6

LINEAR OPERATORS AND ADJOINTS

6.1 Introduction

A study of linear operators and adjoints is essential for a sophisticated approach to many problems of linear vector spaces. The associated concepts and notations of operator theory often streamline an otherwise cumbersome analysis by eliminating the need for carrying along complicated explicit formulas and by enhancing one's insight of the problem and its solution. This chapter contains no additional optimization principles but instead develops results of linear operator theory that make the application of optimization principles more straightforward in complicated situations. Of particular importance is the concept of the adjoint of a linear operator which, being defined in dual space, characterizes many aspects of duality theory.

Because it is difficult to obtain a simple geometric representation of an arbitrary linear operator, the material in this chapter tends to be somewhat more algebraic in character than that of other chapters. Effort is made, however, to extend some of the geometric ideas used for the study of linear functionals to general linear operators and also to interpret adjoints in terms of relations among hyperplanes.

6.2 Fundamentals

A *transformation* T is, as discussed briefly in Chapter 2, a mapping from one vector space to another. If T maps the space X into Y, we write $T: X \to Y$, and if T maps the vector $x \in X$ into the vector $y \in Y$, we write $y = T(x)$ and refer to y as the *image* of x under T. As before, we allow that a transformation may be defined only on a subset $D \subset X$, called the *domain* of T, although in most cases $D = X$. The collection of all vectors $y \in Y$ for which there is an $x \in D$ with $y = T(x)$ is called the *range* of T.

If $T: X \to Y$ and S is a given set in X, we denote by $T(S)$ the *image of S* in Y defined as the subset of Y consisting of points of the form $y = T(s)$

with $s \in S$. Similarly, given any set $P \subset Y$, we denote by $T^{-1}(P)$ the *inverse image of P* which is the set consisting of all points $x \in X$ satisfying $T(x) \in P$.

Our attention in this chapter is focused primarily on linear transformations which are alternatively referred to as *linear operators* or simply operators and are usually denoted by A, B, etc. For convenience we often omit the parentheses for a linear operator and write Ax for $A(x)$. The range of a linear operator $A : X \to Y$ is denoted $\mathscr{R}(A)$ and is obviously a subspace of Y. The set $\{x : Ax = \theta\}$ corresponding to the linear operator A is called the *nullspace* of A and denoted $\mathscr{N}(A)$. It is a subspace of X.

Of particular importance is the case in which X and Y are normed spaces and A is a continuous operator from X into Y. The following result is easily established.

Proposition 1. *A linear operator on a normed space X is continuous at every point in X if it is continuous at a single point.*

Analogous to the procedure for constructing the normed dual consisting of continuous linear functionals on a space X, it is possible to construct a normed space of continuous linear operators on X. We begin by defining the norm of a linear operator.

Definition. A linear operator A from a normed space X to a normed space Y is said to be *bounded* if there is a constant M such that $\|Ax\| \leq M \|x\|$ for all $x \in X$. The smallest such M which satisfies the above condition is denoted $\|A\|$ and called the *norm* of A.

Alternative, but equivalent, definitions of the norm are

$$\|A\| = \sup_{\|x\| \leq 1} \|Ax\|$$

$$\|A\| = \sup_{x \neq \theta} \frac{\|Ax\|}{\|x\|}.$$

We leave it to the reader to prove the following proposition.

Proposition 2. *A linear operator is bounded if and only if it is continuous.*

If addition and scalar multiplication are defined by

$$(A_1 + A_2)x = A_1 x + A_2 x$$

$$(\alpha A)x = \alpha(Ax)$$

the linear operators from X to Y form a linear vector space. If X and Y are normed spaces, the subspace of continuous linear operators can be identified and this becomes a normed space when the norm of an operator is defined according to the last definition. (The reader can easily verify that the requirements for a norm are satisfied.)

Definition. The normed space of all bounded linear operators from the normed space X into the normed space Y is denoted $B(X, Y)$.

We note the following result which generalizes Theorem 1, Section 5.2. The proof requires only slight modification of the proof in Section 5.2 and is omitted here.

Theorem 1. *Let X and Y be normed spaces with Y complete. Then the space $B(X, Y)$ is complete.*

In general the space $B(X, Y)$, although of interest by its own right, does not play nearly as dominant a role in our theory as that of the normed dual of X. Nevertheless, certain of its elementary properties and the definition itself are often convenient. For instance, we write $A \in B(X, Y)$ for, "let A be a continuous linear operator from the normed space X to the normed space Y."

Finally, before turning to some examples, we observe that the spaces of linear operators have a structure not present in an arbitrary vector space in that it is possible to define products of operators. Thus, if $S : X \to Y$, $T : Y \to Z$, we define the operator $TS : X \to Z$ by the equation $(TS)(x) = T(Sx)$ for all $x \in X$. For bounded operators we have the following useful result.

Proposition 3. *Let X, Y, Z be normed spaces and suppose $S \in B(X, Y)$, $T \in B(Y, Z)$. Then $\|TS\| \leq \|T\|\|S\|$.*

Proof. $\|TSx\| \leq \|T\|\|Sx\| \leq \|T\|\|S\|\|x\|$ for all $x \in X$. ∎

Example 1. Let $X = C[0, 1]$ and define the operator $A : X \to X$ by $Ax = \int_0^1 K(s, t)x(t)\,dt$ where the function K is continuous on the unit square $0 \leq s \leq 1$, $0 \leq t \leq 1$. The operator A is clearly linear. We compute $\|A\|$. We have

$$\begin{aligned}\|Ax\| &= \max_{0 \leq s \leq 1} \left| \int_0^1 K(s, t)x(t)\,dt \right| \\ &\leq \max_{0 \leq s \leq 1} \left\{ \int_0^1 |K(s, t)|\,dt \right\} \max_{0 \leq t \leq 1} |x(t)| \\ &= \max_{0 \leq s \leq 1} \int_0^1 |K(s, t)|\,dt \cdot \|x\|.\end{aligned}$$

Therefore,

$$\|A\| \leq \max_{0 \leq s \leq 1} \int_0^1 |K(s, t)|\,dt.$$

We can show that the quantity on the right-hand side is actually the norm of A. Let s_0 be the point at which the continuous function

$$\int_0^1 |K(s, t)|\, dt$$

achieves its maximum. Given $\varepsilon > 0$ let p be a polynomial which approximates $K(s_0, \cdot)$ in the sense that

$$\max_{0 \le t \le 1} |K(s_0, t) - p(t)| < \varepsilon$$

and let x be a function in $C[0, 1)$ with $\|x\| \le 1$ which approximates the discontinuous function sgn $p(t)$ in the sense that

$$\left| \int_0^1 p(t)x(t)\, dt - \int_0^1 |p(t)|\, dt \right| < \varepsilon.$$

This last approximation is easily constructed since p has only a finite number of sign changes.

For this x we have

$$\begin{aligned}
\left| \int_0^1 K(s_0, t)x(t)\, dt \right| &\ge \left| \int_0^1 p(t)x(t)\, dt \right| - \left| \int_0^1 [K(s_0, t) - p(t)]x(t)\, dt \right| \\
&\ge \left| \int_0^1 p(t)x(t)\, dt \right| - \varepsilon \ge \int_0^1 |p(t)|\, dt - 2\varepsilon \\
&\ge \int_0^1 |K(s_0, t)|\, dt - \left| \int_0^1 [|K(s_0, t)| - |p(t)|]\, dt \right| - 2\varepsilon \\
&\ge \int_0^1 |K(s_0, t)|\, dt - 3\varepsilon.
\end{aligned}$$

Thus, since $\|x\| \le 1$,

$$\|A\| \ge \int_0^1 |K(s_0, t)|\, dt - 3\varepsilon.$$

But since ε was arbitrary, and since the reverse inequality was established above, we have

$$\|A\| = \max_{0 \le s \le 1} \int_0^1 |K(s, t)|\, dt.$$

Example 2. Let $X = E^n$ and let $A : X \to X$. Then A is a matrix acting on the components of x. We have $\|Ax\|^2 = (x \,|\, A'Ax)$ where A' is the transpose of the matrix A. Denoting $A'A$ by Q, determination of $\|A\|$ is equivalent to maximizing $(x \,|\, Qx)$ subject to $\|x\|^2 \le 1$. This is a finite-dimensional optimization problem. Since Q is symmetric and positive semidefinite, it has nonnegative eigenvalues and the solution of the optimization problem is given by x equal to the eigenvector of Q corresponding to the largest eigenvalue.

