JOURNAL OF APPROXIMATION THEORY 8, 133-141 (1973)

On an Algorithm for Best Approximate Solutions to $Av = b$ in Normed Linear Spaces

HOWARD ANTON AND C. S. DURIS

Department of Mathematics, Drexel University, Philadelphia, PA. 19104

Communicated by E. W. Cheney

Received March 1, 1971

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

INTRODUCTION

Let V be a real linear space, X a real normed linear space, $A : V \to X$ a linear transformation, and b a fixed vector in X. We assume that the equation

$$Av = b$$

has no solution and that it is of interest to determine a vector $\hat{v} \in V$ such that

$$\| b - A\hat{v} \| \leqslant \| b - Av \|$$

for all $v \in V$. We shall call this the *primal problem*. In absence of further assumptions, the primal problem need not have a solution; and in cases where a solution exists, it need not be unique. An argument of James given in Phelps [4] shows that a space S is reflexive if and only if for each closed subspace E and each $x \in S$, there is a vector $y \in E$ such that $\| x - y \| \leqslant \| x - z \|$ for all $z \in E$. Thus, when A has a closed range and X is reflexive, the primal problem has at least one solution. It is known [2] that a Banach space S is rotund (strictly convex) and reflexive if and only if for each closed, convex set $K \subset S$ and each $x \in S$, there is a unique vector $y \in K$ such that $\| x - y \| \leqslant \| x - z \|$ for all $z \in K$. It follows that if A is 1-1 with a closed range and X is rotund and reflexive, then the primal problem has a unique solution. We shall assume throughout this paper that these conditions are satisfied.

An important special case of the primal problem occurs when $V = R^n$, $X = R^m$ and A is an $m \times n$ matrix for which $m > n$ and rank $(A) = n$. If X has the l^p norm, where $1 < p < \infty$, then the primal problem becomes the l^p-problem for overdetermined systems of linear equations (see [5, 6]).

In Section 1 a dual problem is defined whose solution can be used to solve the primal problem. The remaining sections develop an algorithm for solving this dual problem under the assumption that the range of A has finite codimension. This restrictive condition is satisfied by many operators of interest. In particular, if U is a compact operator on a Banach space, then $I - U$ has a range with finite codimension. More generally, any Fredholm operator has this property (see [3]). Many of the ideas in this paper were motivated by the work of V. P. Sreedharan appearing in [6].

1. The Dual Problem

Throughout this paper we shall adhere to the following notation. X^* denotes the dual space of X, S designates the set $\{x \in X : \|x\| = 1\}$, and S^* is the corresponding set in X^*. If $f \in X^*$ and $x \in X$, then $\langle x, f\rangle$ denotes $f(x)$. The range of A will be written as $R(A)$, and for $M \subset X$ we will designate the set $\{f \in X^* : \langle x, f\rangle = 0$ for all $x \in M\}$ by $M^\perp$. Similarly, if $M \subset X^*$ then $M^\perp = \{x \in X : \langle x, f\rangle = 0$ for all $f \in M\}$.

By the *dual problem* we will mean: Determine $\hat{f} \in R(A)^\perp \cap S^*$ such that $\langle b, \hat{f}\rangle \geqslant \langle b, f\rangle$ for all $f \in R(A)^\perp \cap S^*$. It is known that the dual problem always has a solution and that

$$\inf\{\|b - Av\| : v \in V\} = \max\{\langle b, f\rangle : f \in R(A)^\perp \cap S^*\}$$

(see, e.g., [1]). This common value will be denoted by ρ and since $Av = b$ has no solution, we have $\rho > 0$. In the event that X^* is rotund, the dual problem has a unique solution.

Definition 1. *If $f \in X^* - \{0\}$, then $x \in X \cap S$ is called a dual vector for f* if $\langle x, f\rangle = \|f\|$.

Definition 2. If $x \in X - \{0\}$, then $f \in X^* \cap S^*$ is called a *dual functional for x* if $\langle x, f\rangle = \|x\|$.

Theorem 1.1.

(i) *If X is reflexive and $f \in X^* - \{0\}$, then there is at least one dual vector for f. If, in addition, X is rotund, then the dual vector is unique.*

(ii) *If X is reflexive and X^* is rotund, then for each $x \in X - \{0\}$, there is a unique dual functional for x.*

Proof. These are standard results and can be found, e.g., in [7].

