .$$

and GO TO *step* 8; else continue.

Step 5. Let

$$\mu_k = (\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k \ .$$

Step 6. Find $\alpha_k > 0$ such that

$$\|g_k + \alpha_k u_k\|' = 1 + \alpha_k \mu_k \ .$$

Step 7. Let

$$g_{k+1} = (g_k + \frac{1}{2}\alpha_k u_k)/\|g_k + \frac{1}{2}\alpha_k u_k\|' \ ,$$

and

$$\beta_{k+1} = (\beta_k + \frac{1}{2}\alpha_k \gamma_k)/\|g_k + \frac{1}{2}\alpha_k u_k\|' \ .$$

Step 8. Increase k by 1 and RETURN to *step* 1.

Later in this chapter the stopping rule of step 2 will be used as follows. We will show that the constructed sequence (u_k) converges to zero. Then, because of this convergence, it will be proven that the algorithm converges to the solution of problem (P). However, for implementation purpose, we shall use the more realistic stopping rule, similar to the one used in chapter 1, called the relative duality gap criterion. In other words, we replace the condition $u_k = 0$ of step 2 by the condition $(\|x_k\| - \beta_k)/\|x_k\| \leq \eta$, where $\eta > 0$ is a stopping rule parameter. We record this in

Algorithm 2.2.

Step 0. Find x_0, g_0 and β_0, as in step 0 of algorithm 2.1; let $\eta > 0$ be a stopping parameter.

Step 1. Same as step 1 in algorithm 2.1.

Step 2. If

$$(\|x_{k+1}\| - \beta_k)/\|x_{k+1}\| \leq \eta \ ,$$

stop. x_{k+1} is the solution of (P); else proceed.

Step 3 to 8. Same as in algorithm 2.1.

Our main interest in the upcoming sections of this chapter will be focused on proving that the various steps of the algorithm are valid and that the important step 6 defining $\alpha_k > 0$ is anserwed affirmatively and, finally, to prove that the algorithm leads effectively to the solution of problem (P).

Lemma 2.5 *([6]). Let x_0, g_0, β_0 be defined as in step 0 of the algorithm. Then, there exist $\xi_0 \in \mathbf{R}^n$, $\xi_0 \geq 0$, $y_0 \in \mathbf{R}^m$ such that*

$$g_0 = \xi_0 + A^T y_0 \ , \|g_0\|' = 1 \ , and\ \beta_0 = \langle b, y_0 \rangle > 0 \ . \tag{2.10}$$

Proof. By corollary 2.4, there exist $\xi \in \mathbf{R}^n$, $\xi \geq 0$, $z \in \mathbf{R}^m$ such that

$$x_0 = \xi + A^T z \ , \ and\ \langle \xi, x_0 \rangle = 0 \ .$$

But $g_0 = x_0/\|x_0\|'$, thus $g_0 = \xi_0 + A^T y_0$, where $\xi_0 = \xi/\|x_0\|'$, and $y_0 = z/\|x_0\|'$. To prove the last statement of (2.10), we note that

$$\begin{aligned} \beta_0 &= \langle g_0, x_0 \rangle \\ &= \|x_0\|_2^2/\|x_0\|' > 0 \ , \end{aligned}$$

and

$$\begin{aligned} \langle g_0, x_0 \rangle &= \langle \xi_0 + A^T y_0, x_0 \rangle \\ &= \langle y_0, A x_0 \rangle \qquad (Due\ to\ (\ 2.9)), \\ &= \langle y_0, b \rangle \ , \end{aligned}$$

which completes the proof.

Proposition 2.6 *(a) Let $u_k \neq 0$ be as defined in the algorithm. Then*

$$\forall \alpha > 0 \ , \ g_k + \alpha u_k \ \neq 0 \ . \tag{2.11}$$

(b) Let $\alpha_k > 0$ be as defined by step 6 of the algorithm, then there exist $\xi_{k+1} \in$ $\mathbf{R}^n$, $\xi_{k+1} \geq 0$, $y_{k+1} \in \mathbf{R}^m$ such that

$$g_{k+1} = \xi_{k+1} + A^T y_{k+1} \ , \ \|g_{k+1}\|' = 1 \ , \textit{and} \ \beta_{k+1} \ = \ \langle y_{k+1}, b \rangle \ . \tag{2.12}$$

Proof. To prove (a) we proceed by induction on k. Let $k = 0$ and suppose (2.11) does not hold, i.e, $\exists \alpha > 0$ such that $g_0 = -\alpha u_0$. Then

$$\begin{aligned}
\|u_0\|_2^2 &= \langle u_0, x_1 - a_0 \rangle \\
&= -\alpha^{-1} \langle g_0, x_1 - a_0 \rangle \\
&= -\alpha^{-1} (\langle g_0, x_1 \rangle - \langle g_0, a_0 \rangle) \\
&= -\alpha^{-1} (\langle x_1, \xi_0 + A^T y_0 \rangle - \beta_0) \ , (\textit{because of} \ (\ 2.10) \ \textit{and step} \ 1) \\
&= -\alpha^{-1} (\langle A x_1, y_0 \rangle + \langle x_1, \xi_0 \rangle - \beta_0) \\
&= -\alpha^{-1} (\beta_0 + \langle x_1, \xi_0 \rangle - \beta_0) \ , \ (\textit{since} \ b = A x_1) \ , \\
&= -\alpha^{-1} (\langle x_1, \xi_0 \rangle) \ .
\end{aligned}$$

This yields a contradiction since $x_1 \geq 0, \xi_0 \geq 0$, and $u_0 \neq 0$, which implies (2.11) for $k = 0$, as desired.

To prove (b), by definition of x_1 in step 1 of the algorithm and the corollary 2.4, there exist $\xi \in \mathbf{R}^n, \xi \geq 0$, $z \in \mathbf{R}^m$ such that

$$\begin{aligned}
u_0 &= x_1 - a_0 \\
&= \xi + A^T z \ ,
\end{aligned} \tag{2.13}$$

and

$$\langle z, b \rangle = \langle u_0, x_1 \rangle = \gamma_0 \ .$$

Suppose α_0 is determined by step 6 of the algorithm. Then, because of (2.11),

$$g_0 + \frac{1}{2} \alpha_0 u_0$$

is nonzero. Let

$$y_1 = (y_0 + \frac{1}{2}\alpha_0 z)/\|g_0 + \frac{1}{2}\alpha_0 u_0\|' ,$$

and

$$\xi_1 = (\xi_0 + \frac{1}{2}\alpha_0 \xi)/\|g_0 + \frac{1}{2}\alpha_0 u_0\|' .$$

From (2.10) and (2.13), we obtain

$$\begin{aligned} A^T y_1 + \xi_1 &= (\xi_0 + A^T y_0 + \frac{1}{2}\alpha_0(\xi + A^T z))/\|g_0 + \frac{1}{2}\alpha_0 u_0\|' \\ &= (g_0 + \alpha_0 u_0)/\|g_0 + \alpha_0 u_0\|' \\ &= g_1 . \end{aligned}$$

Now,

$$\begin{aligned} \langle y_1, b\rangle &= (\langle y_0, b\rangle + \frac{1}{2}\alpha_0\langle z, b\rangle)/\|g_0 + \frac{1}{2}\alpha_0 u_0\|' \\ &= (\beta_0 + \frac{1}{2}\alpha_0\gamma_0)/\|g_0 + \frac{1}{2}\alpha_0 u_0\|' \\ &= \beta_1 . \end{aligned}$$

This completes the proof of (2.12), for $k = 0$. The same argument applies for any integer k if we assume the proposition to be true for $k - 1$.

Remark. It will be shown later that the sequence (β_k) generated by the algorithm is strictly increasing. This combined with (2.10) imply that $\beta_k > 0$, for all k. We conclude that step 5 is properly formulated.