We conclude that $\|A\| = \sqrt{\lambda_{\max}}$.

Example 3. The operator $Ax = d/dt\, x(t)$, defined on the subspace M of $C[0, 1]$ consisting of all continuously differentiable functions, has range $C[0, 1]$. A is not bounded, however, since elements of arbitrarily small norm can produce elements of large norm when differentiated. On the other hand, if A is regarded as having domain $D[0, 1]$ and range $C[0, 1]$, it is bounded with $\|A\| = 1$.

INVERSE OPERATORS

6.3 Linearity of Inverses

Let $A : X \to Y$ be a linear operator between two linear spaces X and Y. Corresponding to A we consider the equation $Ax = y$. For a given $y \in Y$ this equation may:

1. have a unique solution $x \in X$,
2. have no solution,
3. have more than one solution.

Many optimization problems can be regarded as arising from cases 2 or 3; these are discussed in Section 6.9. Condition 1 holds for every $y \in Y$ if and only if the mapping A from X to Y is one-to-one and has range equal to Y, in which case the operator A has an *inverse* A^{-1} such that if $Ax = y$, then $A^{-1}(y) = x$.

Proposition 1. *If a linear operator $A : X \to Y$ has an inverse, the inverse A^{-1} is linear.*

Proof. Suppose $A^{-1}(y_1) = x_1$, $A^{-1}(y_2) = x_2$, then

$$A(x_1) = y_1, \qquad A(x_2) = y_2,$$

and the linearity of A implies that $A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 y_1 + \alpha_2 y_2$. Thus $A^{-1}(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 A^{-1}(y_1) + \alpha_2 A^{-1}(y_2)$. ∎

The solution of linear equations and the determination of inverse operators are, of course, important areas of pure and applied mathematics. For optimization theory, however, we are not so much interested in solving equations as formulating the equations appropriate for characterizing an optimal vector. Once the equations are formulated, we may rely on standard techniques for their solution. There are important exceptions to this point of view, however, since optimization theory often provides effective procedures for solving equations. Furthermore, a problem can never really be regarded as resolved until an efficient computational method of solution is derived. Nevertheless, our primary interest in linear operators is their role in optimization problems. We do not develop an extensive theory of linear equations but are content with establishing the existence of a solution.

6.4 The Banach Inverse Theorem

Given a continuous linear operator A from a normed space X onto a normed space Y and assuming that A has an inverse A^{-1}, it follows that A^{-1} is linear but not necessarily continuous. If, however, X and Y are Banach spaces, A^{-1} must be continuous if it exists. This result, known as the Banach inverse theorem, is one of the analytical cornerstones of functional analysis. Many important, deep, and sometimes surprising results follow from it. We make application of the result in Section 6.6 and again in Chapter 8 in connection with Lagrange multipliers. Other applications to problems of mathematical analysis are discussed in the problems at the end of this chapter.

This section is devoted to establishing this one result. Although the proof is no more difficult at each step than that of most theorems in this book, it involves a number of steps. Therefore, since it plays only a supporting role in the optimization theory, the reader may wish to simply scan the proof and proceed to the next section.

We begin by establishing the following lemma which itself is an important and celebrated tool of analysis.

Lemma 1. (*Baire*) *A Banach space X is not the union of countably many nowhere dense sets in X.*

Proof. Suppose that $\{E_n\}$ is a sequence of nowhere dense sets and let F_n denote the closure of E_n. Then F_n contains no sphere in X. It follows that each of the sets $\tilde{F}_n$ is open and dense in X.

Let $S(x_1, r_1)$ be a sphere in $\tilde{F}_1$ with center at x_1 and radius r_1. Let $S(x_2, r_2)$ be a sphere in $\tilde{F}_2 \cap S(x_1, r_1/2)$. (Such a sphere exists since $\tilde{F}_2$ is open and dense.) Proceeding inductively, let $S(x_n, r_n)$ be a sphere in $S(x_{n-1}, r_{n-1}/2) \cap \tilde{F}_n$.

The sequence $\{x_n\}$ so defined is clearly a Cauchy sequence and, thus, by the completeness of X, there is a limit x; $x_n \to x$. This vector x lies in each of the $S(x_n, r_n)$, because, indeed, $x_{n+k} \in S(x_n, r_n/2)$ for $k \geq 1$. Hence x lies in each $\tilde{F}_n$. Therefore, $x \in \bigcap_n \tilde{F}_n$.

It follows that the union of the original collection of sets $\{E_n\}$ is not X since

$$x \in \bigcap_n \tilde{F}_n = \left[\widetilde{\bigcup_n F_n}\right] \subset \left[\widetilde{\bigcup_n E_n}\right]. \blacksquare$$

Theorem 1. (*Banach Inverse Theorem*) *Let A be a continuous linear operator from a Banach space X onto a Banach space Y and suppose that the inverse operator A^{-1} exists. Then A^{-1} is continuous.*

Proof. In view of the linearity of A and therefore of A^{-1}, it is only necessary to show that A^{-1} is bounded. For this it is only necessary to show that the image $A(S)$ in Y of any sphere S centered at the origin in X contains a sphere P centered at the origin in Y, because then the inverse image of P is contained in S. The proof amounts to establishing the existence of a sphere in $A(S)$.

Given a sphere S, for any $x \in X$ there is an integer n such that $x/n \in S$ and hence $A(x/n) \in A(S)$ or, equivalently, $A(x) \in nA(S)$. Since A maps X onto Y, it follows that

$$Y = \bigcup_{n=1}^{\infty} nA(S).$$

According to Baire's lemma, Y cannot be the union of countably many nowhere dense sets and, hence, there is an n such that the closure of $nA(S)$ contains a sphere. It follows that $\overline{A(S)}$ contains a sphere whose center y may be taken to be in $A(S)$. Let this sphere $N(y, r)$ have radius r, and let $y = A(x)$. Now as y' varies over $N(y, r)$, the points $y' - y$ cover the sphere $N(\theta, r)$ and the points of a dense subset of these are of the form $A(x' - x)$ where $A(x') = y'$, $x' \in S$. Since $x', x \in S$, it follows that $x' - x \in 2S$. Hence, the closure of $A(2S)$ contains $N(\theta, r)$ (and by linearity $\overline{A(S)}$ contains $N(\theta, r/2)$).

We have shown that the closure of the image of a sphere centered at the origin contains such a sphere in Y, but it remains to be shown that the image itself, rather than its closure, contains a sphere. For any $\varepsilon > 0$, let $S(\varepsilon)$ and $P(\varepsilon)$ be the spheres in X, Y, respectively, of radii ε centered at the origins. Let $\varepsilon_0 > 0$ be arbitrary and let $\eta_0 > 0$ be chosen so that $P(\eta_0)$ is a sphere contained in the closure of the image of $S(\varepsilon_0)$. Let y be an arbitrary point in $P(\eta_0)$. We show that there is an $x \in S(2\varepsilon_0)$ such that $Ax = y$ so that the image of the sphere of radius $2\varepsilon_0$ contains the sphere $P(\eta_0)$.

