JOURNAL OF APPROXIMATION THEORY 8, 46–61 (1973)

Least-Squares Algorithms for Finding Solutions of Overdetermined Systems of Linear Equations Which Minimize Error in a Smooth Strictly Convex Norm

V. P. SREEDHARAN

Department of Mathematics, Michigan State University, East Lansing, Michigan 48823

Communicated by E. W. Cheney

Received February 1, 1971

DEDICATED TO PROFESSOR I. J. SCHOENBERG ON THE OCCASION OF HIS 70TH BIRTHDAY

1. INTRODUCTION

Let A be an $m \times n$ matrix and b a given m-vector. We shall assume that all matrices and vectors which occur in this paper are over the real field $\mathbb{R}$. Let $\| \cdot \|$ be a norm on $\mathbb{R}^m$. The problem we wish to study is that of giving algorithms for finding $x \in \mathbb{R}^n$, minimizing

$$\{\| \eta(x)\| \mid x \in \mathbb{R}^n\}, \tag{1.1}$$

where

$$\eta(x) = b - Ax. \tag{1.2}$$

We call this problem (P). Assuming that $\| \cdot \|$ is both smooth and strictly convex, we gave a few algorithms for solving the foregoing problem in [5]. Recall that $\| \cdot \|$ is said to be strictly convex, if and only if, $\| x \| = \| y \| = 1$, $\| x + y \| = 2$, implies $x = y$. $\| \cdot \|$ is said to be smooth, if and only if, through each point of unit norm there passes precisely one hyperplane supporting the closed unit ball $B = \{x \in \mathbb{R}^m \mid \| x \| \leqslant 1\}$. (Definitions and properties of smooth norms are given in [1].)

When we equip $\mathbb{R}^m$ with the norm

$$\| y \| = \left(\sum_{i=1}^{m} w_i \mid y_i \mid^p\right)^{1/p}, \tag{1.3}$$

where $w_i > 0$, for $i = 1,..., m$ the algorithms in [5] cover the weighted l^p, $(1 < p < \infty)$ case. An algorithm appears in [3] when $p > 2$. In [4] we considered a dual problem (P'). For easy reference we now state the dual problem (P'). Given $\| \cdot \|$ on $\mathbb{R}^m$, as was done in [2], we define $\| \cdot \|'$, the norm dual to $\| \cdot \|$ by

$$\| u \|' = \max_{\| v \| = 1} (u \mid v), \tag{1.4}$$

where, here and throughout, $(\,|\,)$ denotes the standard inner product on $\mathbb{R}^m$, i.e..

$$(u \mid v) = \sum_{i=1}^{m} u_i v_i . \tag{1.5}$$

The dual problem (P') is:

$$(P') \qquad \begin{cases} \text{Find } y \in \mathbb{R}^m \text{ such that } A'y = 0, \quad \| y \|' = 1, \\ \text{and } (b \mid y) \text{ is a maximum,} \end{cases}$$

where A' is the transpose of A. The algorithms in [5] were directed toward solving (P') so that the duality theorem [4, Theorem 2.1] yielded a solution of problem (P).

Here we show that algorithms can be given to solve the problem (P) directly, and that there is no need to solve the problem (P') initially. The present procedures, therefore, avoid the intermediate minimization sub-problem occuring in each stage of step 3 of the iteration in Algorithm 3.1 [5]. Besides this, the present algorithms should be computationally better behaved than the ones in [5], since in general $n < m$, so that the transposed matrix A' occuring in problem (P') has less rows than columns. Such matrices behave poorly in actual numerical computations. It is very likely, that in the concrete l^p-case (Eq. (1.3)), the present algorithms are much better in the range $2 < p < \infty$; whereas, the algorithms in [5] are more suited for $1 < p < 2$. (We already have some numerical evidence to this effect.)[1]

We will also see later that we can associate another problem to be denoted (P^*), which is also a maximization problem like (P'). The solutions of (P) and (P^*) are more directly connected (see Theorem 2.4 and Corollary 4.5) than the solutions of the pair of problems (P) and (P'). The connection between the solutions of (P) and (P') is the content of the Duality theorem in [4], which depended on strict convexity of the $\| \cdot \|$. That the Theorem 2.4 holds for any norm illustrates again the direct connection between (P) and (P^*).