Proposition 2.7 *Suppose the algorithm is at the stage of executing step 5. Let μ_k be defined by*

$$\mu_k = (\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k .$$

Then, there exists $\alpha_k > 0$ such that

$$\|g_k + \alpha_k u_k\|' = 1 + \alpha_k \mu_k . \tag{2.14}$$

Proof. Let the function $f : \mathbf{R} \longrightarrow \mathbf{R}$ be defined by

$$f(\lambda) = \|g_k + \lambda u_k\|' ,$$

and let

$$l(\lambda) = 1 + \lambda \mu_k .$$

Then, f is a strictly convex function. Moreover, $f(0) = \|g_k\|' = 1 = l(0)$. Also

$$f'(0) = \langle g_k', u_k \rangle ,$$

(see chapter 1). We have

$$\begin{aligned}
\lambda_k/\beta_k &= \langle u_k, x_{k+1}\rangle/\beta_k , \quad (\textit{by step}3) \\
&= \langle u_k, u_k + a_k\rangle/\beta_k , \ (\textit{because of step } 1) \\
&= \|u_k\|_2^2/\beta_k + \langle u_k, g_k'\rangle ,
\end{aligned}$$

(since $a_k = \beta_k g_k'$). If the algorithm is at the stage of executing step 5, then $u_k \neq 0$. So,

$$\begin{aligned}
f'(0) - l'(0) &= \langle g_k', u_k\rangle - (\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k \\
&= \langle g_k', u_k\rangle - \|u_k\|_2^2/\beta_k - \langle u_k, g_k'\rangle + \frac{1}{2}\|u_k\|_2^2/\beta_k \\
&= -\frac{1}{2}\|u_k\|_2^2/\beta_k \\
&< 0 .
\end{aligned}$$

Hence, there must exist $\lambda > 0$ such that $f(\lambda) - l(\lambda) < 0$. Now, $\alpha_k > 0$ in step 5 is sought only if step 4 of the algorithm is answered negatively. This means that γ_k must satisfy

$$\gamma_k < \beta_k\|u_k\|' + \frac{1}{4}\|u_k\|_2^2 . \tag{2.15}$$

Using this inequality, we have

$$
\begin{aligned}
f(\lambda) - l(\lambda) &= \|g_k + \lambda u_k\|' - 1 - \lambda(\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k \\
&\geq \lambda\|u_k\|' - \|g_k\|' - 1 - \lambda(\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k \\
&= \lambda(\beta_k\|u_k\|' - \gamma_k + \frac{1}{2}\|u_k\|_2^2)/\beta_k \ - 2 \\
&> \lambda(-\frac{1}{4}\|u_k\|_2^2 + \frac{1}{2}\|u_k\|_2^2)/\beta_k \ - 2 \ .
\end{aligned}
$$

The last step is due to (2.15). From this we get

$$
f(\lambda) - l(\lambda) \ > \ \frac{\lambda}{4}\|u_k\|_2^2/\beta_k \ - 2 \ \longrightarrow \infty
$$

as $\lambda \longrightarrow \infty$.

Therefore, $f(\lambda) - l(\lambda) > 0$, for some $\lambda > 0$. Since $f - l$ is continuous, there must exists $\alpha_k > 0$, solution of $f(\lambda) - l(\lambda) = 0$. Thus, equation (2.14) holds. The proof is complete.

Proposition 2.8 *Let the norm* $\|.\|$ *be smooth. Then*

(a) The sequence (β_k) *generated by the algorithm is strictly increasing.*

(b) Either the sequence (β_k) *is finite; or it is a convergent infinite sequence.*

Proof. Two cases are to be distinguished. The first case is straightforward. Assume step 4 of the algorithm is executed. This occurs if

$$
\gamma_k \geq \beta_k\|u_k\|' + \frac{1}{4}\|u_k\|_2^2 \ .
$$

So, from the definition of β_{k+1}, we obtain

$$
\begin{aligned}
\beta_{k+1} &= \gamma_k/\|u_k\|' \\
&\geq \beta_k + \frac{1}{4}\|u_k\|_2^2/\|u_k\|' \\
&> \beta_k \ ,
\end{aligned}
$$

(since $u_k \neq 0$). Now, let us assume that step 4 of the algorithm is answered negatively. Then β_{k+1} is defined by step 7. Let $\alpha_k > 0$ be as defined in step 6. Recall that we are assuming the norm $\|.\|$ to be smooth. So, the dual norm $\|.\|'$ is strictly convex. It follows that

$$\begin{aligned}
\|g_k + \frac{1}{2}\alpha_k u_k\|' &< \frac{1}{2}\|g_k + \alpha_k u_k\|' + \frac{1}{2}\|g_k\|' \\
&= 1 + \frac{1}{2}\alpha_k(\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k \, .
\end{aligned}$$

Consequently, we get

$$\begin{aligned}
\beta_k &< ((\beta_k + \frac{1}{2}\alpha_k\gamma_k) - \frac{1}{4}\alpha_k\|u_k\|_2^2)/\|g_k + \frac{1}{2}\alpha_k u_k\|' \\
&< (\beta_k + \frac{1}{2}\alpha_k\gamma_k)/\|g_k + \frac{1}{2}\alpha_k u_k\|' \\
&= \beta_{k+1} \, .
\end{aligned}$$

This proves that, in all cases, the squence (β_k) is strictly increasing, as claimed.

To prove (b), supppose that the sequence (β_k) is infinite. To show its convergence, we need only prove that it is a bounded sequence. This follows immediatly from the weak duality lemma,

$$\begin{aligned}
\beta_k &= \langle y_k, b \rangle \\
&\leq \|x\| \, , \qquad (2.16)
\end{aligned}$$

for any fixed feasible solution $x \in \mathbf{R}^n$. Hence, (β_k) is bounded from above. The proof is complete.

After these preliminaries concerning the feasibility of Algorithm 2.1, we begin the study of its convergence. The next two results will be needed in the convergence theorem.

Lemma 2.9 *Let $\alpha_k > 0$ be as defined in step 6 of the algorithm. Then*

$$\beta_k\langle (g_k + \alpha_k u_k)', u_k \rangle + \frac{1}{2}\|u_k\|_2^2 \geq \gamma_k \, . \qquad (2.17)$$

Proof. First, we recall that by the definition of the dual vectors and the construction of g_k, we have

$$\|(g_k + \alpha_k u_k)'\| = 1 \quad and \quad \|g_k\|' = 1 \ .$$

This implies

$$\langle (g_k + \alpha_k u_k)', g_k \rangle \ \leq \ \|(g_k + \alpha_k u_k)'\| . \|g_k\|' = 1 \ .$$

Due to the definition of the dual norm, we also have the equation

$$\|g_k + \alpha_k u_k\|' = \langle (g_k + \alpha_k u_k)', g_k \rangle + \alpha_k \langle (g_k + \alpha_k u_k)', u_k \rangle \ .$$

This, combined with the equation (2.14), defining α_k, yield

$$\begin{aligned} 1 + \alpha_k(\gamma_k - \frac{1}{2}\|u\|_2^2)/\beta_k &= \|g_k + \alpha_k u_k\|' \\ &\leq 1 + \alpha_k \langle (g_k + \alpha_k u_k)', u_k \rangle \ , \end{aligned}$$

which, since $\alpha_k > 0$, implies (2.17), as desired.