Let $\{\varepsilon_i\}$ be a sequence of positive numbers such that $\sum_{i=1}^{\infty} \varepsilon_i < \varepsilon_0$. Then there is a sequence $\{\eta_i\}$, with $\eta_i > 0$ and $\eta_i \to 0$, such that

$$P(\eta_i) \subset \overline{A[S(\varepsilon_i)]}$$

Since $A(S(\varepsilon_0))$ is dense in $P(\varepsilon_0)$, there is an $x_0 \in S(\varepsilon_0)$ such that $y - Ax_0 \in P(\eta_1)$. It follows that there is an $x_1 \in S(\varepsilon_1)$ with $y - Ax_0 - Ax_1 \in P(\eta_2)$. Proceeding inductively, a sequence $\{x_n\}$ is defined with $x_n \in S(\varepsilon_n)$ and $y - A(\sum_{i=0}^{n} x_i) \in P(\eta_{n+1})$. Let $z_n = x_0 + x_1 + \cdots + x_n$. Then evidently $\{z_n\}$ is a Cauchy sequence since for $m > n$, $\|z_m - z_n\| = \|x_{n-1} + x_{n-2} + \cdots + x_m\| < \varepsilon_{n+1} + \varepsilon_{n+2} + \cdots + \varepsilon_m$. Thus there is an $x \in X$ such that $z_n \to x$. Furthermore, $\|x\| < \varepsilon_0 + \varepsilon_1 + \cdots + \varepsilon_n + \cdots < 2\varepsilon_0$; so $x \in S(2\varepsilon_0)$. Since A is continuous, $Az_n \to Ax$, but since $\|y - Az_n\| < \eta_{n+1} \to 0$, $Az_n \to y$. Therefore, $Ax = y$. ▮

ADJOINTS

6.5 Definition and Examples

The constraints imposed in many optimization problems by differential equations, matrix equations, etc., can be described by linear operators. The resolution of these problems almost invariably calls for consideration of an associated operator: the adjoint. The reason for this is that adjoints provide a convenient mechanism for describing the orthogonality and duality relations which permeate nearly every optimization analysis.

Definition. Let X and Y be normed spaces and let $A \in B(X, Y)$. The *adjoint operator* $A^*: Y^* \to X^*$ is defined by the equation

$$\langle x, A^*y^* \rangle = \langle Ax, y^* \rangle.$$

This important definition requires a bit of explanation and justification. Given a fixed $y^* \in Y^*$, the quantity $\langle Ax, y^* \rangle$ is a scalar for each $x \in X$ and is therefore a functional on X. Furthermore, by the linearity of y^* and A, it follows that this functional is linear. Finally, since

$$|\langle Ax, y^* \rangle| \leq \|y^*\| \|Ax\| \leq \|y^*\| \|A\| \|x\|,$$

it follows that this functional is bounded and is thus an element x^* of X^*. We then define $A^*y^* = x^*$. The adjoint is obviously unique and the reader can verify that it is linear. It is important to remember, as illustrated in Figure 6.1, that $A^*: Y^* \to X^*$.

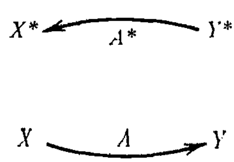

Figure 6.1 An operator and its adjoint

In terms of operator, rather than bracket, notation the definition of the adjoint satisfies the equation

$$y^*(Ax) = (A^*y^*)(x)$$

for each $x \in X$. Thus we may write

$$y^*A = A^*y^*$$

where the left side denotes the functional on X which is the composition of the operators A and y^* and the right side is the functional obtained by operating on y^* by A^*.

Theorem 1. *The adjoint operator A^* of the linear operator $A \in B(X, Y)$ is linear and bounded with $\|A^*\| = \|A\|$.*

Proof. The proof of linearity is elementary and left to the reader. From the inequalities

$$|\langle x, A^*y^*\rangle| = |\langle Ax, y^*\rangle| \leq \|y^*\|\,\|Ax\| \leq \|y^*\|\,\|A\|\,\|x\|$$

it follows that

$$\|A^*y^*\| \leq \|A\|\,\|y^*\|$$

which implies that

$$\|A^*\| \leq \|A\|.$$

Now let x_0 be any nonzero element of X. According to Corollary 2 of the Hahn-Banach theorem, there exists an element $y_0^* \in Y^*$, $\|y_0^*\| = 1$, such that $\langle Ax_0, y_0^*\rangle = \|Ax_0\|$. Therefore,

$$\|Ax_0\| = |\langle x_0, A^*y_0^*\rangle| \leq \|A^*y_0^*\|\,\|x_0\| \leq \|A^*\|\,\|x_0\|$$

from which we conclude that

$$\|A\| \leq \|A^*\|$$

It now follows that $\|A^*\| = \|A\|$. ∎

In addition to the above result, adjoints enjoy the following algebraic relations which follow easily from the basic definition.

Proposition 1. *Adjoints satisfy the following properties:*

1. *If I is the identity operator on a normed space X, then $I^* = I$.*
2. *If $A_1, A_2 \in B(X, Y)$, then $(A_1 + A_2)^* = A_1^* + A_2^*$.*
3. *If $A \in B(X, Y)$ and α is a real scalar, then $(\alpha A)^* = \alpha A^*$.*
4. *If $A_1 \in B(X, Y)$, $A_2 \in B(Y, Z)$, then $(A_2 A_1)^* = A_1^* A_2^*$.*
5. *If $A \in B(X, Y)$ and A has a bounded inverse, then $(A^{-1})^* = (A^*)^{-1}$.*

Proof. Properties 1–4 are trivial. To prove property 5, let $A \in B(X, Y)$ have a bounded inverse A^{-1}. To show that A^* has an inverse, we must show that it is one-to-one and onto. Let $y_1^* \neq y_2^* \in Y^*$, then

$$\langle x, A^*y_1^*\rangle - \langle x, A^*y_2^*\rangle = \langle Ax, y_1^* - y_2^*\rangle \neq 0$$

for some $x \in X$. Thus, $A^*y_1^* \neq A^*y_2^*$ and A^* is one-to-one. Now for any $x^* \in X^*$ and any $x \in X$, $Ax = y$, we have

$$\langle x, x^*\rangle = \langle A^{-1}y, x^*\rangle = \langle y, (A^{-1})^*x^*\rangle$$
$$= \langle Ax, (A^{-1})^*x^*\rangle = \langle x, A^*(A^{-1})^*x^*\rangle$$

which shows that x^* is in $\mathscr{R}(A^*)$ and also that $(A^*)^{-1} = (A^{-1})^*$. ∎

An important special case is that of a linear operator $A : H \to G$ where H and G are Hilbert spaces. If H and G are real, then they are their own duals in the sense of Section 5.3, and the operator A^* can be regarded as mapping G into H. In this case the adjoint relation becomes $(Ax \mid y) = (x \mid A^*y)$. If the spaces are complex, the adjoint, as defined earlier, does not satisfy this relation and it is convenient and customary to redefine the Hilbert space adjoint directly by the relation $(Ax \mid y) = (x \mid A^*y)$. In our study, however, we restrict our attention to real spaces so that difficulties of this nature can be ignored.

Note that in Hilbert space we have the additional property: $A^{**} = A$.

Finally, we note the following two definitions.

Definition. A bounded linear operator A mapping a real Hilbert space into itself is said to be *self-adjoint* if $A^* = A$.

Definition. A self-adjoint linear operator A on a Hilbert space H is said to be *positive semidefinite* if $(x \mid Ax) \geq 0$ for all $x \in H$.