Recall that v' is said to be a $\| \cdot \|$-dual of $v \neq 0$ if, and only if,

$$\| v' \| = 1 \qquad \text{and} \qquad (v' \mid v) = \| v \|'. \tag{1.6}$$

Analogously, let us say that v^* is a $\| \cdot \|'$-dual of $v \neq 0$, if and only if,

$$\| v^* \|' = 1 \qquad \text{and} \qquad (v^* \mid v) = \| v \|. \tag{1.7}$$

[1] The actual numerical implementation of the algorithms appearing in this paper has been carried out by C. S. Duris using Householder transformations. His report with computer programs, respective execution times, iterations, and other information is being prepared for publication in *Numer. Mat.*

We have convenient explicit expressions for the $\|\cdot\|$-dual and $\|\cdot\|'$-dual when the $\|\ \|$ is given by (1.3). In fact, let us write for given weights $w_i > 0$,

$$w = (w_1, \dots, w_m) \tag{1.8}$$

and define

$$w^\alpha = (w_1^\alpha, \dots, w_m^\alpha) \tag{1.9}$$

for $\alpha \in \mathbb{R}$. Suppose, we denote the norm in (1.3) by $\| y \|_{p,w}$, i.e.,

$$\| y \|_{p,w} = \left(\sum_{i=1}^{m} w_i \mid y_i \mid^p \right)^{1/p}. \tag{1.10}$$

It is easy to verify that if

$$\|\cdot\| = \|\cdot\|_{p,w}, \tag{1.11}$$

then

$$\|\cdot\|' = \|\cdot\|_{q,w^{1-q}}, \tag{1.12}$$

where q is the conjugate to p, viz. $(1/p) + (1/q) = 1$. Moreover, if $y \neq 0$, then y', $\|\cdot\|$-dual of y, has the components

$$y_i' = \left(\frac{\mid y_i \mid}{w_i \| y \|'} \right)^{q-1} \operatorname{sgn} y_i, \tag{1.13}$$

where $\| y \|'$ is given by (1.12). Also y^*, $\|\cdot\|'$-dual of y, is given by its components

$$y_i^* = \frac{1}{\| y \|_{p,w}^{p-1}} w_i \mid y_i \mid^{p-1} \operatorname{sgn} y_i. \tag{1.14}$$

We introduce some more notation and terminology which will remain standard throughout the sequel. Primes will denote $\|\cdot\|$-duals, and stars, $\|\cdot\|'$-duals of a given vector. We shall refer to problem (P), when $p = 2$ and $w_i = 1$, for all i, as the l^2-problem (P). The corresponding l^2-norm is denoted by $\|\cdot\|_2$.

Let E be the orthogonal projection (orthogonal, for the inner product (1.5)) of $\mathbb{R}^m$ onto $K = \ker A' = \{x \in \mathbb{R}^m \mid A'x = 0\}$. $\operatorname{Im}(A) = \{Ax \in \mathbb{R}^m \mid x \in \mathbb{R}^n\}$. It is well known that $\operatorname{Im}(A) = K^\perp = \{z \in \mathbb{R}^m \mid (z \mid k) = 0, \forall k \in K\}$. Also we define $d(b, K^\perp) = \inf\{\| b - z \| \mid z \in K^\perp\}$. Let $s = Eb$, so that s is the minimal error of l^2-problem (P). We assume throughout that $s \neq 0$. By (v) we denote the linear span of the vector v.

2.

We proceed to establish a few theorems which explain the structure of problem (P). They will also be used in the proof of convergence of the algorithms to be given in Section 3. We begin with a lemma.

2.1. LEMMA. *If $\|\cdot\|$ is smooth, then $\|\cdot\|'$-duals are unique, i.e., the *-operation is a single valued map of $\mathbb{R}^m\backslash\{0\}$ onto the $\|\cdot\|'$-unit sphere. Moreover, if $v \neq 0$, then*

$$v'^* = (1/\|v\|')v. \tag{2.1.1}$$

*Also the *-map is continuous on $\mathbb{R}^m\backslash\{0\}$.*

Proof. This is a routine verification and is immediate if we recall that $\|\cdot\|$ is smooth, if and only if, $\|\cdot\|'$ is strictly convex. [5, Section 2]. If $u \neq w$ are both $\|\cdot\|'$-duals of $v \neq 0$, so that

$$\|u\|' = 1 = \|w\|' \qquad \text{and} \qquad (u \mid v) = \|v\| = (w \mid v),$$

then

$$\|v\| = (\tfrac{1}{2}(u + w) \mid v) \leqslant \|\tfrac{1}{2}(u + w)\|' \, \|v\| < \|v\|,$$

a contradiction. (2.1.1) is immediate from this uniqueness property. Regarding the continuity of the *-operation; if v_j, $v \in \mathbb{R}^m\backslash\{0\}$ with $v_j \to v$, to show $v_j^* \to v^*$. Since $\|v_j^*\|' = \|v^*\|' = 1$, due to compactness of the $\|\cdot\|'$-unit sphere, it is sufficient to show that (v_j^*) has v^* as a unique adherent point. If $v_{j_k}^* \to u$, then since $(v_{j_k}^* \mid v_{j_k}) = \|v_{j_k}\|$ we see that $(u \mid v) = \|v\|$ and $\|u\|' = 1$, so that $u = v^*$.