When a unique dual vector for $f \in X^* - \{0\}$ exists, we denote it by f^*. Similarly, x^* denotes the unique dual functional for $x \in X - \{0\}$ when it exists.

THEOREM 1.2. *Let X be rotund and reflexive, and suppose that $A : V \to X$ is 1–1 and has a closed range. If $\hat{f}$ is any solution to the dual problem, then*

(i) *$Av = b - \langle b, \hat{f} \rangle \hat{f}^*$ has a unique solution for v.*

(ii) *This solution is the solution to the primal problem.*

Proof. (*i*) We show first that $b - \langle b, \hat{f} \rangle \hat{f}^* \in R(A)$. Let $g \in R(A)^\perp$, let $\hat{v}$ be the solution to the primal problem, and let

$$w = \frac{b - A\hat{v}}{\|b - A\hat{v}\|}.$$

Then

$$\begin{aligned} \langle b - \langle b, \hat{f} \rangle \hat{f}^*, g \rangle &= \langle b, g \rangle - \langle b, \hat{f} \rangle \langle \hat{f}^*, g \rangle \\ &= \langle b - A\hat{v}, g \rangle - \langle b, \hat{f} \rangle \langle \hat{f}^*, g \rangle \\ &= \rho \langle w, g \rangle - \rho \langle \hat{f}^*, g \rangle \\ &= \rho \langle w - \hat{f}^*, g \rangle \end{aligned}$$

Letting $g = \hat{f}$ gives

$$0 = \langle b - \langle b, \hat{f} \rangle \hat{f}^*, \hat{f} \rangle = \rho \langle w - \hat{f}^*, \hat{f} \rangle.$$

Therefore

$$\langle w, \hat{f} \rangle = \langle \hat{f}^*, \hat{f} \rangle = \|\hat{f}\|.$$

By uniqueness of the dual vector for $\hat{f}$, we obtain $w = \hat{f}^*$, so that

$$\langle b - \langle b, \hat{f} \rangle \hat{f}^*, g \rangle = 0 \qquad \text{for all } g \in R(A)^\perp$$

and consequently $b - \langle b, \hat{f} \rangle \hat{f}^* \in R(A)$. The uniqueness will follow from the uniqueness of solutions to the primal problem once part (ii) is proved.

(*ii*) If v_0 is any solution to $Av = b - \langle b, \hat{f} \rangle \hat{f}^*$,

$$\|Av_0 - b\| = \|\langle b, \hat{f} \rangle \hat{f}^*\| = |\langle b, \hat{f} \rangle| = \rho$$

so that v_0 is a solution to the primal problem.

2. Statement of the Algorithm

It follows from Theorem 1.2 that once a solution $\hat{f}$ to the dual problem is found, the solution to the primal problem can be obtained by computing

$$A^{-1}(b - \langle b, \hat{f}\rangle \hat{f}^*).$$

In this section we give an algorithm for solving the dual problem under the assumption that $R(A)$ has finite codimension, and both X and X^* are uniformly rotund (uniformly convex). This algorithm is, in general, infinite and generates a sequence (f_i) which converges strongly to the solution of the dual problem.

Definition 2.1. If W is any finite dimensional subspace of X^* having basis $B = \{g_1, g_2, ..., g_n\}$, then the *projection of X onto W relative to B* is the linear transformation $F : X \to W$ defined by

$$F(x) = \sum_{i=1}^{n} \langle x, g_i\rangle g_i .$$

Note that $F(x) = 0$ if and only if $x \in W^{\perp}$. Also, in the case where X is a Hilbert space (so that X can be identified with X^*) and $\{g_1, ..., g_n\}$ is an orthonormal basis for W, the mapping F is the orthogonal projection of X onto W and does not depend on the orthonormal basis chosen.

We are now in a position to state the algorithm. Both the proof of convergence and the verification that the steps in the algorithm can always be carried out will be established in the next section.