Theorem 2.10 *Let a_k , x_k and u_k be as defined in the algorithm. Then, the sequences (a_k) , (x_k) and (u_k) are bounded.*

Proof. If the algorithm terminates in a finite number of iterations, we are done. Consider the case the sequences are infinite. Let $d > 0$ be the value of the minimization problem (P). Then as mentioned above,

$$\begin{aligned} \beta_k &= \langle b, y_k \rangle \\ &\leq d \ , \quad for \ all \ k \ . \end{aligned} \tag{2.18}$$

Now, from step 1 of the algorithm, we see that

$$\begin{aligned} \|a_k\|_2 &= \beta_k \|g_k'\|_2 \\ &\leq M\beta_k \|g_k'\| \\ &= M\beta_k \quad (since \ \|v'\| = 1 \ , for \ all \ v \neq 0) \\ &\leq Md \ , \end{aligned}$$

where $M > 0$ is such that $\|v\|_2 \leq M\|v\|$ for all $v \in \mathbf{R}^n$. Hence (a_k) is a bounded sequence.

To prove that the sequence (x_k) is bounded, let $\tilde{x}$ be any fixed feasible solution of problem (P). Because x_{k+1} is the minimizer of the problem

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

(this minimization is done in step 1 of the algorithm at each iteration cycle), we have

$$\begin{aligned} \|x_{k+1}\| &\leq \|x_{k+1} - a_k\|_2 + \|a_k\|_2 \\ &\leq \|\tilde{x} - a_k\|_2 + \|a_k\|_2 \\ &\leq \|\tilde{x}\|_2 + 2\|a_k\|_2 \\ &\leq \|\tilde{x}\|_2 + 2Md \ . \end{aligned}$$

Thus, (x_k) is a bounded sequence. From this it follows clearly that the sequence (u_k), $u_k = x_{k+1} - a_k$, is also bounded. This completes the proof.

We can now put everything together to prove that the algorithm 2.2 converges to the solution of problem (P).

Proposition 2.11 *([6]). Let (u_k) be the sequence generated by the algorithm. If $u_k = 0$, for some k, then x_{k+1} is the solution of problem (P).*

Proof. $u_k = 0$ implies $x_{k+1} = a_k = \beta_k g_k'$ (because of step 1 of the algorithm). By virtue of proposition 2.6,

$$\beta_k = \langle b, y_k \rangle \quad and \quad g_k = \xi_k + A^T y_k \ .$$

This gives us the relation

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

Moreover,

$$\|\xi_k + A^T y_k\|' = 1 \quad \textit{and} \quad \langle b, y_k \rangle > 0 \ .$$

Using theorem 2.3, it is obvious now that x_{k+1} solves (P).

Theorem 2.12 *(a) If the algorithm terminates after a finite number k of iterations, then x_{k+1} is the solution of problem (P).*

(b) If the algorithm generates an infinite sequence (x_k), then it converges to the solution of (P).

Proof. Algorithm 2.2 terminates in a finite number k of iterations iff $u_k = 0$. In this case, by proposition 2.11, the algorithm computes the unique solution x_{k+1} to the problem (P). This proves (a).

To prove (b), assume that the algorithm is executed for an infinite number of indices (k). By theorem 2.10, the sequence (γ_k) defined by

$$\gamma_k = \langle u_k, x_{k+1} \rangle \ ,$$

where (x_k) and (u_k) are generated by the algorithm 2.1, is a bounded sequence. Hence, by passing to a subsequence if necessary, we may assume that $\gamma_k \longrightarrow \gamma$. Since $0 < \beta_k \leq d$ (due to (2.18)) and $\|g_k\|' = 1$, let us pass to a further subsequence, denoted again by (k), such that

$$\beta_k \longrightarrow \beta \quad \textit{and} \quad g_k \longrightarrow g \ .$$

Our first goal is to establish that

$$\lim u_k = 0 \ . \tag{2.19}$$

Suppose the claim were false. Then, once more by theorem 2.10, there exists a subsequence, denoted (k), such that

$$u_k \longrightarrow u \neq 0 \ . \tag{2.20}$$

Case 1. Step 6 is executed for an infinite number of indices (k). We claim that the corresponding sequence of positive numbers (α_k) is bounded from above. Because of (2.20), we can pick a subsequence, denoted once more by (k), such that

$$\|u_k\|_2^2 \geq \delta\ ,\quad \forall k\ ,\quad for \quad some \quad \delta > 0\ . \tag{2.21}$$

Now, using (2.14) it follows

$$\begin{aligned} \alpha_k\|u_k\|' - 1 &\leq \|g_k + \alpha_k u_k\|' \\ &= 1 + \alpha_k(\gamma_k - \frac{1}{2}\|u_k\|_2^2)/\beta_k\ . \end{aligned}$$

This shows that

$$\alpha_k(\beta_k\|u_k\|' - \gamma_k + \frac{1}{2}\|u_k\|_2^2) \leq 2\beta_k\ . \tag{2.22}$$

Recall that step 6 is executed only in case

$$\gamma_k < \beta_k\|u_k\|' + \frac{1}{4}\|u_k\|_2^2\ . \tag{2.23}$$

Combining (2.22) and (2.23) we get

$$\frac{1}{4}\alpha_k\|u_k\|_2^2 < 2\beta_k\ .$$

Together with (2.21) and (2.18), this shows that

$$0 < \alpha_k \ < 8d/\delta\ .$$

Thus, the sequence (α_k) is bounded, as claimed. Passing to a further subsequence, if necessary, we may assume that there exists $\alpha\ \geq\ 0$ such that

$$\alpha_k \longrightarrow \alpha \quad as \quad k \longrightarrow \infty\ .$$

If we let $k \longrightarrow \infty$ in (2.17), then by continuity of the map $z \mapsto z'$ on $\mathbf{R}^n \setminus \{0\}$, it follows that

$$\beta\langle(g + \alpha u)', u\rangle + \frac{1}{2}\|u\|_2^2\ \geq\ \gamma\ . \tag{2.24}$$

We distinguish two possibilities, $\alpha = 0$ or $\alpha > 0$. We prove that each of these possibilities leads to a contradiction. We then conclude that $\lim u_k = 0$.

If $\alpha = 0$, then (2.24) becomes

$$\beta\langle g', u\rangle + \frac{1}{2}\|u\|_2^2 \geq \gamma \ . \tag{2.25}$$

From the definition of u_k in step 1 of the algorithm, we obtain

$$\begin{aligned} \|u_k\|_2^2 &= \langle x_{k+1} - a_k, u_k\rangle \\ &= \langle x_{k+1}, u_k\rangle - \langle a_k, u_k\rangle \\ &= \gamma_k - \beta_k\langle g_k', u_k\rangle \ . \end{aligned} \tag{2.26}$$

Passing to the limit on both sides of (2.26) leads to

$$\|u\|_2^2 = \gamma - \beta\langle g', u\rangle \ . \tag{2.27}$$

Inequality (2.25) and equation (2.27) force u to satisfy

$$\frac{1}{2}\|u\|_2^2 \leq 0 \ ,$$

which is a contradiction to (2.20). So $u = 0$, as claimed.

Suppose now that $\alpha > 0$. Using step 7 of the algorithm and allowing $k \longrightarrow \infty$, we get

$$\beta_{k+1} \longrightarrow (\beta + \frac{1}{2}\alpha\gamma)/\|g + \frac{1}{2}\alpha u\|' = \hat{\beta} \ .$$

Note that $\beta_k = \langle y_k, b\rangle$ and $\beta_{k+1} = \langle y_{k+1}, b\rangle$ both have the same limit. So $\beta = \hat{\beta}$. This yields the equation

$$\beta\|g + \frac{1}{2}\alpha u\|' = \beta + \frac{1}{2}\alpha\gamma \ . \tag{2.28}$$

Once more allowing $k \longrightarrow \infty$ in the equation (2.14) defining α_k, gives

$$\beta\|g + \alpha u\|' = \beta + \alpha(\gamma - \frac{1}{2}\|u\|_2^2) \ . \tag{2.29}$$

Applying the strict convexity of the dual norm $\|.\|'$ implies

$$\begin{aligned}\beta\|g + \frac{1}{2}\alpha u\|' &< \beta(\frac{1}{2}\|g\|' + \frac{1}{2}\|g + \alpha u\|') \\ &= \beta(\frac{1}{2} + \frac{1}{2}(1 + \alpha(\gamma - \frac{1}{2}\|u\|_2^2)/\beta)) \quad (\textit{because of } (\ 2.29) \\ &= \beta + \frac{1}{2}\alpha(\gamma - \frac{1}{2}\|u\|_2^2)\ . \end{aligned} \tag{2.30}$$

Inserting this in (2.28) shows that

$$\beta + \frac{1}{2}\alpha\gamma < \beta + \frac{1}{2}\alpha(\gamma - \frac{1}{2}\|u\|_2^2)\ ,$$

which implies

$$\frac{1}{4}\|u\|_2^2 < 0$$

leading to the sought contradiction with (2.20). So $u = 0$ and the whole sequence (u_k) converges to zero, as desired.