2.2. THEOREM. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$. Suppose*

$$z \in K^\perp \oplus (s), \qquad \|z\| = 1 \tag{2.2.1}$$

and that $z^* \in K$.

Then

$$b - (b \mid z^*)z \in \operatorname{Im}(A), \tag{2.2.2}$$

and ρ the $\|\cdot\|$-minimal error of problem (P) is given by

$$\rho = d(b, K^\perp) = (b \mid z^*). \tag{2.2.3}$$

Proof. A proof of this can be obtained from part of Theorem 5.1 of [5]. Here is a more direct proof. Since $s \in K$, we have

$$Ez = \frac{(z \mid s)}{(s \mid s)}\, s. \tag{2.2.4}$$

$$\begin{aligned} 1 = \|z\| &= (z^* \mid z) \\ &= (Ez^* \mid z), \qquad \text{since} \quad z^* \in K, \\ &= (z^* \mid Ez) = \frac{(z^* \mid s)(z \mid s)}{(s \mid s)}, \qquad \text{by (2.2.4)}. \end{aligned} \tag{2.2.5}$$

Since $\mathrm{Im}(A) = K^\perp$, to prove (2.2.2) is to show that

$$(b \mid k) - (b \mid z^*)(z \mid k) = 0, \qquad \forall k \in K. \tag{2.2.6}$$

Since $Eb = s$ and $z^* \in K$, this is same as requiring to show that

$$(s \mid k) - (s \mid z^*)(z \mid k) = 0, \qquad \forall k \in K. \tag{2.2.7}$$

Now (2.2.7) is clear if $k \in K$ and $k \perp s$, for then $z \perp k$. Hence, to establish (2.2.7) we need to check it only when $k = s$. But this is (2.2.5), which proves (2.2.2). Also,

$$\rho = d(b, K^\perp) \leqslant \| b - (b - (b \mid z^*)z)\| = (b \mid z^*), \tag{2.2.8}$$

whereas for every $k \in K^\perp$,

$$\| b - k \| \geqslant (b - k \mid z^*) = (b \mid z^*), \tag{2.2.9}$$

completing the proof of (2.2.3).

2.3. Corollary. *Let z be as in the theorem with $z^* \in K$. Then*

$$(b \mid z^*) = \frac{(s \mid s)}{(s \mid z)}. \tag{2.3.1}$$

Proof. Since $z^* \in K$, $(b \mid z^*) = (s \mid z^*)$; but by (2.2.5) $(s \mid z^*)(s \mid z) = (s \mid s)$, which yields (2.3.1). Q.E.D.

Given problem (P), as remarked earlier, we associated a dual problem (P') in terms of $\|\cdot\|'$, the dual norm of $\|\cdot\|$. As shown by the following theorem, we can also associate a dual problem (P^*) directly in terms of $\|\cdot\|$ itself. The theorems also make precise how a solution of (P^*) yields a solution of (P).

2.4. Theorem. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$ and $s \in K$ be the l^2-error vector of problem (P), i.e., $s = Eb$. Consider the problem*

$$(P^*) \qquad M = \max_w (s \mid w), \qquad \textit{subject to} \quad \| w \| = 1, \quad w \in K^\perp \oplus (s).$$

Then

$$\rho M = (s \mid s), \tag{2.4.1}$$

where

$$\rho = \|\cdot\|\text{-minimal error of problem } (P). \tag{2.4.2}$$

Moreover, the equation system

$$Ax = b - M^{-1}(s \mid s)z, \tag{2.4.3}$$

where z is a solution of problem (P^*), has an exact solution which is a $\|\cdot\|$-minimal solution of problem (P).

Proof. Evidently $M < \infty$ and the max is attained for some z. Now

$$\begin{aligned} M &= \max\{(s \mid k + \beta s) \mid \|k + \beta s\| = 1, k \in K^\perp, \beta \in \mathbb{R}\} \\ &= (s \mid s) \max\{\beta \mid \|k + \beta s\| = 1, k \in K^\perp, \beta \in \mathbb{R}\}, \text{ since } s \in K, \\ &= (s \mid s) \max\{\beta \mid \|k + s\| = (1/\beta), k \in K^\perp, \beta \neq 0\} \\ &= (s \mid s) \max\{(1/\|k + s\|) \mid k \in K^\perp\} \\ &= (s \mid s)[\min\{\|k + s\| \mid k \in K^\perp\}]^{-1} \\ &= (s \mid s)/\rho, \text{ which is (2.4.1), cf. (2.2.3).} \end{aligned}$$

To show that $Ax = b - M^{-1}(s \mid s)z$ has an exact solution is to show that

$$(b \mid k) - M^{-1}(s \mid s)(z \mid k) = 0, \qquad \forall k \in K, \tag{2.4.4}$$

i.e.,

$$(s \mid k) - M^{-1}(s \mid s)(z \mid k) = 0, \qquad \forall k \in K. \tag{2.4.5}$$

This is clear if $k \in K$ is such that $k \perp s$, for then $k \perp z$, since $z \in K^\perp \oplus (s)$. So we need to verify (2.4.5) only when $k = s$. Due to (P^*), $(z \mid s) = M$, so that (2.4.5) follows when $k = s$.