Example 1. Let $X = Y = E^n$. Then $A : X \to X$ is represented by an $n \times n$ matrix. Thus the i-th component of Ax is

$$(Ax)_i = \sum_{j=1}^{n} a_{ij} x_j .$$

We compute A^*. For $y \in Y$ we have

$$(Ax \mid y) = \sum_{i=1}^{n} \sum_{j=1}^{n} y_i a_{ij} x_j = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij} y_i = (x \mid A^*y)$$

where A^* is the matrix with elements $a_{ij}^* = a_{ji}$. Thus A^* is the transpose of A.

Example 2. Let $X = Y = L_2[0, 1]$ and define

$$Ax = \int_0^1 K(t, s)x(s)\, ds, \qquad t \in [0, 1]$$

where

$$\int_0^1 \int_0^1 |K(t, s)|^2 \, ds \, dt < \infty.$$

Then

$$(Ax \mid y) = \int_0^1 y(t) \int_0^1 K(t, s)x(s) \, ds \, dt$$
$$= \int_0^1 x(s) \int_0^1 K(t, s)y(t) \, dt \, ds.$$

Or, by interchanging the roles of s and t,

$$(Ax \mid y) = \int_0^1 x(t) \int_0^1 K(s, t)y(s) \, ds \, dt = (x \mid A^*y)$$

where

$$A^*y = \int_0^1 K(s, t)y(s) \, ds.$$

Therefore, the adjoint of A is obtained by interchanging s and t in K.

Example 3. Again let $X = Y = L_2[0, 1]$ and define

$$Ax = \int_0^t K(t, s)x(s) \, ds, \qquad t \in [0, 1],$$

with

$$\int_0^1 \int_0^1 |K(t, s)|^2 \, dt \, ds < \infty.$$

Then

$$(Ax \mid y) = \int_0^1 y(t) \int_0^t K(t, s)x(s) \, ds \, dt$$
$$= \int_0^1 \int_0^t y(t)K(t, s)x(s) \, ds \, dt.$$

The double integration represents integration over the triangular region shown in Figure 6.2*a*, integrating vertically and then horizontally. Alternatively, the integration may be performed in the reverse order as in Figure 6.2*b*, leading to

$$(Ax \mid y) = \int_0^1 \int_s^1 y(t)K(t, s)x(s) \, dt \, ds$$
$$= \int_0^1 x(s)\left(\int_s^1 K(t, s)y(t) \, dt\right) ds.$$

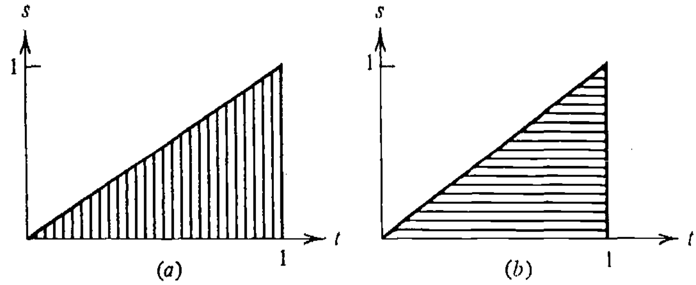

Figure 6.2 Region of integration

Or, interchanging the roles of t and s,

$$(Ax \mid y) = \int_0^1 x(t)\left(\int_t^1 K(s, t)y(s)\,ds\right) dt = (x \mid A^*y)$$

where

$$A^*y = \int_t^1 K(s, t)y(s)\,ds.$$

This example comes up frequently in the study of dynamic systems.

Example 4. Let $X = C[0, 1]$, $Y = E^n$ and define $A : X \to Y$ by the equation

$$Ax = (x(t_1), x(t_2), \ldots, x(t_n))$$

where $0 \leq t_1 < t_2 < t_2 < \cdots < t_n \leq 1$ are fixed. It is easily verified that A is continuous and linear. Let $y^* = (y_1, y_2, \ldots, y_n)$ be a linear functional on E^n. Then

$$\langle Ax, y^* \rangle = \sum_{i=1}^{n} y_i x(t_i) = \int_0^1 x(t)\,dv(t) = \langle x, A^*y^* \rangle$$

where $v(t)$ is constant except at the points t_i where it has a jump of magnitude y_i, as illustrated in Figure 6.3. Thus $A^*: E^n \to NBV[0, 1]$ is defined by $A^*y^* = v$.

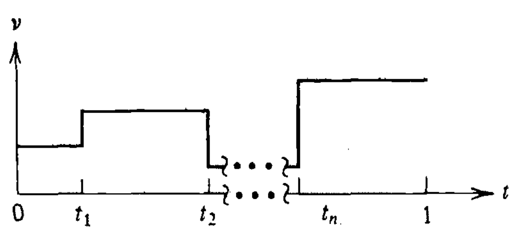

Figure 6.3 The function v

6.6 Relations Between Range and Nullspace

Adjoints are extremely useful in our recurring task of translating between the geometric properties and the algebraic description of a given problem. The following theorem and others similar to it are of particular interest.

Theorem 1. *Let X and Y be normed spaces and let $A \in B(X, Y)$. Then*

$$[\mathscr{R}(A)]^{\perp} = \mathscr{N}(A^*).$$

Proof. Let $y^* \in \mathscr{N}(A^*)$ and $y \in \mathscr{R}(A)$. Then $y = Ax$ for some $x \in X$. The calculation $\langle y, y^* \rangle = \langle Ax, y^* \rangle = \langle x, A^*y^* \rangle = 0$ shows that $\mathscr{N}(A^*) \subset [\mathscr{R}(A)]^{\perp}$.

Now assume $y^* \in [\mathscr{R}(A)]^{\perp}$. Then for every $x \in X$, $\langle Ax, y^* \rangle = 0$. This implies $\langle x, A^*y^* \rangle = 0$ and hence that $[\mathscr{R}(A)]^{\perp} \subset \mathscr{N}(A^*)$. ∎

Example 1. Let us consider the finite-dimensional version of Theorem 1. Let A be a matrix; $A : E^n \to E^m$. A consists of n column vectors a_i, $i = 1, 2, \ldots, n$, and $\mathscr{R}(A)$ is the subspace of E^m spanned by these vectors. $[\mathscr{R}(A)]^{\perp}$ consists of those vectors in E^m that are orthogonal to each a_i.

On the other hand, the matrix A^* (which is just the transpose of A) has the a_i's as its rows; hence the vectors in E^m orthogonal to a_i's comprise the nullspace of A^*. Therefore, both $[\mathscr{R}(A)]^{\perp}$ and $\mathscr{N}(A^*)$ consist of all vectors orthogonal to each a_i.

Our next theorem is a dual to Theorem 1. It should be noted, however, that the additional hypothesis that $\mathscr{R}(A)$ be closed, is required. Moreover, the dual theorem is much deeper than Theorem 1, since the proof requires both the Banach inverse theorem and the Hahn-Banach theorem.

Lemma 1. *Let X and Y be Banach spaces and let $A \in B(X, Y)$. Assume that $\mathscr{R}(A)$ is closed. Then there is a constant K such that for each $y \in \mathscr{R}(A)$ there is an x satisfying $Ax = y$ and $\|x\| \le K\|y\|$.*

Proof. Let $N = \mathscr{N}(A)$ and consider the space X/N consisting of equivalence classes $[x]$ modulo N. Define $\bar{A} : X/N \to \mathscr{R}(A)$ by $\bar{A}[x] = Ax$. It is easily verified that $\bar{A}$ is one-to-one, onto, linear, and bounded. Since $\mathscr{R}(A)$ closed implies that $\mathscr{R}(A)$ is a Banach space, it follows from the Banach inverse theorem that $\bar{A}$ has a continuous inverse. Hence, given $y \in \mathscr{R}(A)$, there is $[x] \in X/N$ with $\|[x]\| \le \|\bar{A}^{-1}\|\|y\|$. Take $x \in [x]$ with $\|x\| \le 2\|[x]\|$ and then $K = 2\|\bar{A}^{-1}\|$ satisfies the conditions stated in the lemma. ∎

Now we give the dual to Theorem 1.