Case 2. Assume that the condition in step 4 of the algorithm is satisfied for all but a finite number of indices (k). As in the first case we proceed by contradiction. So, let us suppose that (u_k) does not converge to zero. Then there exist a subsequence which we denote again by (k) and a nonzero vector u such that

$$\lim u_k = u \neq 0\ . \tag{2.31}$$

Since step 4 is answered affirmatively, we have

$$\gamma_k \geq \beta_k\|u_k\|' + \frac{1}{4}\|u_k\|_2^2\ , \tag{2.32}$$

for all k. Letting $k \longrightarrow \infty$ in the above inequality implies

$$\gamma \geq \beta\|u\|' + \frac{1}{4}\|u\|_2^2\ . \tag{2.33}$$

By step 4,

$$\beta_{k+1} = \gamma_k/\|u_k\|'\ .$$

As $k \longrightarrow \infty$, this converges to

$$\beta = \gamma / \|u\|' . \tag{2.34}$$

Combining inequality (2.33) and equation (2.34), we obtain

$$\frac{1}{4}\|u\|_2^2 \leq 0 ,$$

a contradiction with (2.31). We have thus proved that in all cases, the sequence (u_k) generated by the algorithm converges to zero, establishing (2.19), as claimed.

We now prove that the algorithm converges to the unique solution of problem (P). Let x^* be any cluster point of the sequence (x_k) and let $(x_{k'})$ be a subsequence converging to x^*. Writing the relation in step 1 of the algorithm for all k', we have

$$x_{k'+1} = u_{k'} + \beta_{k'} g'_{k'} , \tag{2.35}$$

where

$$g_{k'} = A^T y_{k'} + \xi_{k'} \quad , \quad \|g_{k'}\|' = 1 ,$$

and

$$\beta_{k'} \; = \; \langle y_{k'}, b \rangle \; , \; \; \forall \; \; k' \; ,$$

(by proposition 2.6). Since $\|g_{k'}\|' = 1$, by passing to a further subsequence that we denote (k'), we have

$$g_{k'} \longrightarrow g \quad , \quad \|g\|' = 1 \; .$$

It has been proven that $\lim u_{k'} \; = \; 0$, so if we let $k' \longrightarrow \infty$ in (2.35) we get

$$x^* = \beta g' \quad , \quad \|g\|' = 1 \; . \tag{2.36}$$

Clearly, x^* is feasible since $K = \{ \; x \in \mathbf{R}^n \mid x \geq 0, \; Ax = b \; \}$ is closed. In other words, $Ax^* = b$ *and* $x^* \geq 0$.

We have seen earlier in the proof that $\beta = \lim\langle y_{k'}, b\rangle$, where $y_{k'} \in \mathbf{R}^m$, $\xi \geq 0$ *and* $\|A^T y_{k'} + \xi_{k'}\|' = 1$. By the weak duality lemma and equation (2.36) we have

$$\begin{aligned} \langle y_{k'}, b\rangle &\leq \|x^*\| \\ &= \beta. \end{aligned} \tag{2.37}$$

Letting $k' \longrightarrow \infty$ in (2.37) yields equality. This shows that every cluster point of the seqeunce (x_k) is a solution of problem (P). Due to the uniqueness of the solution, we conclude that (x_k) converges to the unique solution of (P). This completes the proof of the theorem.

Chapter 3

On The Minimum Norm Solution Of A System Of Linear Inequalities

Lawson and Hanson proposed in their book ([5],chapter 23) an algorithm for solving the so called LDP problem (Least Distance Programming). The algorithm computes the vector of minimal norm solution of a system of linear inequalities. The norm used by the two authors for the objective function was the l^2-norm and the problem was

$$Minimize\ \ \|x\|_2\ \ subject\ to\ \ Gx \geq h\ .$$

The formulation of the solution to the problem was based on the Kuhn-Tucker optimality conditions. When the norm $\|.\|_2$ is replaced by the l^p-norm, $1 < p < \infty$, in the objective function, the problem is no longer linear. A different characterization of the solution has to be considered.

We shall propose a generalization of the LDP algorithm, given by Lawson and Hanson, to the l^p-case, for any p, $1 < p < \infty$. Keeping in mind the difference of the two problems, we will follow closely, whenever possible the approach in [5]. This is made possible, simpler and elegant because of the availability of the least distance algorithm presented in chapter 2 of this work.

The problem considered, and refered to as problem (P) is

$$(P) \qquad \|x\| \; (min) \; , \quad Ax \geq b \; .$$

Here A an $m \times n$-real matrix, b an m-real vector, x an n-real vector. The norm $\|.\|$ is assumed to be strictly convex and smooth. To solve the given problem (P) and with an eye toward the characterization of the solution, a dual problem, referred to as (P'), will be introduced. We will investigate the relation between (P) and (P'). A related least distance problem, similar to the one studied in Chapter 2, will then be used to solve (P). The main contribution of this work resides in giving an explicit formula to compute the vector of minimal $\|.\|_p$-norm solution of (P).

We introduce briefly the notion of the dual vector for the norm $\|.\|$ on $\mathbf{R}^n$. Let $\|.\|'$ denote the dual norm. For any vector $v \in \mathbf{R}^n$, $v \neq 0$, a $\|.\|$ – *dual vector*, v', is defined by

$$\|v'\| = 1 \; , \quad \langle v', v \rangle = \|v\|' \; . \tag{3.1}$$

Similarly, the $\|.\|'$ – *dual vector*, v^* , is defined by

$$\|v^*\|' = 1 \; , \quad \langle v^*, v \rangle = \|v\| \; . \tag{3.2}$$

When $\|.\| = \|.\|_p$, $1 < p < \infty$, is the usual l^p-norm, then $\|.\|' = \|.\|_q$, where $p+q = pq$. In terms of coordinates, the dual vectors are given by

$$v_i' = (|v_i|/\|v\|_q)^{q-1} sgn v_i \quad i = 1, ..., n \; ,$$

$$v_i^* = (|v_i|/\|v\|_p)^{p-1} sgn v_i \quad i = 1, ..., n \; .$$

The following two identities are useful and will be often reffered to later.

$$v^{*'} = v/\|v\| \quad and \quad v^{'*} = v/\|v\|' \; , \tag{3.3}$$

for any $v \neq 0$.

Set

$$K = \{\ x \in \mathbf{R}^n \ \mid \ Ax \geq b\ \}\ .$$

If the region K is nonempty, then due to the compactness of the norm $\|.\|$ unit ball in $\mathbf{R}^n$ a solution of problem (P) exists; it is unique due to the strict convexity of the norm $\|.\|$.

Given problem (P), we assotiate a dual problem

$$(P') \qquad max\ \{\ \langle b, y\rangle \mid\ y \in \mathbf{R}^m\ ,\ y \geq 0\ , \|A^T y\|' \leq 1\ \}$$

where A^T is the transpose of the matrix A.

