Journal of Approximation Theory **117**, 179–184 (2002)
doi:10.1006/jath.2002.3685

NOTE

Dykstra's Algorithm and a Representation of the Moore–Penrose Inverse

C. W. Groetsch

Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221-0025, USA
E-mail: groetsch@uc.edu

Communicated by Frank Deutsch

Received July 30, 2001; accepted January 9, 2002

The convergence of Lardy's series representation of the Moore–Penrose inverse of a closed unbounded linear operator is proved via Dykstra's alternating projection algorithm. © 2002 Elsevier Science (USA)

1. INTRODUCTION

In [3] Lardy used spectral techniques to prove the convergence of a series representation of the Moore–Penrose inverse of a closed linear operator A that is defined on a dense subspace $\mathscr{D}(A)$ of a Hilbert space H_1 and takes values in a Hilbert space H_2. (In this discussion the symbols $\langle\cdot,\cdot\rangle$, $\|\cdot\|$, and I are used indiscriminately to denote the inner product, norm, and identity operator, respectively, in either Hilbert space.) The Moore–Penrose inverse of A is the operator $A^\dagger$ defined on the dense subspace $\mathscr{D}(A^\dagger) = R(A) + R(A)^\perp$ of H_2 which maps $y \in \mathscr{D}(A^\dagger)$ to the unique vector $x = A^\dagger y \in \mathscr{D}(A) \cap N(A)^\perp$ satisfying $Ax = Py$, where P is the orthogonal projector of H_2 onto $\overline{R(A)}$, the closure of the range of A ($N(A)$ is the nullspace of A). In particular, $\mathscr{D}(A^\dagger)$ consists of those $y \in H_2$ for which $Py \in R(A)$. The vectors $x \in \mathscr{D}(A)$ satisfying $Ax = Py$ are called least-squares solutions of the equation $Az = y$ since any such least-squares solution x satisfies $\|Ax - y\| \leqslant \|Az - y\|$ for any $z \in \mathscr{D}(A)$. Therefore, $y \in \mathscr{D}(A^\dagger)$ if and only if the equation $Az = y$ has least-squares solutions and $A^\dagger y$ is that least-squares solution having smallest norm. It is well known that $A^\dagger : \mathscr{D}(A^\dagger) \to H_1$ is itself a closed densely defined linear operator which is bounded if and only if $R(A)$ is closed.

0021-9045/02 $35.00

Lardy's series representation of the Moore–Penrose inverse makes use of the remarkable theorem of von Neumann (see [4]) which asserts that the linear operators

$$\hat{A} = (I + AA^*)^{-1}, \qquad \check{A} = (I + A^*A)^{-1}, \qquad A^*\hat{A} \quad \text{and} \quad A\check{A}$$

are each *bounded* linear operators that are defined *everywhere* on the appropriate Hilbert spaces. Lardy's theorem asserts that

$$A^\dagger y = \sum_{k=1}^{\infty} A^*\hat{A}^k y \tag{1}$$

for each $y \in \mathscr{D}(A^\dagger)$. As an alternative to his spectral theory approach, we apply Dykstra's algorithm (see [1, p. 207]) to prove the convergence of this representation of the Moore–Penrose inverse. Our line of argument is this: the series representation is converted into an iterative method; the iterative method is characterized as a multi-stage optimization procedure; and the optimization procedure is interpreted as alternately projecting onto two closed affine subsets of a product Hilbert space. The validity of representation (1) is then an immediate consequence of the convergence of Dykstra's algorithm.

2. AN ITERATIVE METHOD

We begin by proving a simple identity that relates the bounded operators $A^*\hat{A}$, $\hat{A}$ and $\check{A}$ (see also [2, Lemma 2.2]).

Lemma. $(A^*\hat{A})\hat{A} = \check{A}(A^*\hat{A})$.

Proof. Note the operators indicated in the statement of the lemma are defined everywhere. Given $y \in H_2$, let $z = A^*\hat{A}\hat{A}y$, and note that $z \in \mathscr{D}(A)$ since $R(\hat{A}) \subseteq \mathscr{D}(AA^*)$. We then have

$$Az = (-I + I + AA^*)\hat{A}\hat{A}y = -\hat{A}\hat{A}y + \hat{A}y$$

and hence, since the right-hand side is in $\mathscr{D}(AA^*) \subseteq \mathscr{D}(A^*)$, $Az \in \mathscr{D}(A^*)$, and further

$$(I + A^*A)z = A^*\hat{A}\hat{A}y - A^*\hat{A}\hat{A}y + A^*\hat{A}y.$$

Therefore, $z = \check{A}A^*\hat{A}y$. ∎

Let $x_0 = 0$ and $x_n = \sum_{k=1}^{n} A^*\hat{A}^k y$. Then, by the lemma,

$$x_{n+1} = A^*\hat{A}y + \sum_{k=1}^{n} A^*\hat{A}\hat{A}^k y = A^*\hat{A}y + \check{A}x_n \tag{2}$$

and hence the partial sums of the series representation (1) satisfy iterative relation (2).