Theorem 2. *Let X and Y be Banach spaces and let $A \in B(X, Y)$. Let $\mathscr{R}(A)$ be closed. Then*

$$\mathscr{R}(A^*) = [\mathscr{N}(A)]^\perp.$$

Proof. Let $x^* \in \mathscr{R}(A^*)$. Then $x^* = A^*y^*$ for some $y^* \in Y^*$. For any $x \in \mathscr{N}(A)$, we have

$$\langle x, x^* \rangle = \langle x, A^*y^* \rangle = \langle Ax, y^* \rangle = 0.$$

Thus $x^* \in [\mathscr{N}(A)]^\perp$ and it follows that $\mathscr{R}(A^*) \subset [\mathscr{N}(A)]^\perp$.

Now assume that $x^* \in [\mathscr{N}(A)]^\perp$. For $y \in \mathscr{R}(A)$ and each x satisfying $Ax = y$, the functional $\langle x, x^* \rangle$ has the same value. Hence, define $f(y) = \langle x, x^* \rangle$ on $\mathscr{R}(A)$. Let K be defined as in the lemma. Then for each $y \in \mathscr{R}(A)$ there is an x with $\|x\| \leq K\|y\|$, $Ax = y$. Therefore, $|f(y)| \leq K\|x^*\|\|y\|$ and thus f is a bounded linear functional on $\mathscr{R}(A)$. Extend f by the Hahn-Banach theorem to a functional $y^* \in Y^*$. Then from

$$\langle x, A^*y^* \rangle = \langle Ax, y^* \rangle = \langle x, x^* \rangle,$$

it follows that $A^*y^* = x^*$ and thus $\mathscr{R}(A^*) \supset [\mathscr{N}(A)]^\perp$. ∎

In many applications the range of the underlying operator is finite dimensional, and hence satisfies the closure requirement. In other problems, however, this requirement is not satisfied and this generally leads to severe analytical difficulties. We give an example of an operator whose range is not closed.

Example 2. Let $X = Y = l_1$ with $A: X \to Y$ defined by

$$A\{\xi_1, \xi_2, \ldots, \xi_n, \ldots\} = \left\{\xi_1, \frac{1}{2}\xi_2, \frac{1}{3}\xi_3, \cdots, \frac{1}{n}\xi_n, \cdots\right\}.$$

Then $\mathscr{R}(A)$ contains all finitely nonzero sequences and thus $\overline{\mathscr{R}(A)} = Y$. However,

$$y = \left\{1, \frac{1}{2^2}, \frac{1}{3^2}, \cdots, \frac{1}{n^2}, \cdots\right\} \notin \mathscr{R}(A)$$

and thus $\mathscr{R}(A)$ is not closed.

In Hilbert space there are several additional useful relations between range and nullspace similar to those which hold in general normed space. These additional properties are a consequence of the fact that in Hilbert space an operator and its adjoint are defined on the same space.

Theorem 3. *Let A be a bounded linear operator acting between two real Hilbert spaces. Then*

1. $[\mathscr{R}(A)]^{\perp} = \mathscr{N}(A^*)$.
2. $\overline{\mathscr{R}(A)} = [\mathscr{N}(A^*)]^{\perp}$.
3. $[\mathscr{R}(A^*)]^{\perp} = \mathscr{N}(A)$.
4. $\overline{\mathscr{R}(A^*)} = [\mathscr{N}(A)]^{\perp}$.

Proof. Part 1 is just Theorem 1. To prove part 2, take the orthogonal complement of both sides of 1 obtaining $[\mathscr{R}(A)]^{\perp\perp} = [\mathscr{N}(A^*)]^{\perp}$. Since $\overline{\mathscr{R}(A)}$ is a subspace, the result follows. Parts 3 and 4 are obtained from 1 and 2 by use of the relation $A^{**} = A$. ∎

Example 3. Let $X = Y = l_2$. For $x = \{\xi_1, \xi_2 \ldots,\}$, define $Ax = \{0, \xi_1, \xi_2, \ldots\}$. A is a shift operator (sometimes referred to as the creation operator because a new component is created). The adjoint of A is easily computed to be the operator taking $y = \{\eta_1, \eta_2, \ldots\}$ into $A^*y = \{\eta_2, \eta_3, \ldots\}$, which is a shift in the other direction (referred to as the destruction operator). It is clear that $[\mathscr{R}(A)]^{\perp}$ consists of all those vectors in l_2 that are zero except possibly in their first component; this subspace is identical with $\mathscr{N}(A^*)$.

6.7 Duality Relations for Convex Cones

The fundamental algebraic relations between nullspace and range for an operator and its adjoint derived in Section 6.6 have generalizations which often play a role in the analysis of problems described by linear inequalities analogous to the role of the earlier results to problems described by linear equalities.

Definition. Given a set S in a normed space X, the set $S^{\oplus} = \{x^* \in X^*: \langle x, x^* \rangle \geq 0 \text{ for all } x \in S\}$ is called the *positive conjugate cone of S.* Likewise the set $S^{\ominus} = \{x^* \in X^* : \langle x, x^* \rangle \leq 0 \text{ for all } x \in S\}$ is called the *negative conjugate cone of S.*

It is a simple matter to verify that $S^{\oplus}$ and $S^{\ominus}$ are in fact convex cones. They are nonempty since they always contain the zero functional. If S is a subspace of X, then obviously $S^{\oplus} = S^{\ominus} = S^{\perp}$; hence, the conjugate cones can be regarded as generalizations of the orthogonal complement of a set. The definition is illustrated in Figure 6.4 for the Hilbert space situation where $S^{\oplus}$ and $S^{\ominus}$ can be regarded as subsets of X. The basic properties

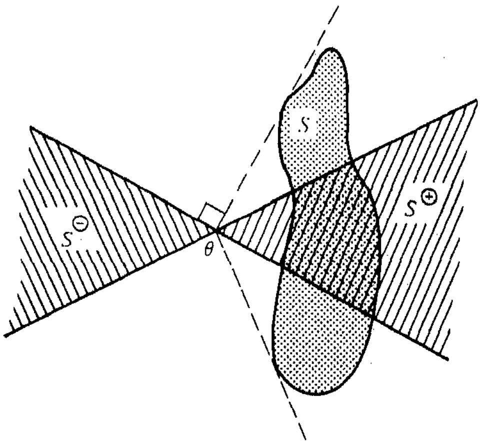

Figure 6.4 A set and its conjugate cones

of the operation of taking conjugate cones are given in the following proposition.

Proposition 1. *Let S and T be sets in a normed space X. Then*

1. *$S^{\oplus}$ is a closed convex cone in X^*.*
2. *If $S \subset T$, then $T^{\oplus} \subset S^{\oplus}$.*

In the general case the conjugate cone can be interpreted as a collection of half-spaces. If $x^* \in S^{\oplus}$, then clearly $\inf_{x \in S} \langle x, x^* \rangle \geq 0$ and hence the hyperplane $\{x : \langle x, x^* \rangle = 0\}$ has S in its positive half-space. Conversely, if x^* determines a hyperplane having S in its positive half-space, it is a member of $S^{\oplus}$. Therefore, $S^{\oplus}$ consists of all x^* which contain S in their positive half-spaces.

The following theorem generalizes Theorem 1 of Section 6.6.