The next theorem generalizes slightly the duality theorem 3.1 in [6] to the case of linear inequality constraints. It establishes the relation between problem (P) and its dual (P'). Since $0 \in K$ iff $b \leq 0$, in which case the value of (P) is zero, we explicitly exclude this easy case.

Theorem 3.1 *Assume that K is non-empty and that $0 \notin K$. Then, problems (P) and (P') have the same value.*

Proof. We begin by showing the classical weak duality inequality. Let $x \in K$, $y \in \mathbf{R}^m$, $y \geq 0$ such that $\|A^T y\|' \leq 1$ .Then

$$\begin{aligned}
\langle b, y\rangle &\leq \langle Ax, y\rangle \qquad (since\ y \geq 0\ ) \\
&= \langle x, A^T y\rangle \\
&\leq \|x\|.\|A^T y\|' \\
&\leq \|x\| \qquad\qquad\qquad\qquad\qquad\qquad (3.4)
\end{aligned}$$

Hence, value of $(P') \leq$ value of (P). To prove the desired, let

$$d = d(0, K) = \inf\ \{\|x\| \mid\ x \in K\}\ .$$

By the duality theorem in Nirenberg [7],

$$d(0, K) = \max \{ -\sigma(z) \mid \|z\|' \leq 1 \} , \tag{3.5}$$

where σ denotes the support function of the convex set K and where we have adopted the notation $d(a, X) = \inf\{ \|a - x\| \mid x \in X \}$. Let z_0 be the maximizer of (3.5). By the definition of σ, we see that

$$\begin{aligned} -\sigma(z_0) &= -\sup \{\langle z_0, x\rangle \mid x \in K \} \\ &= -\sup \{\langle z_0, x\rangle \mid Ax \geq b \} \\ &= \inf \{\langle -z_0, x\rangle \mid Ax \geq b \} \\ &= \sup \{\langle b, y\rangle \mid A^T y = -z_0 \ , \ y \geq 0 \} . \end{aligned}$$

The last step is due to the duality theory in linear programming.

Now, since

$$\sup \{\langle b, y\rangle \mid A^T y = -z_0 \ , \ y \geq 0 \}$$

is finite, there exists $\bar{y} \geq 0$ such that

$$A^T \bar{y} = -z_0 \ , \ \bar{y} \geq 0 \ ,$$

and

$$\langle b, \bar{y}\rangle = \sup \{\langle b, y\rangle \mid A^T y = -z_0 \ , \ y \geq 0 \} .$$

Furtheremore,

$$d(0, K) = \langle b, \bar{y}\rangle . \tag{3.6}$$

Equation(3.6) combined with (3.4) imply the claimed equality.

The following theorem is due to Sreedharan-Nikolopoulos [6] in a slightly different form. Only a few changes are needed in the proof. We include it for completness. It states a characterization of the solution of problem (P) and establishes the relation between the solutions of (P) and (P'). Because of theorem 3.1, only a minor modification is needed for its use in our case.

Theorem 3.2 *Assume that K is non-empty and that $0 \notin K$. Then the following are equivalent*

(i) $\bar{x}$ solves problem (P).

(ii) $A\bar{x} \geq b$ and $\exists y \in \mathbf{R}^m$, $y \geq 0$ such that

$$\|A^T y\|' = 1, \quad \langle y, b\rangle > 0 \ ,$$

and

$$\bar{x} = \langle b, y\rangle (A^T y)' \ .$$

Proof. Due to theorem 3.1, it is easy to see that (ii) implies (i). Since $\|A^T y\|' = 1$,

$$\|\bar{x}\| = \langle b, y\rangle \|(A^T y)'\| = \langle b, y\rangle \ .$$

To prove the converse, let $y \geq 0$ be a solution of problem (P'). Then

$$\|\bar{x}\| = \langle b, y\rangle = \rho \ .$$

Since $A\bar{x} \geq b$ and $y \geq 0$, we have

$$\begin{aligned} \langle A^T y, \bar{x}\rangle &= \langle y, A\bar{x}\rangle \\ &\geq \langle y, b\rangle \\ &= \rho \end{aligned}$$

On the other hand

$$\langle A^T y, \bar{x}\rangle \leq \|A^T y\|' \|\bar{x}\| \leq \rho \ .$$

Thus

$$\langle A^T y, \bar{x}\rangle = \rho \ \ \textit{and} \ \ \|A^T y\|' = 1 \ .$$

So,

$$\langle A^T y, \bar{x}/\rho\rangle = \|A^T y\|' \ .$$

By uniqueness of the $\|.\|$-dual and since $\|\bar{x}/\rho\| = 1$, we have

$$\bar{x}/\rho = (A^T y)' ,$$

which implies

$$\bar{x} = \langle b, y\rangle (A^T y)'$$

completing the proof.

Theorem 3.3 *Assume that K is non-empty and that $0 \notin K$. Then, $\bar{x}$ is the solution of problem (P) iff $A\bar{x} \geq b$ and $\exists y \in \mathbf{R}^m$, $y \geq 0$ such that*

$$\|A^T y\|' = 1 , \tag{3.7}$$

and

$$\bar{x}^* = A^T y . \tag{3.8}$$

Furthermore,

$$\langle b, y\rangle \;=\; \langle \bar{x}^*, \bar{x}\rangle$$

and y is the solution of (P').

Proof. If $\bar{x}$ is the solution of (P), then by theorem 3.2

$$\bar{x}^* = (A^T y)'^* = A^T y ,$$

with $y \geq 0$ and $\|A^T y\|' = 1$, which proves the "only if" part.

"If" part: By (3.8)

$$\bar{x}/\|\bar{x}\| = \bar{x}^{*\prime} = (A^T y)' .$$

Thus

$$1 = \|\bar{x}^*\| = \|A^T y\|' = \langle A^T y, (A^T y)'\rangle = \langle A^T y, \bar{x}/\|\bar{x}\|\rangle .$$

So,

$$\begin{aligned} \|\bar{x}\| &= \langle A^T y, \bar{x} \rangle \\ &= \langle y, A\bar{x} \rangle \\ &\geq \langle b, y \rangle \end{aligned}$$

Now,

$$\langle b, y \rangle = \langle A^T y, \bar{x} \rangle \leq \|A^T y\|' \|\bar{x}\| = \|\bar{x}\| \ .$$

This implies

$$\|\bar{x}\| = \langle b, y \rangle \ .$$

Therefore, $\bar{x}$ and y solve (P) and (P') respectively, as desired.

As mentiond in the discussion at the beginning of this chapter, a key step toward the solution of problem (P) lies in our ability to compute effectively a least distance solution of a related problem. Before we formulate this problem, we set down some notations that will be used throughout this chapter.

Let E be an $m \times n$ real matrix and let $c \in \mathbf{R}^m$. We consider the convex cone

$$C \ = \ \{z \in \mathbf{R}^m \mid E^T z \ \leq 0 \ \}$$

and its *negative* polar

$$C^0 \ = \ \{Ex \mid x \geq 0 \ , \ x \in \mathbf{R}^n \ \} \ .$$

Given E and b, we introduce the following minimization problem

$$\min \{ \ \|c - y\|' \mid y \in C^0 \ \} \tag{3.9}$$

With (3.9) is associated the following dual problem

$$\max \{\langle c, z \rangle \mid z \in C \ , \ \|z\| = 1 \ \} \ . \tag{3.10}$$

Problem (3.9) and its dual (3.10) have the same value (see chapter 1).

The relation between (3.9) and its dual (3.10) was studied in Sreedharan [20] where the following theorem is proved

Theorem 3.4 *Assume that $c \notin C^0$. Let $\bar{x} \geq 0$ be a solution of problem (3.9) and let y be the maximizer of (3.10). Then*

$$E\bar{x} = c - \langle c, y \rangle y^* \tag{3.11}$$

and

$$\langle E\bar{x}, y \rangle = 0 \tag{3.12}$$

Proof. See theorems 3.5 and 3.7 in Sreedharan [20]. Note the interchange of primes and stars since our minimization problem uses the dual norm.