Theorem 2.1. *If X and X^* are uniformly rotund and $R(A)$ has finite codimension, then the following algorithm yields a solution to the dual problem in a finite number of steps or else generates a sequence (f_i) which converges strongly to the solution of the dual problem.*

Step (0): *Select a fixed basis $B = \{g_1, g_2, ..., g_n\}$ for $W = \{R(A) \cup b\}^{\perp}$ and let F be the projection of X onto W relative to B.*

Step (1): *Choose $f_0 \in R(A)^{\perp}$ such that $\|f_0\| = 1$ and $\langle b, f_0\rangle > 0$.*

Step (2): *Set $i = 0$*

Step (3): *Compute $h_i = F(f_i^*)$*

Step (4): *If $h_i = 0$, then stop since f_i is the solution of the dual problem. If $h_i \neq 0$, go to Step* (5).

Step (5): *Determine* α_i *such that* $\|f_i - \alpha_i h_i\| \leqslant \|f_i - \lambda h_i\|$ *for all* λ.

Step (6): *Set* $f_{i+1} = (f_i - \alpha_i h_i)/\|f_i - \alpha_i h_i\|$, *increase* i *by* 1, *and return to* *Step* (3).

3. Convergence of the Algorithm

Throughout this section we assume that X and X^* are uniformly rotund and $R(A)$ has finite codimension. (We note that some relaxation of the rotundity conditions is possible in some of the lemmas of this section although we have no need for the added generality here.)

Lemma 3.1. *If* $f \in R(A)^\perp \cap S^*$, $\langle b,f\rangle > 0$, *and* $F: X \to \{R(A) \cup b\}^\perp$ *is a projection with respect to any basis* B *for* $\{R(A) \cup b\}^\perp$, *then* f *is a solution of the dual problem if and only if* $F(f^*) = 0$.

Proof. If $F(f^*) = 0$, then $f^* \in \text{lin}\{R(A) \cup b\}$ so that $f^* = Av + \alpha b$ for some $v \in V$ and real number α. Therefore,

$$1 = \langle f^*, f\rangle = \langle Av + \alpha b, f\rangle = \alpha\langle b, f\rangle.$$

Thus,

$$f^* = Av + b/\langle b, f\rangle.$$

For every $g \in R(A)^\perp \cap S^*$ we have

$$\begin{aligned} 1 = \|f^*\| \|g\| \geqslant |\langle f^*, g\rangle| &= |\langle Av + (b/\langle b,f\rangle), g\rangle| \\ &= |\langle b, g\rangle|/\langle b, f\rangle. \end{aligned}$$

Therefore, $\langle b,f\rangle \geqslant |\langle b, g\rangle|$ for any $g \in R(A)^\perp \cap S^*$ so that f is a solution of the dual problem.

Conversely, assume $\hat{f}$ is the solution of the dual problem. By Theorem 1.2, there is a $\hat{v} \in V$ such that

$$A\hat{v} = b - \langle b, \hat{f}\rangle \hat{f}^*.$$

It follows that

$$0 = F(A\hat{v}) = F(b) - \langle b, f\rangle F(f^*) = \langle b, f\rangle (Ff^*)$$

so that $F(f^*) = 0$.

Lemma 3.2. *If* f *and* h *are linearly independent functionals in* X^*, *then* $\|f\| \leqslant \|f - \lambda h\|$ *for all real* λ *if and only if* $\langle f^*, h\rangle = 0$.

Proof. Assume $\langle f^*, h\rangle = 0$. Then

$$\|f\| = \langle f^*, f\rangle = \langle f^*, f - \lambda h\rangle \leq \|f - \lambda h\|$$

for all real λ. Conversely, assume that for all real λ we have $\|f\| \leq \|f - \lambda h\|$. It follows from the Hahn–Banach theorem that there is a linear functional $L \in X^{**}$ such that $L(h) = 0$, $L(f) = \|f\|$ and $\|L\| = 1$. Since uniform rotundity implies reflexivity, there is a vector $y \in X$ such that $L(\cdot) = \langle y, \cdot\rangle$. Therefore $\langle y, f\rangle = \|f\|$ and $\|y\| = 1$. By uniqueness of the dual vector for f, $y = f^*$. Thus $0 = L(h) = \langle f^*, h\rangle$.