Theorem 1. *Let X and Y be normed linear spaces and let $A \in B(X, Y)$. Let S be a subset of X. Then*

$$[A(S)]^{\oplus} = A^{*-1}(S^{\oplus})$$

(where the inverse denotes the inverse image of $S^{\oplus}$).

Proof. Assume $y^* \in [A(S)]^{\oplus}$ and $s \in S$. Then $\langle As, y^* \rangle \geq 0$ and hence $\langle s, A^*y^* \rangle \geq 0$. Thus, since s is arbitrary in S, $y^* \in A^{*-1}(S^{\oplus})$. The argument is reversible. ∎

Note that by putting $S = X$, $S^{\oplus} = \{\theta\}$, the above result reduces to $[\mathscr{R}(A)]^{\perp} = \mathscr{N}(A^*)$.

*6.8 Geometric Interpretation of Adjoints

It is somewhat difficult to obtain a clear simple visualization of the relation between an operator and its adjoint since if $A : X \to Y$, $A^* : Y^* \to X^*$, four spaces and two operators are involved. However, in view of the unique correspondence between hyperplanes not containing the origin in a space and nonzero elements of its dual, the adjoint A^* can be regarded as mapping hyperplanes in Y into hyperplanes in X. This observation can be used to consolidate the adjoint relations into two spaces rather than four. We limit our discussion here to invertible operators between Banach spaces. The arguments can be extended to the more general case, but the picture becomes somewhat more complex.

Let us fix our attention on a given hyperplane $H \subset X$ having $\theta \notin H$. The operator A maps this hyperplane point by point into a subset L of Y. It follows from the linearity of A that L is a linear variety, and since A is assumed to be invertible, it follows that L is in fact a hyperplane in Y not containing $\theta \in Y$. Therefore, A maps the hyperplane H point by point into a hyperplane L.

The hyperplanes H and L define unique elements $x_1^* \in X^*$ and $y_1^* \in Y^*$ through the relations $H = \{x : \langle x, x_1^* \rangle = 1\}$, $L = \{y : \langle y, y_1^* \rangle = 1\}$. The adjoint operator A^* can then be applied to y_1^* to produce an x_1^* or, equivalently, A^* maps L into a hyperplane in X. In fact, A^* maps L back to H. For if $A^* y_1^* = x_1^*$, it follows directly from the definition of adjoints that $\{x : \langle x, x_1^* \rangle = 1\} = \{x : \langle x, A^* y_1^* \rangle = 1\} = \{x : \langle Ax, y_1^* \rangle = 1\} = H$. Therefore, A^* maps the hyperplane L, as a unit, back to the hyperplane H. This interpretation is illustrated in Figure 6.5 where the dotted line arrows symbolize elements of a dual space.

Another geometric interpretation is discussed in Problem 13.

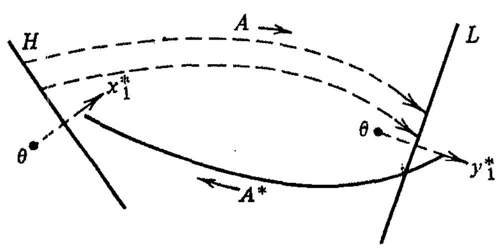

Figure 6.5 Geometric interpretation of adjoints

OPTIMIZATION IN HILBERT SPACE

Suppose A is a bounded linear operator from a Hilbert space G into a Hilbert space H; $A : G \to H$. Then, as pointed out previously, the linear equation $Ax = y$ may, for a given $y \in H$,

1. possess a unique solution $x \in G$,
2. possess no solution,
3. possess more than one solution.

Case 1 is in many respects the simplest. We found in Section 6.4 that in this case A has a unique bounded inverse A^{-1}. The other two cases are of interest in optimization since they allow some choice of an optimal x to be made. Indeed, most of the problems that were solved by the projection theorem can be viewed this way.

6.9 The Normal Equations

When no solution exists (case 2), we resolve the problem by finding an approximate solution.

Theorem 1. *Let G and H be Hilbert spaces and let $A \in B(G, H)$. Then for a fixed $y \in H$ the vector $x \in G$ minimizes $\|y - Ax\|$ if and only if $A^*Ax = A^*y$.*

Proof. The problem is obviously equivalent to that of minimizing $\|y - \hat{y}\|$ where $\hat{y} \in \mathscr{R}(A)$. Thus, by Theorem 1, Section 3.3 (the projection theorem without the existence part), $\hat{y}$ is a minimizing vector if and only if $y - \hat{y} \in [\mathscr{R}(A)]^\perp$. Hence, by Theorem 3 of Section 6.6, $y - \hat{y} \in \mathscr{N}(A^*)$. Or $\theta = A^*(y - \hat{y}) = A^*y - A^*Ax$. ∎

Theorem 1 is just a restatement of the first form of the projection theorem applied to the subspace $\mathscr{R}(A)$. There is no statement of existence in the theorem since in general $\mathscr{R}(A)$ may not be closed. Furthermore, there is no statement of uniqueness of the minimizing vector x since, although $\hat{y} = Ax$ is unique, the preimage of $\hat{y}$ may not be unique. If a unique solution always exists, i.e., if A^*A is invertible, the solution takes the form

$$x = (A^*A)^{-1}A^*y.$$

Example 1. We consider again the basic approximation problem in Hilbert space. Let $\{x_1, x_2, \ldots, x_n\}$ be an independent set of vectors in a real Hilbert space H. We seek the best approximation to $y \in H$ of the form $\hat{y} = \sum_{i=1}^n a_i x_i$.

Define the operator $A : E^n \to H$ by the equation

$$Aa = \sum_{i=1}^{n} a_i x_i,$$

where $a = (a_1, \ldots, a_n)$. The approximation problem is equivalent to minimizing $\|y - Aa\|$. Thus, according to Theorem 1, the optimal solution must satisfy $A^*Aa = A^*y$.

It remains to compute the operator A^*. Clearly, $A^*: H \to E^n$. For any $x \in H$, $a \in E^n$, we have

$$(x \mid Aa) = (x \mid \sum_{i=1}^{n} a_i x_i) = \sum_{i=1}^{n} a_i (x \mid x_i) = (z \mid a)_{E^n}$$

where $z = ((x \mid x_1), \ldots, (x \mid x_n))$. Thus, $A^*x = ((x \mid x_1), (x \mid x_2), \ldots, (x \mid x_n))$.

The operator A^*A maps E^n into E^n and is therefore represented by an $n \times n$ matrix. It is then easily deduced that the equation $A^*Aa = A^*y$ is equivalent to

$$\begin{bmatrix} (x_1 \mid x_1) & (x_2 \mid x_1) & \cdots & (x_n \mid x_1) \\ (x_1 \mid x_2) & & & \\ \vdots & & & \\ (x_1 \mid x_n) & & \cdots & (x_n \mid x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (y \mid x_1) \\ (y \mid x_2) \\ \vdots \\ (y \mid x_n) \end{bmatrix},$$

the normal equations.

The familiar arguments for this problem show that the normal equations possess a unique solution and that the Gram matrix A^*A is invertible. Thus, $a = (A^*A)^{-1}A^*y$.

The above example illustrates that operator notation can streamline an optimization analysis by supplying a compact notational solution. The algebra required to compute adjoints and reduce the equations to expressions involving the original problem variables is, however, no shorter.

6.10 The Dual Problem

If the equation $Ax = y$ has more than one solution, we may choose the solution having minimum norm.