Also

$$\|\eta(x)\| = \|b - Ax\| = M^{-1}(s \mid s)\|z\| = \rho, \qquad \text{by (2.4.1). Q.E.D.}$$

2.5. Theorem. *Let $\|\cdot\|$ be a smooth, strictly convex norm and z a solution of problem (P^*). Then*

$$z^* \in K, \tag{2.5.1}$$

and

$$\rho = (b \mid z^*). \tag{2.5.2}$$

Proof. We shall give two proofs; the first will depend on Theorem 5.1 of [5]. The second proof is independent of this theorem and is presented in Section 4.3.

First only assume that $\|\cdot\|$ is strictly convex. Let $y \in K$ be such that $\|y\|' = 1$ and $(s \mid y) = \rho$. Such a y exists since problem (P') has a solution. By Theorem 5.1 of [5] then $Ey' \in (s)$. (Recall primes denote $\|\cdot\|$-duals.) In other words,

$$y' \in K^\perp \oplus (s).$$

Also Eq. (5.3) of [5] shows that

$$(s \mid y)(y' \mid s) = (s \mid s).$$

Hence

$$(y' \mid s) = (s \mid s)/\rho = M \qquad \text{by (2.4.1)}. \tag{2.5.3}$$

Since $\| y' \| = 1$, $y' \in K^{\perp} \oplus (s)$ and (2.5.3) holds, we see that y' is a solution of problem (P^*). Due to strict convexity of $\| \cdot \|$, $z = y'$.

Now assume that $\| \cdot \|$ is also smooth, besides being strictly convex. Then by Lemma 2.1, $*$ is a single-valued map so that

$$z^* = y'^*.$$

Again by Lemma 2.1, $y'^* = y/\| y \|' = y$. This shows that $z^* \in K$.

Equation (2.5.2) is now clear. Q.E.D.

2.6. Lemma. *Let*

$$\rho = d(b, K^{\perp}) > 0. \tag{2.6.1}$$

Suppose $\beta > 0$ *and* $v \in K^{\perp}$ *such that*

$$\| v + \beta s \| \leqslant 1. \tag{2.6.2}$$

Then

$$\beta \leqslant (1/\rho). \tag{2.6.3}$$

Proof.

$$\rho = \inf_{k \in K^{\perp}} \| b - k \| = \inf_{k \in K^{\perp}} \| s + (I - E)b - k \|$$

$$= \inf_{k \in K^{\perp}} \| s - k \|, \qquad \text{since} \quad (I - E)b \in K^{\perp}.$$

If (2.6.2) holds then $\| (1/\beta)v + s \| \leqslant (1/\beta)$. Now since, $-(1/\beta)v \in K^{\perp}$ we get $(1/\beta) \geqslant \rho$. Q.E.D.

2.7. Lemma. *Let* $\| \cdot \|$ *be a smooth norm on* $\mathbb{R}^m$ *and* $z, w \in \mathbb{R}^m$ *be linearly independent. Then* $\alpha \in \mathbb{R}$ *is such that*

$$\| z - \alpha w \| \leqslant \| z - \lambda w \|, \qquad \forall \lambda \in \mathbb{R} \tag{2.7.1}$$

if and only if

$$((z - \alpha w)^* \mid w) = 0. \tag{2.7.2}$$

Proof. Refer Lemma 5.1 of [5] and note that primes in that lemma go over into stars here, since $\| \cdot \|$ is used in place of $\| \cdot \|'$. This change also explains the use of smoothness in the hypothesis of the present lemma.

3.

In this section the algorithms for solving (P) are given. Throughout this section we assume that $\|\cdot\|$ is both strictly convex and smooth.

3.1. ALGORITHM. *Step* 1. Let $y_1 = s/\|s\|$, where $s = Eb$, the usual minimal error of l^2-problem (P). Or, more generally, let $y_1 \in K^{\perp} \oplus (s)$, with $\|y_1\| = 1$ and $(y_1 \mid s) > 0$.

Step 2. Find y_1^* the $\|\cdot\|'$-dual of y_1. In the weighted l^p-case, it is easy to find the y_1^*. (See (1.14)).

Step 3. Put $v_1 = Fy_1^*$, where $F = I - E$. If $\|v_1\|_2$, is small, solve the equation system,

$$Ax = b - \frac{(s \mid s)}{(s \mid y_1)} y_1,$$

in l^2-sense and take this as a solution of problem (P). If $\|v_1\|_2$ is not small, proceed as follows. Other tests for a solution to be acceptable may be formulated using the so-called "duality gap," since $\|\eta(x)\| \geqslant (b \mid y)$ for every $y \in K$, $\|y\|' = 1$ with equality only when we have solutions of (P) and (P').