Corollary 3.5 *([20]) If $\bar{x} \geq 0$ is a solution of problem (3.9), then*

$$E^T r' \leq 0 \ , \tag{3.13}$$

where $r = c - E\bar{x}$.

Remark. Since the map $z \mapsto z'$ is odd, positively homogenous of of degree zero on $\mathbf{R}^m \setminus \{0\}$, it is easily seen from (3.11) that

$$\begin{aligned} r' &= (\langle c, y \rangle y^*)' \\ &= y/\|y\| \ , \quad (\textit{by } (3.3) \ ) \\ &= y \qquad (\textit{since } \|y\| = 1) \end{aligned} \tag{3.14}$$

We are now ready to give the algorithm for solving (P) when $\|.\| = \|.\|_p$. The reader will note that for the validity of the present algorithm, it will not be sufficient to assume that the norm $\|.\|$ is just strictly convex and smooth. The special structure of the l^p norm will be used. A careful look at our proof will reveal that the new requirement is the following. Let $v \neq 0$ and v' its $\|.\|$ dual. Then $v'_i = 0$ implies $v_i = 0$, where $v = (v_1, ...v_n)$ and $v' = (v'_1, ..., v'_n)$. The algorithm starts by solving a problem of type (3.9). Then it proceeds to compute the solution of (P). We will state this algorithm for the norms $\|.\| = \|.\|_p$ and $\|.\|' = \|.\|_q$, where $p + q = pq$

Given the matrix A and the vector b, defining problem (P), let

$$E = \begin{bmatrix} A^T \\ b^T \end{bmatrix}$$

and $c = [0, ..., 0, 1]^T$. E is an $(m+1) \times n$-matrix and c an $(n+1)$-vector.

Algorithm 3.1

Step 0. If $b \leq 0$, Then $\bar{x} = 0$ solves (P). GO TO step 6.

Step 1. Find $\bar{u} \geq 0$, a solution of the problem

$$\|c - Ex\|_q \ (min) \ , \quad x \geq 0 \ .$$

Step 2. Compute the residual $r = c - E\bar{u}$.

Step 3. If $r = 0$, the feasible region of (P) is empty. GO TO step 5; else proceed.

Step 4. Compute $r' = (r'_1, ..., r'_{n+1})$, the $\|.\|_p$ dual of r. Let

$$\bar{x}_j = -r'_j / r'_{n+1} \ , \quad j = 1, ..., n \ .$$

Step 5. Accept $\bar{x}$ as the solution of (P).

Step 6. The computation is complete.

Before proceeding any further, some comments are in order here. To find $\bar{u} \geq 0$ in step 1, Algorithm 1.2 of chapter 1 can be used. Step 3 answers the question of the feasibility of the system $Ax \geq b$. To determine feasibility we may use the l^2-residual. If the problem is feasible we actually start all over from step 1 with the given l^q norm. If the region K is non-empty (the system has a solution), then the algorithm will compute the solution of the system, which has minimum norm. If step 3 is answered affirmatively, the region K is empty. In this case we exit the algorithm.

We start by proving the feasibility of the algorithm in

Proposition 3.6 *Let $r = c - E\bar{u}$ be the residual as given by step 1 of the algorithm. If step 3 is answered negatively, then*

$$r'_{n+1} = \|r\|_q \ > \ 0 \ .$$

Proof. Note that since the case $b \leq 0$ has already been handled, $0 \notin K$. In this case let y be the solution of the dual problem (3.10). It follows from (3.12) and (3.14) that

$$\begin{aligned} \langle E^T r', \bar{u} \rangle &= \langle E^T y, \bar{u} \rangle \\ &= \langle y, E\bar{u} \rangle \\ &= 0 \ , \quad (\textit{Due to} \ (\ 3.12)) \ . \end{aligned} \tag{3.15}$$

This implies that

$$\begin{aligned} 0 = \langle E^T r', \bar{u} \rangle &= \langle r', E\bar{u} \rangle \\ &= \langle r', c - r \rangle \ , \quad (\textit{due to step } 1) \\ &= \langle r', c \rangle - \|r\|_q \ . \end{aligned} \tag{3.16}$$

Step 3 of the algorithm is answered negatively if $r \ \neq \ 0$. Hence from (3.16), we obtain

$$r'_{n+1} = \|r\|_q \ > 0, \quad (\textit{because of} \ (\ 3.1)) \tag{3.17}$$

This completes the proof.

In the next result, we show that the stopping criterion of the proposed algorithm is well formulated. We also show the feasibility of the system $Ax \geq b$, if $r \neq 0$.

Theorem 3.7 *Let r be the residual vector, with (n+1) components, given by step 1 of the algorithm, then*

(a) If $\bar{x}$ is as defined in step 4, $\bar{x}$ is the minimal norm solution of problem (P).

(b) If $\|r\|_p = 0$, the system $Ax \geq b$ is inconsistent.

Proof. Let us prove (a). From step 4 of the algorithm, we have

$$\begin{aligned} \bar{x}_j &= -r_j'/r_{n+1}' \\ &= -r_j'/\|r\|_q , \quad j = 1, \dots n . \end{aligned} \tag{3.18}$$

To verify the feasibility of $\bar{x}$, we need to show $A\bar{x} \geq b$. By (3.17), we have

$$\begin{aligned} -\|r\|_q[\bar{x}, -1]^T &= (-r_{n+1}')[\bar{x}, -1]^T \\ &= (-r_{n+1}')[-r_1'/r_{n+1}', \dots, -r_n'/r_{n+1}', -1]^T \\ &= [r_1', \dots, r_n', r_{n+1}']^T \\ &= r'^T \end{aligned}$$

Hence

$$-[A, b]\begin{bmatrix} \bar{x} \\ -1 \end{bmatrix}\|r\|_q = E^T r' . \tag{3.19}$$

This combined with the inequality (3.13) of corollary 3.4 implies

$$(b - A\bar{x})\|r\|_q = E^T r' \leq 0 .$$

Thus, $A\bar{x} \geq b$, as claimed.

If $\bar{x} = 0$, then due to the feasibility we just proved, $b \leq 0$ and we are done. So assume that $\bar{x} \neq 0$. Then, by step 4, $r_j' \neq 0$ for some j, $1 \leq j \leq n$. As observed earlier, this implies $r_j \neq 0$, due to the special form of the $\|.\|_p$ dual in the case of the l^p norm. So, $\tilde{r} = (r_1, \dots, r_n) \neq 0$. Now we have

$$\begin{aligned} r_j' &= (|r_j|/\|r\|_q)^{q-1} sgn\ r_j \\ &= |r_j|^{q-1} sgn\ r_j/\|r\|_q^{q-1} , \quad j = 1, \dots n . \end{aligned}$$

From this it follows that

$$
\begin{aligned}
-\bar{x}_j &= |r_j|^{q-1} sgn\ r_j / \|r\|_q \|r\|_q^{q-1} \\
&= |r_j|^{q-1} sgn\ r_j / \|r\|_q^q \\
&= (|r_j|^{q-1} / \|\tilde{r}\|_q^{q-1}) sgn\ r_j . (\|\tilde{r}\|_q^{q-1} / \|r\|_q^q) , \quad j = 1, \dots, n .
\end{aligned}
$$

So,

$$-\bar{x} = \tilde{r}' . (\|\tilde{r}\|_q^{q-1} / \|r\|_q^q) .$$

Using the fact that the map $z \mapsto z^*$ is odd, positively homogenous of degree zero on $\mathbf{R}^n \setminus \{0\}$, we get

$$-\bar{x}^* = \tilde{r}'^* = \tilde{r} / \|\tilde{r}\|_q . \tag{3.20}$$

By definition of E and c, we have $\tilde{r} = -A^T \bar{u}$. Thus

$$
\begin{aligned}
\bar{x}^* &= A^T \bar{u} / \|\tilde{r}\|_q \\
&= A^T (\bar{u} / \|\tilde{r}\|_q) , \quad \bar{u} \geq 0 .
\end{aligned} \tag{3.21}
$$

This implies that $\bar{x}$ satisfies equation (3.8) with $y = \bar{u}/\|r\|_q$, $y \geq 0$ and $\|A^T y\|_q = 1$. We have verified that $\bar{x}$ satisfies the conditions of theorem 3.3, so that $\bar{x}$ is the solution of problem (P). The proof of (a) is now complete.