Theorem 1. *Let G and H be Hilbert spaces and let $A \in B(G, H)$ with range closed in H. Then the vector x of minimum norm satisfying $Ax = y$ is given by $x = A^*z$ where z is any solution of $AA^*z = y$.*

Proof. If x_1 is a solution of $Ax = y$, the general solution is $x = x_1 + u$ where $u \in \mathcal{N}(A)$. Since $\mathcal{N}(A)$ is closed, it follows that there exists a unique vector x of minimum norm satisfying $Ax = y$ and that this vector is orthogonal to $\mathcal{N}(A)$. Thus, since $\mathcal{R}(A)$ is assumed closed,

$$x \in [\mathcal{N}(A)]^{\perp} = \mathcal{R}(A^*).$$

Hence $x = A^*z$ for some $z \in H$, and since $Ax = y$, we conclude that $AA^*z = y$. ∎

Note that if, as is frequently the case, the operator AA^* is invertible, the optimal solution takes the form

$$x = A^*(AA^*)^{-1}y.$$

Example 1. Suppose a linear dynamic system is governed by a set of differential equations of the form

$$\dot{x}(t) = Fx(t) + bu(t)$$

where x is an $n \times 1$ vector of time functions, F is an $n \times n$ matrix, b is an $n \times 1$ vector, and u is a scalar control function.

Assume that $x(0) = \theta$ and that it is desired to transfer the system to $x(T) = x_1$ by application of suitable control. Of the class of controls which accomplish the desired transfer, we seek the one of minimum energy $\int_0^T u^2(t)\,dt$. The problem includes the motor problem discussed in Chapter 3.

The explicit solution to the equation of motion is

$$x(T) = \int_0^T e^{F(T-t)}bu(t)\,dt.$$

Thus, defining the operator $A : L_2[0, T] \to E^n$ by

$$Au = \int_0^T e^{F(T-t)}bu(t)\,dt,$$

the problem is equivalent to that of determining the u of minimum norm satisfying $Au = x_1$.

Since $\mathcal{R}(A)$ is finite dimensional, it is closed. Thus the results of Theorem 1 apply and we write the optimal solution as

$$u = A^*z$$

where

$$AA^*z = x_1.$$

It remains to calculate the operators A^* and AA^*. For any $u \in L_2$, $y \in E^n$

$$(y \mid Au)_{E^n} = y' \int_0^T e^{F(T-t)}bu(t)\,dt = \int_0^T y'e^{F(T-t)}bu(t)\,dt$$

$$= (A^*y \mid u)_{L_2}$$

where

$$A^*y = b'e^{F'(T-t)}y.$$

Also, AA^* is the $n \times n$ matrix,

$$AA^* = \int_0^T e^{F(T-t)}bb'e^{F'(T-t)}\,dt.$$

If the matrix AA^* is invertible, the optimal control can be found as

$$u = A^*(AA^*)^{-1}x_1.$$

6.11 Pseudoinverse Operators

We now develop a more general and more complete approach to the problem of finding approximate or minimum norm solutions to $Ax = y$. The approach leads to the concept of the pseudoinverse of an operator A.

Suppose again that G and H are Hilbert spaces and that $A \in B(G, H)$ with $\mathscr{R}(A)$ closed. (In applications the closure of $\mathscr{R}(A)$ is usually supplied by the finite dimensionality of either G or H).

Definition. Among all vectors $x_1 \in G$ satisfying

$$\|Ax_1 - y\| = \min_x \|Ax - y\|,$$

let x_0 be the unique vector of minimum norm. The *pseudoinverse* $A^\dagger$ of A is the operator mapping y into its corresponding x_0 as y varies over H.

To justify the above definition, it must be verified that there is a unique x_0 corresponding to each $y \in H$. We observe first that $\min_x \|Ax - y\|$ is achieved since this amounts to approximating y by a vector in the closed subspace $\mathscr{R}(A)$. The approximation $\hat{y} = Ax_1$ is unique, although x_1 may not be.

The set of vectors x_1 satisfying $Ax_1 = \hat{y}$ is a linear variety, a translation of the subspace $\mathscr{N}(A)$. Thus, since this variety is closed, it contains a unique x_0 of minimum norm. Thus $A^\dagger$ is well defined. We show below that it is linear and bounded.

The above definition of the pseudoinverse is somewhat indirect and algebraic. We can develop a geometric interpretation of $A^\dagger$ so that certain of its properties become more transparent.

According to Theorem 1, Section 3.4, the space G can be expressed as

$$G = \mathscr{N}(A) \oplus \mathscr{N}(A)^\perp.$$

Likewise, since $\mathscr{R}(A)$ is assumed closed,

$$H = \mathscr{R}(A) \oplus \mathscr{R}(A)^\perp.$$

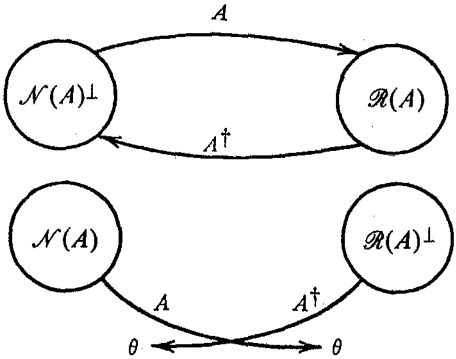

Figure 6.6 The pseudoinverse

The operator A restricted to $\mathcal{N}(A)^\perp$ can be regarded as an operator from the Hilbert space $\mathcal{N}(A)^\perp$ onto the Hilbert space $\mathcal{R}(A)$. Between these spaces A is one-to-one and onto and hence has a linear inverse which, according to the Banach inverse theorem, is bounded. This inverse operator defines $A^\dagger$ on $\mathcal{R}(A)$. Its domain is extended to all of H by defining $A^\dagger y = \theta$ for $y \in \mathcal{R}(A)^\perp$. Figure 6.6 shows this schematically and Figure 6.7 gives a geometric illustration of the relation of the various vectors in the problem.

It is easy to verify that this definition of $A^\dagger$ is in agreement with that of the last definition. Any $y \in H$ can be expressed uniquely as $y = \hat{y} + y_1$ where $\hat{y} \in \mathcal{R}(A)$, $y_1 \in \mathcal{R}(A)^\perp$. Thus $\hat{y}$ is the best approximation to y in $\mathcal{R}(A)$. Then $A^\dagger y = A^\dagger(\hat{y} + y_1) = A^\dagger \hat{y}$. Define $x_0 = A^\dagger y$. Then by definition $Ax_0 = \hat{y}$. Furthermore, $x_0 \in \mathcal{N}(A)^\perp$ and is therefore the minimum norm solution of $Ax_1 = \hat{y}$.

The pseudoinverse possesses a number of algebraic properties which are generalizations of corresponding properties for inverses. These properties are for the most part unimportant from our viewpoint of optimization; therefore they are not proved here but simply stated below.

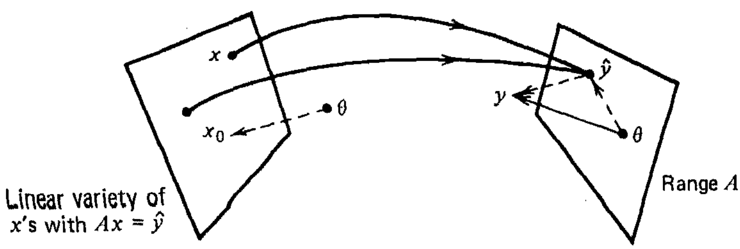

Figure 6.7 Relation between y and x_0

Proposition 1. *Let A be a bounded linear operator with closed range and let $A^\dagger$ denote its pseudoinverse. Then*

1. $A^\dagger$ *is linear.*
2. $A^\dagger$ *is bounded.*
3. $(A^\dagger)^\dagger = A$.
4. $(A^*)^\dagger = (A^\dagger)^*$.
5. $A^\dagger A A^\dagger = A^\dagger$.
6. $A A^\dagger A = A$.
7. $(A^\dagger A)^* = A^\dagger A$.
8. $A^\dagger = (A^* A)^\dagger A^*$.
9. $A^\dagger = A^*(AA^*)^\dagger$.