Step 4. Choose $\alpha > 0$ such that

$$\|y_1 - \alpha v_1\| = 1.$$

Step 5. Choose $\beta > 0$ such that

$$\|y_1 - (\alpha/2)\, v_1 + \beta s\| = 1.$$

Let $y_2 = y_1 - (\alpha/2)\, v_1 + \beta s$. Replace y_1 by y_2 and go to Step 2 onwards carrying out the iteration.

The existence of $\alpha > 0$ satisfying Step 4 is proved in Lemma 4.1. Any one of the procedures in [5] for solving $\|z_1 - \alpha w_1\|' = 1$, in the notation of that paper, is applicable here. Hence, we shall omit a description of an algorithm for finding α.

The existence of $\beta > 0$ in Step 5 follows from the fact that

$$\varphi(\lambda) = \|y_1 - (\alpha/2)\, v_1 + \lambda s\|$$

is a strictly convex function of λ and $\varphi(0) < 1$, by the strict convexity of $\|$

3.2. ALGORITHM. *Step* 1. Let $y_1 = s/\|s\|$, or more generally, let $y_1 \in K^{\perp} \oplus (s)$, with $\|y_1\| = 1$ and $(y_1 \mid s) > 0$.

Step 2. Find y_1^*.

Step 3. Put $v_1 = Fy_1{}^*$ and do as in Step 3 of Algorithm 3.1 if $\| v_1 \|_2$ is small. If $\| v_1 \|_2$ is not small proceed as follows.

Step 4. Choose $\alpha > 0$ such that

$$\| y_1 - \alpha v_1 \| = 1.$$

Step 5. Put

$$y_2 = \frac{y_1 - (\alpha/2)\, v_1}{\| y_1 - (\alpha/2)\, v_1 \|}.$$

Replace y_1 by y_2 and go to Step 2 onwards carrying out the iteration.

4.

In this section we show that the algorithms of Section 3 converge. Moreover, they have some natural generalizations which we state as theorems. We need a preliminary lemma.

4.1. LEMMA. *Let* $\| \cdot \|$ *be a strictly convex, smooth norm on* $\mathbb{R}^m$. *If* $y_1 \in K^\perp \oplus (s)$ *with* $\| y_1 \| = 1$, *and if* $v_1 = Fy_1{}^* \neq 0$, *then there is a positive* α *such that*

$$\| y_1 - \alpha v_1 \| < 1. \tag{4.1.1}$$

Moreover, if $\{y_1, v_1\}$ *are linearly independent then every* $\alpha > 0$ *satisfying* (4.1.1) *is such that*

$$\| y_1 - \alpha v_1 \| > 0. \tag{4.1.2}$$

Proof. We have

$$(y_1{}^* \mid v_1) = (y_1{}^* \mid Fy_1{}^*) = \| Fy_1{}^* \|_2^2 = \| v_1 \|_2^2, \tag{4.1.3}$$

and

$$\begin{aligned} \| y_1 - \lambda v_1 \| &\geqslant (y_1{}^* \mid y_1 - \lambda v_1) \\ &= 1 - \lambda (y_1{}^* \mid v_1) = 1 - \lambda \| v_1 \|_2^2 . \end{aligned} \tag{4.1.4}$$

From (4.1.4) it is clear that $\| y_1 - \lambda v_1 \| > 1$, whenever $\lambda < 0$.

If $\{y_1, v_1\}$ are linearly dependent there is a scalar α for which $\| y_1 - \alpha v_1 \| = 0$. By the conclusion we just made, $\| y_1 - \lambda v_1 \| > 1$ if $\lambda < 0$. Hence, α satisfying $\| y_1 - \alpha v_1 \| = 0$, must be positive, which shows that $\alpha = 1/\| v_1 \|$.

Hence, suppose that $\{y_1, v_1\}$ are linearly independent. If the lemma in question is false then we should, therefore, have

$$\| y_1 - \lambda v_1 \| \geqslant 1, \qquad \forall \lambda \in \mathbb{R}, \tag{4.1.5}$$

i.e.,

$$\| y_1 - \lambda v_1 \| \geqslant \| y_1 \|, \qquad \forall \lambda \in \mathbb{R}. \tag{4.1.6}$$

By Lemma 2.7, then $(y_1^* \mid v_1) = 0$, which by (4.1.3) implies that $v_1 = 0$; a contradiction.