The iterate x_{n+1} is characterized as follows:

PROPOSITION 1. *Let $x_0 = 0$. Then x_{n+1} is the unique solution $z \in \mathscr{D}(A)$ of the equation*

$$A^*(Az - y) + z = x_n, \qquad n = 0, 1, 2, \dots \,.$$

Proof. Since $R(\check{A}) \subseteq \mathscr{D}(A^*A)$ and $R(\hat{A}) \subseteq \mathscr{D}(AA^*)$, we see that

$$x_{n+1} = A^*\hat{A}y + \check{A}x_n \in \mathscr{D}(A).$$

Also,

$$Ax_{n+1} = AA^*\hat{A}y + A\check{A}x_n = -\hat{A}y + y + A\check{A}x_n.$$

Therefore,

$$Ax_{n+1} - y = -\hat{A}y + A\check{A}x_n \in \mathscr{D}(A^*)$$

and

$$A^*(Ax_{n+1} - y) = -A^*\hat{A}y + A^*A\check{A}x_n = -A^*\hat{A}y - \check{A}x_n + x_n = -x_{n+1} + x_n,$$

that is,

$$A^*(Ax_{n+1} - y) + x_{n+1} = x_n.$$

If $z_1, z_2 \in \mathscr{D}(A)$ satisfy the equation of the proposition, then $w = z_1 - z_2$ satisfies $A^*Aw + w = 0$ and hence $w = 0$ since $(I + A^*A)$ is invertible. ∎

3. MULTI-STAGE OPTIMIZATION

The iterative process of the previous section may be viewed as a multi-stage optimization procedure in the product Hilbert space $\mathscr{H} = H_1 \times H_2$ (endowed with the usual product norm and inner product). Since A is a closed linear operator, the graph

$$\mathscr{G} = \{(x, Ax) \colon x \in \mathscr{D}(A)\}$$

is a closed subspace of $\mathscr{H}$. Given a vector $(x_n, y) \in \mathscr{H}$, let (x_{n+1}, Ax_{n+1}) be the metric projection (in $\mathscr{H}$) of (x_n, y) onto $\mathscr{G}$. The vector x_{n+1} is then the unique vector z in $\mathscr{D}(A)$ that minimizes the quantity

$$\|z - x_n\|^2 + \|Az - y\|^2$$

and hence

$$\langle Ax_{n+1} - y, Au \rangle = \langle x_n - x_{n+1}, u \rangle$$

for all $u \in \mathscr{D}(A)$. Therefore, $Ax_{n+1} - y \in \mathscr{D}(A^*)$ and

$$A^*(Ax_{n+1} - y) = x_n - x_{n+1}. \tag{3}$$

The result of the previous section therefore characterizes the iterates which form the partial sums of the series representation (1) as the unique solutions of the multi-stage optimization process

$$x_{n+1} = \operatorname{argmin}\{\|Az - y\|^2 + \|z - x_n\|^2 \colon z \in \mathscr{D}(A)\}. \tag{4}$$

4. DYKSTRA'S ALGORITHM APPLIED

Let $K_1 = H_1 \times \{Py\}$, where P is the orthogonal projector of H_2 onto $\overline{R(A)}$, and let $K_2 = \mathscr{G}$, the graph of A. Then K_1 and K_2 are closed affine subsets of $\mathscr{H}$ and $K_1 \cap K_2 \neq \emptyset$, if and only if there is a $x \in \mathscr{D}(A)$ with $Ax = Py$, that is, if and only if $y \in \mathscr{D}(A^\dagger)$. To put it another way,

$$K_1 \cap K_2 = L_y \times \{Py\},$$

where L_y is the set of least-squares solutions of the equation $Az = y$.

PROPOSITION 2. *If $y \in \mathscr{D}(A^\dagger)$, then $x_n \to A^\dagger y$, as $n \to \infty$, where $\{x_n\}$ is defined by* (4) (*equivalently* (3) *or* (2)).