LEMMA 3.3. *Assume $f \in R(A)^{\perp} \cap S^*$, $\langle b, f\rangle > 0$ and $F: X \to \{R(A) \cup b\}^{\perp}$ is a projection with respect to any basis B for $\{R(A) \cup b\}^{\perp}$. If $h = F(f^*) \neq 0$, then there is a unique real number $\alpha > 0$ such that*

$$0 < \|f - \alpha h\| = \min_{\lambda} \|f - \lambda h\| < 1.$$

Further, α is the unique solution of $\langle (f - \alpha h)^*, h\rangle = 0$.

Proof. The existence of α is clear and the uniqueness follows from the rotundity of X^*. We show next that h and f are linearly independent. If $k_1 f + k_2 h = 0$ then

$$0 = \langle b, k_1 f + k_2 h\rangle = k_1 \langle b, f\rangle.$$

The independence is a consequence of the facts $\langle b, f\rangle > 0$ and $h = F(f^*) \neq 0$. It now follows that

$$0 < \|f - \alpha h\| = \min_{\lambda} \|f - \lambda h\| \leq \|f\| = 1.$$

We show next that $\|f - \alpha h\| < 1$. If $1 = \|f\| \leq \|f - \lambda h\|$ for all λ, it follows from Lemma 3.2 that $\langle f^*, h\rangle = 0$.
Therefore,

$$0 = \langle f^*, F(f^*)\rangle = \sum_{i=1}^{n} \langle f^*, g_i\rangle^2.$$

This implies

$$h = F(f^*) = \sum_{i=1}^{n} \langle f^*, g_i\rangle\, g_i = 0$$

which is a contradiction, so that $\|f - \alpha h\| < 1$.

Since $\langle f^*, h\rangle > 0$ it follows that the minimizing α must be positive since

$$\begin{aligned}\|f - \alpha h\| &= \|f^*\|\,\|f - \alpha h\| \\ &\geq \langle f^*, f - \alpha h\rangle \\ &= 1 - \alpha\langle f^*, h\rangle.\end{aligned}$$

Further, it follows from Lemma 3.2 and the fact $\|f - \alpha h\| \leq \|(f - \alpha h) - \lambda h\|$ for all λ that α is the unique solution of $\langle (f - \alpha h)^*, h\rangle = 0$.

LEMMA 3.4. *If (f_i) is any sequence in S^* such that $\|f_i - f\| \to 0$, then $\|f_i^* - f^*\| \to 0$.*

Proof. We show first that $f_i^* \to f^*$ weakly. Due to the weak sequential compactness of the closed unit ball in X, it suffices to show that every weakly convergent subsequence of (f_i^*) converges weakly to f^*. Let $(f_{i_m}^*)$ be any weakly convergent subsequence of (f_i^*) and assume $(f_{i_m}^*)$ converges weakly to z. We have

$$0 \leq \langle z, f\rangle - 1 \leq |\langle z, f\rangle - \langle f_{i_m}^*, f\rangle| + |\langle f_{i_m}^*, f\rangle - \langle f_{i_m}^*, f_{i_m}\rangle|$$

$$\leq |\langle z, f\rangle - \langle f_{i_m}^*, f\rangle| + \|f - f_{i_m}\|.$$

Taking the limit as $m \to \infty$ yields

$$\langle z, f\rangle = 1.$$

But $f_{i_m}^* \xrightarrow{w} z$ implies $\|z\| \leq 1$ so that

$$1 \geq \langle (z/\|z\|), f\rangle = (1/\|z\|) \geq 1.$$

Thus $\|z\| = 1$, and by uniqueness of the dual vector, $z = f^*$.

Since $\|f_i^*\| \to \|f^*\|$ and $f_i^* \xrightarrow{w} f^*$, it follows from the uniform rotundity of X that $\|f_i^* - f^*\| \to 0$ (see, e.g. [7, p. 111, problem 9]).

We are now in a position to prove the main result, Theorem 2.1.

Proof of Theorem 2.1.

Case 1 (The algorithm does not terminate). Let $\rho_i = \langle b, f_i\rangle$. Since $f_i \in R(A)^\perp \cap S^*$, $i = 1, 2, \ldots$ we have

$$\rho_i = \langle b, f_i\rangle = \langle b - Av, f_i\rangle \leq \|b - Av\| \leq \rho$$

for all $v \in V$. From the Lemma 3.3 and the fact that $\rho_0 = \langle b, f_0\rangle > 0$, we have

$$\begin{aligned}\rho_{i+1} - \rho_i &= \langle b, f_{i+1}\rangle - \langle b, f_i\rangle \\ &= ((1/\|f_i - \alpha_i h_i\|) - 1)\langle b, f_i\rangle \\ &\geq 0.\end{aligned}$$

Since (ρ_i) is a monotone sequence bounded above by ρ, there is a real number σ such that

$$\lim_{i\to\infty} \rho_i = \sigma \leq \rho.$$

We now show that $\sigma = \rho$. Assume $\sigma < \rho$.