4.2. Corollary. *If y_1 and v_1 are linearly independent and are as in the lemma then there exists a unique $\alpha > 0$ such that*

$$\| y_1 - \alpha v_1 \| = 1. \tag{4.2.7}$$

Proof. The function $\varphi(\lambda) = \| y_1 - \lambda v_1 \|$ is strictly convex and $\varphi(0) = 1$. By the foregoing lemma there is $\lambda > 0$ such that $\varphi(\lambda) < 1$. Since $\varphi(\lambda) \to \infty$ as $\lambda \to \infty$, there is $\alpha > 0$ such that $\varphi(\alpha) = 1$. Uniqueness of α follows from strict convexity of φ.

Before proceeding further we can utilize the lemma to give an alternate proof of Theorem 2.5. We use the notation of that theorem.

4.3. *Alternate Proof of Theorem* 2.5. If $z^* \notin K$, then $v = Fz^* \neq 0$. Now $\| z \| = 1$, $z \in K^\perp \oplus (s)$ and $(z \mid s) > 0$, since z is a solution of problem (P^*). Note that $(v \mid s) = 0$; whereas, $(z \mid s) > 0$, so that $\{z, v\}$ are linearly independent. By Lemma 4.1, there is $\alpha > 0$ such that

$$0 < \| z - \alpha v \| < 1. \tag{4.3.8}$$

Then

$$w = \frac{z - \alpha v}{\| z - \alpha v \|} \in K^\perp \oplus (s), \qquad \text{and} \qquad \| w \| = 1; \tag{4.3.9}$$

whereas, since $s \perp v$

$$(s \mid w) = \frac{(s \mid z)}{\| z - \alpha v \|} > (s \mid z), \tag{4.3.10}$$

a contradiction to the definition of z, as a solution of problem (P^*). Q.E.D.

4.4. Theorem. *Assume that $\| \cdot \|$ is both strictly convex and smooth. Then the Algorithm* 3.1 *converges to a solution of problem* (P).

Proof. If $v_1 = 0$, then due to the definition of F, $y_1^* \in K$. Also since $y_1 \in K^\perp \oplus (s)$, $\| y_1 \| = 1$, by Theorem 2.2 we conclude that the equation system $Ax = b - (b \mid y_1^*) y_1$ is exactly solvable, yielding a solution of problem (P) with $\|\cdot\|$-error $= (b \mid y_1^*) = (s \mid s)/(s \mid y_1)$. (The last equality follows from Corollary 2.3.)

Denote by y_k, v_k, α_k, β_k the corresponding elements defined at the kth stage of the iteration. By construction $y_k \in K^\perp \oplus (s)$, $k = 1, 2, \ldots$. It would be convenient to set $\beta_0 = (y_1 \mid s)/(s \mid s)$ and only assume that $y_1 \in K^\perp \oplus (s)$, $\| y_1 \| = 1$, in place of the explicit assumption in Step 1 of Algorithm 2.1, where we took $y_1 = s/\| s \|$. We shall let

$$y_0 = y_1 - \beta_0 s. \tag{4.4.1}$$

Then $y_0 \in K^\perp$. Note that $\beta_0 > 0$ and due to the way we defined β_k, $\beta_k > 0$ for every k. We have

$$\sum_{k=0}^{\infty} \beta_k < \infty, \tag{4.4.2}$$

a fortiori, $\beta_k \to 0$. In fact,

$$\sum_{k=0}^{\infty} \beta_k = \frac{1}{\rho}, \tag{4.4.3}$$

where ρ is the minimal $\|\cdot\|$-error of problem (P). This stronger assertion will be evident in the course of proof of this theorem, though we now explicitly establish only the estimate

$$\sum_{k=0}^{\infty} \beta_k \leqslant \frac{1}{\rho}. \tag{4.4.4}$$

By the definition of (y_k),

$$y_k = y_{k-1} - (\alpha_{k-1}/2)\, v_k + \beta_{k-1} s. \tag{4.4.5}$$

Iterating (4.4.5) we get

$$y_k = y_0 - \tfrac{1}{2} \sum_{j=1}^{k-1} \alpha_j v_j + \left(\sum_{j=0}^{k-1} \beta_j \right) s. \tag{4.4.6}$$

Note that $y_0 \in K^\perp$ and $v_j \in K^\perp$ for each j. Moreover, $\| y_k \| = 1$, so that by Lemma 2.6

$$\sum_{j=0}^{k-1} \beta_j \leqslant \frac{1}{\rho}, \tag{4.4.7}$$

which establishes (4.4.4).