To prove (b), assume that $r = 0$ and that there exists a solution $\tilde{x}$ of the system $Ax \geq b$. We have

$$c - \begin{bmatrix} A^T \\ b^T \end{bmatrix} \bar{u} = r = 0 .$$

Since $c = [0, \dots 0, 1]^T$, we easily see that

$$A^T \bar{u} = 0 , \quad b^T \bar{u} = 1 . \tag{3.22}$$

By step 1 of the algorithm, $\bar{u} \geq 0$, so

$$1 = b^T \bar{u}$$

$$
\begin{aligned}
&\leq \langle A\tilde{x}, \bar{u}\rangle \\
&= \langle \tilde{x}, A^T\bar{u}\rangle \\
&= 0\ , \quad (by\ (\ 3.22))\ ,
\end{aligned}
$$

a contradiction. Thus, the system $Ax \geq b$ has no solution, as claimed. This completes the proof.

We close this chapter with the following observations.

1. With the algorithm of chapter 1 at hand, the present algorithm for finding the minimal norm solution of a system of linear inequalities is easy to implement.

2. A consequence of this algorithm is to determine whether the system of linear inequalities under consideration is consistent or not, as shown in step 3. But if this is all that one is interested in then one would use the l^2-norm in place of the l^q-norm in solving the problem stated in step 1.

Chapter 4

Numerical results

In this chapter we discuss the computational aspects of the algorithm presented in chapter 1 of this work. At this time we will not investigate the numerical results of the algorithm of chapter 2. The whole implementation of this algorithm will be presented elsewhere.

The main computational difficulty encountered in this algorithm is finding $\alpha_k > 0$ such that

$$\|y_k + \alpha_k r_k\| = 1 + \alpha_k \mu_k \ .$$

This search occurs in step 5 at each iteration cycle.

Suppose that we are at the stage of entering step 5. Let the function F be defined by

$$F(\alpha) = 1 + \alpha\mu_k - \|y_k + \alpha r_k\|' \ . \tag{4.1}$$

We are searching for $\alpha_k > 0$ such that

$$F(\alpha_k) = 0 \ .$$

It is well known that

$$\frac{d}{d\alpha}\|y_k + \alpha r_k\|' = \langle (y_k + \alpha r_k)', r_k \rangle \tag{4.2}$$

(see for example [14].).

Assume that the search for α_k has been reduced to an interval (β, γ), $\gamma > \beta \geq 0$, with $F(\beta) > 0$ and $F(\gamma) < 0$. We begin by fitting a quadratic $q(\alpha)$ on the interval $[\beta, \gamma]$ as follows

$$\begin{cases} q(\beta) = F(\beta) := F_1 \\ q(\gamma) = F(\gamma) := F_2 \\ q'(\beta) = F'(\beta) := F_1' \end{cases} \tag{4.3}$$

We seek the roots $\hat{\alpha}$ of q. If $|F(\hat{\alpha})| \leq \eta$ and $\hat{\alpha} \in (\beta, \gamma)$, where η is a given tolerance parameter, then we set

$$\alpha_k = \hat{\alpha}$$

and return to the main algorithm. If the stopping condition $|F(\hat{\alpha})| \leq \eta$ is not met but $\hat{\alpha}$ belongs to (β, γ), we reduce the interval of search by setting

$$\begin{cases} \gamma = \hat{\alpha} & \text{if } F(\hat{\alpha}) < 0 \\ \beta = \hat{\alpha} & \text{if } F(\hat{\alpha}) > 0 \end{cases}$$

and then apply the routine to the new reduced interval (β, γ), till an acceptable α_k is obtained.

In the case when the root $\hat{\alpha}$ is not in the interval of search (β, γ) or if the quadratic interpolation has no real root, we consider the quadratic fitting as not suitable. A linear interpolation is then performed to determine $\hat{\alpha}$, i.e

$$\hat{\alpha} = (\beta F_2 - \gamma F_1)/(F_2 - F_1) \ , \tag{4.4}$$

followed by an update of the interval (β, γ), as was done in the quadratic fitting case.

Let

$$q(\alpha) = A(\alpha - \beta)^2 + F_1'(\alpha - \beta) + F_1 \tag{4.5}$$

be the quadratic interpolation defined via (4.3). It is easily seen that

$$A = (F_2 - F_1 - F_1'(\gamma - \beta))/(\gamma - \beta)^2 \ . \tag{4.6}$$

The roots of q are

$$\hat{\alpha} - \beta = (-F_1' \pm \sqrt{F_1'^2 - 4AF_1}\ )/2A \ . \tag{4.7}$$

Recall that the quadratic interpolation is considered under the conditions $F_1 > 0$ and $F_2 < 0$. So, q has its maximum in $(-\infty, \hat{\alpha})$. It follows that

$$q'(\hat{\alpha}) = 2A(\hat{\alpha} - \beta) + F_1' \leq 0 \ .$$

Thus, the only relevant root in (4.7) is

$$\hat{\alpha} - \beta = (-F_1' - \sqrt{F_1'^2 - 4AF_1}\ )/2A \ .$$

The following subalgorithm is based on the above discussion. For a further refinement of the interval of search (β, γ), we included a bisection to be performed at each iteration cycle of the subalgorithm.

4.1 Subalgorithm

Step 0. Let $\beta = 0$ and γ be such that $F(\gamma) < 0$. Let $\epsilon > 0$

Step 1. Let $F_1 = F(\beta)$, $F_1' = F'(\beta)$ and $F_2 = F(\gamma)$.

Step 2. Let $h = \gamma - \beta$.

Step 3. Compute $A = (F_2 - F_1 - F_1'h)/h^2$. If $A = 0$, GO TO step 7; else proceed.

Step 4. Let $\Delta = F_1'^2 - 4AF_1$. If $\Delta < 0$, GO TO step 7; else proceed.

Step 5. Set

$$\hat{\alpha} = \beta + (-F_1' - \sqrt{\Delta})/2A \ .$$

Step 6. If $\beta < \hat{\alpha} < \gamma$, GO TO step 8; else proceed.

Step 7. Let

$$\hat{\alpha} = (\beta F_2 - \gamma F_1)/F_2 - F_1) \ .$$

Step 8. If $|F(\hat{\alpha})| \leq \epsilon$, set

$$\alpha_k = \hat{\alpha}$$

and RETURN to the main program; else proceed.

Step 9. If $F(\hat{\alpha}) < -\epsilon$, set $\gamma = \hat{\alpha}$; else proceed.

Step 10. Set $\beta = \hat{\alpha}$.

Step 11. Let $\hat{\alpha} = (\beta + \gamma)/2$.

Step 12. If $|F(\hat{\alpha})| \leq \epsilon$, set $\alpha_k = \hat{\alpha}$ and return to the main program; else proceed.

Step 13. If $F(\hat{\alpha}) < -\epsilon$, set $\gamma = \hat{\alpha}$; else proceed

Step 14. Set $\beta = \hat{\alpha}$ and RETURN to step 1.