In certain limiting cases it is possible to give a simple explicit formula for $A^\dagger$. For instance, if A^*A is invertible, then $A^\dagger = (A^*A)^{-1}A^*$. If AA^* is invertible, then $A^\dagger = A^*(AA^*)^{-1}$. In general, however, a simple formula does not exist.

Example 1. The pseudoinverse arises in connection with approximation problems. Let $\{x_1, x_2, \dots, x_n\}$ be a set of vectors in a Hilbert space H. In this example, however, these vectors are not assumed to be independent. As usual, we seek the best approximation of the form $\hat{y} = \sum_{i=1}^n a_i x_i$ to the vector y. Or by defining $A : E^n \to H$ by

$$Aa = \sum_{i=1}^{n} a_i x_i,$$

we seek the approximation to $Aa = y$. If the vector a achieving the best approximation is not unique, we then ask for the a of smallest norm which gives $\hat{y}$. Thus

$$a_0 = A^\dagger y.$$

The computation of $A^\dagger$ can be reduced by Proposition 1 to

$$A^\dagger = (A^*A)^\dagger A^*$$

so that the problem reduces to computing the pseudoinverse of the $n \times n$ Gram matrix $A^*A = G(x_1, x_2, \dots, x_n)$.

6.12 Problems

1. Let $X = L_2[0, 1]$ and define A on X by

$$Ax = \int_0^1 K(t, s)x(s)\,ds$$

where

$$\int_0^1 \int_0^1 |K(t, s)|^2 \, dt \, ds < \infty.$$

Show that $A : X \to X$ and that A is bounded.

2. Let $X = L_p[0, 1]$, $1 < p < \infty$, $Y = L_q[0, 1]$, $1/p + 1/q = 1$. Define A by

$$Ax = \int_0^1 K(t, s)x(s) \, ds.$$

Show that $A \in B(X, Y)$ if $\int_0^1 \int_0^1 |K(t, s)|^q \, dt \, ds < \infty$.

3. Let A be a bounded linear operator from c_0 to l_∞. Show that corresponding to A there is an infinite matrix of scalars α_{ij}, $i, j = 1, 2, \ldots$, such that $y = Ax$ is expressed by the equations

$$\eta_i = \sum_{j=1}^{\infty} \alpha_{ij} \xi_j,$$

where $y = \{\eta_i\}$, $x = \{\xi_i\}$, and the norm of A is given by

$$\|A\| = \sup_i \sum_{j=1}^{\infty} |\alpha_{ij}|.$$

4. Prove the two-norm theorem: If X is a Banach space when normed by $\| \; \|_1$ and by $\| \; \|_2$ and if there is a constant c such that $\|x\|_1 \leq c \|x\|_2$ for all $x \in X$, then there is a constant C such that $\|x\|_2 \leq C \|x\|_1$ for all $x \in X$.
5. The *graph* of a transformation $T : X \to Y$ with domain $D \subset X$ is the set of points $(x, Tx) \in X \times Y$ with $x \in D$. Show that a bounded linear transformation with closed domain has a closed graph.
6. Let $X = C[a, b] = Y$ and let D be the subspace of X consisting of all continuously differentiable functions. Define the transformation T on D by $Tx = dx/dt$. Show that the graph of T is closed.
7. Prove the closed graph theorem: If X and Y are Banach spaces and T is a linear operator from X to Y with closed domain and closed graph, then T is bounded.
8. Show that a linear transformation mapping one Banach space into another is bounded if and only if its nullspace is closed.
9. Let H be the Hilbert space of n-tuples with inner product $(x \mid y)_Q = x'Qy$ where Q is a symmetric positive-definite matrix. Let an operator A on H be defined by an $n \times n$ matrix $[a_{ij}]$ in the usual sense. Find the matrix representation of A^*.

10. Let $X = L_p[0, 1]$, $1 < p < \infty$, $Y = L_q[0, 1]$, $1/p + 1/q = 1$. Let $A \in B(X, Y)$ be defined by $Ax = \int_0^1 K(t, s)x(s)\, ds$ where

$$\int_0^1 \int_0^1 |K(t, s)|^q \, dt \, ds < \infty.$$

(See Problem 2.) Find A^*.

11. Let X and Y be normed spaces and let $A \in B(X, Y)$. Show that ${}^\perp[\mathscr{R}(A^*)] = \mathscr{N}(A)$. (See Section 5.7.)

12. Let X, Y be Banach spaces and let $A \in B(X, Y)$ have closed range. Show that

$$\inf_{Ax=b} \|x\| = \max_{\|A^*y^*\| \leq 1} \langle b, y^* \rangle.$$

Use this result to reinterpret the solution of the rocket problem of Example 3, Section 5.9.

13. Let X and Y be normed spaces and let G be the graph in $X \times Y$ of an operator $A \in B(X, Y)$. Show that $G^\perp$ is the graph of $-A^*$ in $X^* \times Y^*$.

14. Prove the Minkowski-Farkas lemma: Let A be an $m \times n$ matrix and b an n-dimensional vector. Then $Ax \leq \theta$ implies $b'x \leq 0$ if and only if $b = A'\lambda$ for some m-dimensional vector $\lambda \geq \theta$. Give a geometric interpretation of this result. (Inequalities among vectors are to be interpreted componentwise.)

15. Let M be a closed subspace of a Hilbert space H. The operator P defined by $Px = m$, where $x = m + n$ is the unique representation of $x \in H$ with $m \in M$, $n \in M^\perp$, is called the projection operator onto M. Show that a projection operator is linear and bounded with $\|P\| = 1$ if M is at least one dimensional.

16. Show that a bounded linear operator on a Hilbert space H is a projection operator if and only if:

    1. $P^2 = P$ (idempotent)
    2. $P^* = P$ (self-adjoint).

17. Two projection operators P_1 and P_2 on a Hilbert space are said to be orthogonal if $P_1 P_2 = 0$. Show that two projection operators are orthogonal if and only if their ranges are orthogonal.

18. Show that the sum of two projection operators is a projection operator if and only if they are orthogonal.

19. Let G and H be Hilbert spaces and suppose $A \in B(G, H)$ with $\mathscr{R}(A)$ closed. Show that

$$A^\dagger = \lim_{\varepsilon \to 0+} (A^*A + \varepsilon I)^{-1} A^* = \lim_{\varepsilon \to 0+} A^*(AA^* + \varepsilon I)^{-1}$$

where the limits represent convergence in $B(H, G)$.

20. Find the pseudoinverse of the operator A on E^3 defined by the matrix

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

21. Let G, H, K be Hilbert spaces and let $B \in B(G, K)$ with range equal to K and $C \in B(K, H)$ with nullspace equal to $\{\theta\}$ (i.e., B is onto and C is one-to-one). Then for $A = CB$ we have $A^\dagger = B^\dagger C^\dagger$.

REFERENCES

§6.4. The Banach inverse theorem is intimately related to the closed graph theorem and the open mapping theorem. For further developments and applications, see Goffman and Pedrick [59].

§6.7. The concept of conjugate cone is related to various alternative concepts including dual cones, polar cones, and polar sets. See Edwards [46], Hurwicz [75], Fenchel [52], or Kelley, Namioka, et al. [85].

§6.10. See Balakrishnan [14].

§6.11. The concept of a pseudoinverse (or generalized inverse) was originally introduced for matrices by Moore [105], [106] and Penrose [117], [118]. See also Greville [67]. For an interesting introduction to the subject, see Albert [5]. Pseudoinverses on Hilbert space have been discussed by Tseng [146] and Ben Israel and Charnes [20]. Our treatment closely parallels that of Desoer and Whalen [38].