We claim that the sequence (v_k) converges to 0. If not, there is a $\delta > 0$ and a subsequence of this sequence, which we denote again by (v_k) such that

$$\| v_k \| \geq \delta, \qquad \forall k \geq 1. \tag{4.4.8}$$

Since

$$\| y_k - \alpha_k v_k \| = 1, \qquad \forall k \geq 1, \tag{4.4.9}$$

we see that

$$1 \geq \alpha_k \| v_k \| - \| y_k \| \geq \delta\alpha_k - 1, \tag{4.4.10}$$

or

$$\alpha_k \leq (2/\delta); \tag{4.4.11}$$

that is, the positive subsequence (α_k) is bounded. Since $\| y_k \| = 1$, for every k, by passing to appropriate subsequence of subsequences, which we again denote by (y_k) and (α_k), we may assume that (y_k) and (α_k) converge, say to y and α, respectively. By the continuity of the map * on $\mathbb{R}^m \backslash \{0\}$ we also have

$$v_k = Fy_k^* \to Fy^* = v, \qquad \text{say}. \tag{4.4.12}$$

But by Step 5 of Algorithm 3.1

$$\| y_k - (\alpha_k/2)\, v_k + \beta_k s \| = 1. \tag{4.4.13}$$

Allowing $k \to \infty$, since $\beta_k \to 0$, we get

$$\| y - (\alpha/2)v \| = 1; \tag{4.4.14}$$

whereas, by (4.4.9)

$$\| y - \alpha v \| = 1. \tag{4.4.15}$$

Since $\| \cdot \|$ is strictly convex and $\| y \| = 1$, (4.4.14) and (4.4.15) show that $\alpha = 0$.

We now show that

$$\| y - \lambda v \| \geq 1, \qquad \forall \lambda \in \mathbb{R}. \tag{4.4.16}$$

This inequality follows from (4.1.4) for $\lambda \leq 0$. If $\lambda > 0$, then there exists j such that $0 < \alpha_k < \lambda$, for all $k \geq j$. The function,

$$\varphi(t) = \| y_k - tv_k \|, \tag{4.4.17}$$

is strictly convex in t and $\varphi(0) = 1 = \varphi(\alpha_k)$. Hence,

$$\| y_k - tv_k \| \geq 1, \qquad \text{if} \quad t \notin (0, \alpha_k), \tag{4.4.18}$$

so that

$$\| y_k - \lambda v_k \| \geq 1, \qquad \forall k \geq j. \tag{4.4.19}$$

Allowing $k \to \infty$ in (4.4.19) we get (4.4.16) for $\lambda > 0$.

But since $\| v \| \geqslant \delta$ and $\| y \| = 1$, (4.4.16) is in contradiction to Lemma 4.1. Hence, we have to conclude that the original sequence (v_k) converges to 0.

Since $\| y_k \| = 1$, for every k, to show (y_k) is convergent, due to compactness of the unit $\| \cdot \|$-sphere, it would be sufficient to show that (y_k) has a unique adherent point. If y is any adherent point of (y_k) so that a subsequence (y_{k_j}) converges to y, then by the foregoing, $v_{k_j} \to 0$. Hence, by Theorem 2.2

$$b - (b \mid y^*) y \in \operatorname{Im}(A),$$

which shows that

$$\rho = d(b, K^\perp) = (b \mid y^*); \tag{4.4.20}$$

$\rho y = (b \mid y^*) y$ is a point in $K^\perp$ nearest to b. Due to strict convexity of $\| \cdot \|$, nearest points are unique; in other words, y is the only adherent point of (y_k).
Q.E.D.

4.5. COROLLARY. *$y = \lim y_k$ is the unique solution of problem (P^*).*

Proof. $y \in K^\perp \oplus (s)$, $\| y \| = 1$ and $y^* \in K$. By (4.4.20) and (2.3.1) we get $(s \mid y) = M$, where M is defined in Theorem 2.4. Q.E.D.

We now state and prove a theorem which is a generalization of the previous theorem and is in itself, therefore, an algorithm for solving problem (P).

4.6. THEOREM. *Let $\| \cdot \|$ be a strictly convex, smooth norm on $\mathbb{R}^m$ and C a nonempty compact subset of the open unit interval $(0, 1)$. Define the sequence (y_k) recursively by requiring y_1 to be any element such that*

$$y_1 \in K^\perp \oplus (s), \qquad \| y_1 \| = 1, \qquad (y_1 \mid s) > 0. \tag{4.6.1}$$

Assume that y_k is defined. Let

$$v_k = F y_k^*, \qquad k \geqslant 1. \tag{4.6.2}$$

Choose $\beta_k > 0$ such that

$$\| y_k - \lambda_k \alpha_k v_k + \beta_k s \| = 1, \tag{4.6.3}$$

where $\alpha_k > 0$ is such that

$$\| y_k - \alpha_k v_k \| = 1 \tag{4.6.4}$$

and $\lambda_k \in C$ is arbitrary. Put

$$y_{k+1} = y_k - \lambda_k \alpha_k v_k + \beta_k s. \tag{4.6.5}$$

Then the sequence (y_k) *converges. Denoting its limit by* y, *any exact solution of the equation system*

$$Ax = b - \frac{(s \mid s)}{(s \mid y)} y, \tag{4.6.6}$$

yields a solution of problem (P).