A Newton method can also be incorporated within this subalgorithm. As noted in chapter 1, in the l^p-case the dual of a given non-zero vector z is given by

$$z_i^{'} = (|z_i|/\|z\|_q)^{q-1} sgn z_i \ , \quad i = 1, ...m \ .$$

This particulary simple formula of z' makes the calculation of the derivative in (4.2) immediate. We use the Newton iterations as follows. Via the quadratic model, we determine $\hat{\alpha}$ belonging to (β, γ), then we start the Newton iteration at $\hat{\alpha}$. We compute

$$\alpha^* = \hat{\alpha} - F(\hat{\alpha})/F'(\hat{\alpha}) \ . \tag{4.8}$$

Having this new approximation, we check if α^* belongs to (β, γ) and if it yields an actual decrease the value of F. Only under these conditions we continue the Newton iterations until an acceptable $\hat{\alpha}$ is reached. If one of the conditions

$$(a) \ \ \hat{\alpha} \in (\beta, \gamma)$$

$$(b) \ \ |F(\alpha^*) < |F(\hat{\alpha})|$$

is not met, we exit the Newton iteration and return to the quadratic model.

The algorithm 1.2 suggested in chapter 1 was coded in Fortran 77 for a SPARC station. The norm considered is the l^p-norm

$$\|x\|_p = (\sum |x_i|^p)^{1/p}$$

for various values of p. The dual norm is the $\|.\|_q$ norm, $q = p/(p-1)$. The stopping rule parameter whithin which we consider the tolerance for the duality gap as acceptable is $\epsilon = 10^{-6}$. The code was run in double precision.

We calculated the sequence (α_k) using the subalgorithm outlined in this chapter. We also incorporated a Newton iteration scheme, as discussed earlier, to determine each α_k.

The coded version of the main algorithm seems to do much better for $p \geq 2$ than for p in the range $(1, 2)$.

When $p > 2$ and not far away from 2, the convergence seems to do well compared to [20]. We also found that the sequence $(\langle b, y_k \rangle)$ increases and the duality gap decreases monotonically. However, in some cases, the sequence (α_k) poses more problems, e.g, it may converge to two different limits.

The case $1 < p < 2$ does not do as well. For example for $p = 1.8$ a significantly larger amount of iterations were needed to reach the same acceptable tolerance of 10^{-6}. The sequence $(\langle b, y_k \rangle)$ still increased monotonically. The duality gap decreased in the same way.

In conclusion, compared to the algorithm in [20], the present algorithm seems to perform well for values of p larger than 2, but not as well for p in the range $(1, 2)$ with regard to the number of iterations.

The following linear system is taken from Barrodale and Young [1] and was used by Sreedharan [20] and others ([8], [10]) for $1 < p < \infty$.

$$\begin{aligned}
x_1 &= 1.52 \\
x_1 + x_2 &= 1.025 \\
x_1 + 2x_2 &= 0.475 \\
x_1 + 3x_2 &= 0.01 \\
x_1 + 4x_2 &= -0.475 \\
x_1 + 5x_2 &= -1.005
\end{aligned}$$

We record the results in the following table.

p	x_1	x_2	ρ	iterations
5.0	0.258716	0.000334	1.472222	149
4.5	0.253714	0.001339	1.507273	59
4.0	0.261537	0.000031	1.546597	28
3.8	0.261783	0.000000	1.568348	27
3.5	0.261952	0.000000	1.607494	24
3.0	0.261791	0.000000	1.697914	20
2.5	0.260704	0.000000	1.842740	15
2.0	0.258333	0.000000	2.102851	1
1.8	0.239018	0.0019	2.280894	165

Table 4.1: $1.8 \leq p \leq 5$

Algorithm 1.2 was also coded in Fortran 77 for a SPARC station. The results are presented for $p = 3$ (table 4.1) and for various of $\lambda_k = 1/\delta_k$. The tolerence parameter is $\epsilon = 10^{-6}$. For $p = 3$, the numerical results suggest that the algorithm converges faster for λ_k near $\frac{1}{2}$. The convergence tends to be slower for λ_k far away from $\frac{1}{2}$. We used the same example as above.

δ_k	x_1	x_2	ρ	iterations
1.8	0.26179422	0.0	1.69791476	16
1.9	0.26179173	0.0	1.69791476	18
2.0	0.26179170	0.0	1.69791476	20
2.2	0.26179069	0.0	1.69791476	20
2.5	0.26179026	0.0	1.69791476	23
2.8	0.26179038	0.0	1.69791476	26
3.0	0.26179090	0.0	1.69791476	29
4.0	0.26179037	0.0	1.69791476	39
5.0	0.26179041	0.0	1.69791476	50
10.0	0.26179035	0.0	1.69791476	104

Table 4.2: $p = 3$

Bibliography

[1] Barrodale, I. and Young, A., Algorithms for best L_1 and L_∞ linear approximation on a discrete set, *Numer. Math.*, 8, (1966), 295-306.

[2] Cheney, E. W.,"Introduction to Approximation theory",2nd ed., McGraw-Hill, New York, 1982.

[3] Fletcher, R., Grant,J. A. and Hebden, M. D., The calculation of linear best L_p approximation, *Computer J.*, 14, (1971), 276-279.

[4] Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H., Linearly constrained optimization, in *Nonlinear Optimization 1981*, M. J. D. Powell, ed., Academic Press, London (1982), 123-139.

[5] Lawson, C.L. and Hanson, R. J.,"Solving Least Squares Problems", Prentice Hall, Englewood Cliffs, N.J. (1974).

[6] Nikolopoulos, P. V. and Sreedharan V. P., An algorithm for computing non negative minimal norm solutions, *Numer. Funct. Anal. and Optimiz.*. To appear.

[7] Nirenberg, L., "Functional Analysis", Lectures given in 1960-61, notes by Lesley Sibner, New York University, (1961).

[8] Owens, R. W., An algorithm for best approximate solutions of $Ax = b$ with a smooth strictly convex norm, *Numer. Math.*, 29, (1977), 83-91.

[9] Owens, R. W. and Sreedharan, V. P., An Algorithm for approximation by elements of a cone in a Banach space, *Numer. Funct. Anal. and Optimiz*, 10, (1989), 1161-1189.

[10] Owens, R. W. and Sreedharan V. P., Least squares methods to minimize errors in a smooth, strictly convex norm, *J. Approx. Theory*, 2, (1993), 180-198.

[11] Polak, E., "Computational Methods in Optimization", Academic Press, New york, N.Y (1971).

[12] Rockafellar, R. T., "Convex Analysis", Princeton University Press, Princeton, N.J, 1969.

[13] Späth, H."Mathematical Algorithms for Linear Regression", Academic Press, Orlando, FL, in press.

[14] Sreedharan, V. P., Least squares algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.*, 17, (1971), 387-401.

[15] 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. Approx. Theory,* 8, (1973), 46-61.

[16] Sreedharan, V. P. , A subgradient projection algorithm, *J. Approx. Theory*, 35, (1982), 111-126.

[17] Sreedharan, V. P., Subgradient projection algorithm II , *J. Approx. Theory*, 41, (1984), 217-243.

[18] Sreedharan, V. P., Extensions of subgradient projection algorithms, *J. Approx. Theory*, 47,(1986), 228-239.

[19] Sreedharan, V. P., ϵ-Subgradient projection algorithm, *J. Approx. Theory*, 51, (1987), 27-46.

[20] Sreedharan, V. P., An algorithm for a non negative norm minimal solutions, *Numer. Funct. Anal. and Optimiz.*, 9, (1987), 193-232.

[21] Watson, G. A., "Approximation Theory and Numerical Methods", John Wiley and Sons, New York, N.Y. (1980).

[22] Zangwill, W. I., "Nonlinear Programming: A unified approach", Prentice Hall, Englewood Cliffs, N. J., (1969).

[23] Zoutendijk, G., "Methods of Feasible Directions", Elsevier, Amsterdam (1960).