Since (f_i) is a bounded sequence in the finite dimensional subspace $R(A)^{\perp}$, we can pick a subsequence (f_{m_i}) converging strongly to a functional $f \in X^*$. Since F is a continuous linear transformation, we have by Lemma 3.4 that (h_{m_i}) converges strongly to $h = F(f^*)$. If $h = 0$, then by Lemma 3.1, f is a solution of the dual problem so that

$$\sigma = \lim_{i\to\infty} \rho_{m_i} = \lim_{i\to\infty} \langle b, f_{m_i} \rangle = \langle b, f \rangle = \rho$$

which is a contradiction. If $h \neq 0$, let $\| h \| = d > 0$. For i sufficiently large we have $\| h_{m_i} \| > d/2$. Therefore,

$$1 \geq \| f_{m_i} - \alpha_{m_i} h_{m_i} \| \geq \alpha_{m_i} \| h_{m_i} \| - 1$$

so that

$$0 < \alpha_{m_i} < 4/d$$

for i sufficiently large. By passing to a subsequence, if necessary, we can assume α_{m_i} converges to a real number $\bar{\alpha}$. Since

$$\| f_{m_i} - \alpha_{m_i} h_{m_i} \| \leq \| f_{m_i} - \lambda h_{m_i} \|$$

for all λ and

$$\rho_{m_i} / \rho_{m_i+1} = \| f_{m_i} - \alpha_{m_i} h_{m_i} \|$$

we obtain

$$\| f \| = 1 = \lim_{i\to\infty} \| f_{m_i} - \alpha_{m_i} h_{m_i} \| = \| f - \bar{\alpha} h \| \leq \| f - \lambda h \|$$

for all λ. Since $h \neq 0$, f and h are linearly independent so that $\langle f^*, h \rangle = 0$ by Lemma 3.2. It follows that $h = 0$, which is a contradiction. Thus $\sigma = \rho$.

Suppose now that (f_{m_i}) is any strongly convergent subsequence of (f_i), and let $f = \lim_{i\to\infty} f_{m_i}$. Since $\| f \| = 1$ and $\langle b, f \rangle = \rho$, it follows that f is the unique solution $\hat{f}$ to the dual problem. Since $\hat{f}$ is the only strong cluster point for the bounded sequence (f_i), we have $\lim_{i\to\infty} \| f_i - \hat{f} \| = 0$.

Case II (The algorithm terminates). Suppose the algorithm terminates at the ith step, i.e., $F(f_i^*) = 0$. As in Case I,

$$0 < \rho_0 < \rho_1 < \cdots < \rho_i = \langle b, f_i \rangle.$$

By Lemma 3.1, f_i is a solution of the dual problem.

We conclude by noting that the functional f_0 required to start the algorithm can be obtained by selecting any basis for $R(A)^{\perp}$ (an extension of the basis B in Step (0) is the logical choice) and projecting f_0^* onto $R(A)^{\perp}$ relative to this basis.

References

1. R. C. Buck, Applications of duality in approximation theory, Approximation of Functions, pp. 27–42, Elsevier, Amsterdam, 1965.
2. D. F. Cudia, Rotundity, pp. 73–97 in Convexity, Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society, Providence, Rhode Island, 1963.
3. S. Lang, "Analysis II," Addison Wesley, Reading, MA, 1969.
4. R. R. Phelps, Uniqueness of Hahn–Banach extensions and unique best approximation, *Trans. Amer. Math. Soc.* **95** (1960), 238–255.
5. J. R. Rice and K. H. Usow, The Lawson algorithm and extensions, *Math. Comp.*, **22** (1968), 118–127.
6. V. P. Sreedharan, Least square algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.* **17** (1971), 387–401.
7. A. Wilansky, "Functional Analysis," Blaisdell, New York, 1964.