Moreover, y *is a solution of problem* (P^*).

Proof. Algorithm 3.1 and Theorem 4.4 corresponds to the choice of $C = \{\frac{1}{2}\}$. The proof of Theorem 4.4 carries over to this case with a minor modification.

We have the convergence (4.4.2) as before, and, hence, $\beta_k \to 0$. We show that $v_k \to 0$, by aiming for a contradiction assuming that (v_k) is not convergent to zero. The steps are as in Theorem 4.4, though now we have to pick an additional subsequence of (λ_k) which converges to λ (say). Then $0 < \lambda < 1$. We pass to appropriate subsequences, which will be again denoted by v_k, α_k such that $v_k \to v \neq 0$, and $\alpha_k \to \alpha$. In the limit, from (4.6.3) and (4.6.4) we get

$$\| y - \lambda\alpha v \| = 1 = \| y - \alpha v \|. \tag{4.6.7}$$

Since $\| y \| = 1$, $0 < \lambda < 1$ and $\| \cdot \|$ is strictly convex, (4.6.7) implies that $\alpha = 0$.

Now as in Theorem 4.4 we show that $v_k \to 0$ and then complete the proof exactly as in that theorem.

That y is a solution of problem (P^*) was shown in Corollary 4.5. Q.E.D.

The next theorem shows that Algorithm 3.2 is convergent. In fact, the following theorem is a generalization of Algorithm 3.2 and is in the same spirit as Theorem 4.6, which is a generalization of Algorithm 3.1.

4.7. THEOREM. *Let* $\| \cdot \|$ *be a smooth and strictly convex norm on* $\mathbb{R}^m$ *and* C *a nonempty compact subset of the open unit interval* $(0, 1)$. *Define the sequence* (y_k) *recursively by choosing* y_1 *such that*

$$y_1 \in K^\perp \ominus (s), \qquad \| y_1 \| = 1, \qquad (y_1 \mid s) > 0. \tag{4.7.1}$$

Assume y_k *is defined. Let*

$$v_k = F{y_k}^*, \qquad k \geqslant 1. \tag{4.7.2}$$

Define y_{k+1} *by*

$$y_{k+1} = \frac{y_k - \lambda_k \alpha_k v_k}{\| y_k - \lambda_k \alpha_k v_k \|}, \tag{4.7.3}$$

where $\alpha_k > 0$ *is chosen such that*

$$\|y_k - \alpha_k v_k\| = 1, \tag{4.7.4}$$

and $\lambda_k \in C$ *is arbitrary. Then the sequence* (y_k) *converges. Denoting* $y = \lim y_k$, *a solution of problem* (P) *is given by any exact solution of*

$$Ax - b = \frac{(s \mid s)}{(s \mid y)}\, y. \tag{4.7.5}$$

Moreover, y *is the unique solution of problem* (P^*).

Proof. By the construction of y_k we see that $y_k \in K \cap (s)$ and $\|y_k\| = 1$. Note that

$$(v_k \mid s) = (Fy_k^* \mid s) = (y_k^* \mid Fs) = 0. \tag{4.7.6}$$

Due to the strict convexity of $\|\ \|$,

$$0 < \|y_k - \lambda_k \alpha_k v_k\| < 1. \tag{4.7.7}$$

By (4.7.3), (4.7.6), and (4.7.7) we get

$$(y_{k+1} \mid s) = \frac{(y_k \mid s)}{\|y_k - \lambda_k \alpha_k v_k\|} > (y_k \mid s). \tag{4.7.8}$$

This shows that $((y_k \mid s))$ is a strictly increasing sequence, which is clearly bounded by M (where M was defined in Theorem 2.4). Hence,

$$\frac{(y_k \mid s)}{(y_{k+1} \mid s)} \to 1, \tag{4.7.9}$$

which by (4.7.7) shows that

$$\lim_k \|y_k - \lambda_k \alpha_k v_k\| = 1. \tag{4.7.10}$$

Proceeding as in Theorem 4.6 we get the relations (4.6.7) from (4.7.10) and (4.7.4). The rest of the proof proceeds as that of Theorem 4.6.

Remark. Obviously, when $C = \{\frac{1}{2}\}$ in the foregoing theorem we get Algorithm 3.2.

References

1. M. M. Day, "Normed Linear Spaces," Springer Verlag, New York, 1958.
2. A. S. Householder, "The Theory of Matrices in Numerical Analysis," Blaisdell Publishing Company, New York, 1964.

3. J. R. RICE AND K. H. USOW, The Lawson algorithm and extensions, *Math. Comp.* **22** (1968), 118–127.
4. V. P. SREEDHARAN, Solutions of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.* **13** (1969), 146–151.
5. V. P. SREEDHARAN, Least-squares algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.* **17** (1971), 387–401.