Proof. Let P_i be the metric projection of $\mathscr{H}$ onto K_i, $i = 1, 2$. Note that since

$$\|Az - y\| = \|Az - Py\|,$$

$P_2(u, v) = P_2(u, Pv)$ for any $(u, v) \in \mathscr{H}$. The variational characterization (4) of x_1 gives

$$(x_1, Ax_1) = P_2(0, y) = P_2(0, Py) = P_2P_1(0, y).$$

Also, $P_1(x_1, Ax_1) = (x_1, Py)$ and hence,

$$P_2P_1(x_1, Ax_1) = P_2(x_1, Py) = P_2(x_1, y) = (x_2, Ax_2)$$

and, in general,

$$(x_n, Ax_n) = (P_2P_1)^n(0, y) \tag{5}$$

and hence, by Dykstra's theorem ([1, p. 216]),

$$(x_n, Ax_n) \to P_{K_1 \cap K_2}(0, y) \qquad \text{as} \quad n \to \infty,$$

where $P_{K_1 \cap K_2}$ is the metric projector of $\mathscr{H}$ onto $L_y \times \{Py\}$. In particular, $x_n \to x$ where x is the least-squares solution nearest to 0. In other words, $x = A^\dagger y$, the least-squares solution of smallest norm. ∎

Representation (1) of the Moore–Penrose inverse requires that iterative method (2) starts at $x_0 = 0$. However, iterative method (2) is well-defined for an arbitrary initial approximation x_0. In the case of an arbitrary x_0, process (5) becomes

$$(x_n, Ax_n) = (P_2P_1)^n(x_0, y), \qquad n = 1, 2, 3, \ldots .$$

As above, Dykstra's result assures that x_n converges to the least-squares solution x which is nearest to the initial approximation x_0. A priori information on the desired least-squares solution, in the form of x_0, may therefore be allowed to influence the particular least-squares solution to which the Dykstra algorithm converges.

The formulation of iterative method (2) in terms of Dykstra's algorithm in the product space $\mathscr{H}$ also provides a justification of the *regularity* (in the sense of Tikhonov and Arsenin [5]) of the method. Suppose that, instead of the exact data $y \in \mathscr{D}(A^\dagger)$, only an approximation $y^\delta \in H_2$ is available satisfying $\|y - y^\delta\| \leqslant \delta$. Suppose x_n^δ is defined by (2) using y^δ instead of y (and $x_0^\delta = x_0$). Also, let $K_1^\delta = H_1 \times \{Py^\delta\}$ where, as before, P is the orthogonal projector of H_2 onto $\overline{R(A)}$. We then have

$$(x_n^\delta, Ax_n^\delta) = (P_2P_1^\delta)(x_{n-1}^\delta, Ax_{n-1}^\delta) = (P_2P_1^\delta)^n(x_0, y^\delta),$$

where P_1^δ is the metric projector of $\mathscr{H}$ onto K_1^δ. Finally, let $|\cdot|$ be the norm on $\mathscr{H}$, that is,

$$|(u, v)|^2 = \|u\|^2 + \|v\|^2 \qquad \text{for} \quad (u, v) \in \mathscr{H}.$$

One then has

$$
\begin{aligned}
\|x_n^\delta - x_n\|^2 &\leqslant |(x_n^\delta, Ax_n^\delta) - (x_n, Ax_n)|^2 \\
&= |P_2\{P_1^\delta(x_{n-1}^\delta, Ax_{n-1}^\delta) - P_1(x_{n-1}, Ax_{n-1})\}|^2 \\
&\leqslant |P_1^\delta(x_{n-1}^\delta, Ax_{n-1}^\delta) - P_1(x_{n-1}, Ax_{n-1})|^2 \\
&= |(x_{n-1}^\delta, Py^\delta) - (x_{n-1}, Py)|^2 \\
&\leqslant \|x_{n-1}^\delta - x_{n-1}\|^2 + \delta^2.
\end{aligned}
$$

Since $x_0^\delta = x_0$, it follows that $\|x_n^\delta - x_n\| \leqslant \sqrt{n}\delta$. Combining this with the previous proposition we arrive at:

PROPOSITION 3. *Suppose* $y \in \mathscr{D}(A^\dagger)$ *and* $\|y - y^\delta\| \leqslant \delta$. *If* $n = n(\delta)$ *satisfies* $n(\delta) \to \infty$ *and* $\sqrt{n(\delta)}\delta \to 0$ *as* $\delta \to 0$, *then* $x_{n(\delta)}^\delta \to A^\dagger y$ *as* $\delta \to 0$.

REFERENCES

1. F. Deutsch, "Best Approximation in Inner Product Spaces", Springer-Verlag, New York, 2001.
2. C. W. Groetsch, Spectral methods for linear inverse problems involving unbounded operators, *J. Approx. Theory* **70** (1992), 16–28.
3. L. J. Lardy, A series representation of the generalized inverse of a closed linear operator, *Atti Accad. Naz. Lincei* (*Ser. VIII*) **58** (1975), 152–157.
4. F. Riesz and B. Sz.-Nagy, "Functional Analysis", Ungar, New York, 1955.
5. A. N. Tikhonov and V. Y. Arsenin, "Solutions of Ill-Posed Problems" (translated from the Russian), Wiley, New York, 1977.