

## THE SOLUTION OF SINGULAR EQUATIONS, I: LINEAR EQUATIONS IN HILBERT SPACE

THOMAS I. SEIDMAN

**The theorem on which this paper is based is an easy generalization of the fact that the nullspace of an operator is the orthocomplement of (the closure of) the range of its adjoint. Its significance, here, is the observation that this may be applied to give a computationally feasible algorithm for the problem of the title.**

Consider a pair of hilbert spaces  $U, V$  and a linear transformation  $A: U \rightarrow V$ . For any  $b$  in the range of  $A$  there is, by definition, a solution of the equation:

$$(1) \quad Ax = b .$$

It is specifically *not* assumed here that the range of  $A$  is closed in  $V$  so, in general, the solution cannot be taken to depend continuously on  $b$ . Since, in many contexts, the inhomogeneous term  $b$  is known only through 'measurement'—hence, only to 'arbitrary but finite' accuracy—this lack of continuous dependence has heretofore been taken to preclude useful computation; see, e.g., Hadamard's discussion of a 'well-posed' problem [4]. There has been, however, considerable interest (some references are noted in §5) in computational approaches to various problems which are ill-posed in Hadamard's sense. Even for well-posed problems it has proved desirable to distinguish in principle (cf. [9], p. 224) between the notions of *solvability* (existence) and *approximation-solvability* (obtaining a solution — granting solvability — as a limit of solutions of finite dimensional problems).

Numerous examples of such singular situations as we consider might easily be adduced at this point: the backward heat equation, integral equations of the first kind, analytic continuation, inversion of the Laplace transform, etc. The particular application through which the author came to the algorithm described here concerned synthesis of a boundary null-control for the heat equation. This has already been discussed in greater detail elsewhere [2] but will be treated briefly here §6 as an example. The author would like to dedicate this paper to the memory of W. C. Chewning, who initiated that work, with thanks and in regret for his untimely death.

2. Basic theorem. We consider the following general setting:

- a pair of hilbert spaces  $U, V$ ,
- a densely defined linear map  $A: D \subset U \rightarrow V$ ,
- an element  $b \in R(A)$  so  $b = Ax_*$  for some  $x_* \in D$ ,
- a subspace  $Y$  of  $D_*$  (the domain of the adjoint  $A^*$ ).

We let  $X$  be the closure in  $U$  of  $A^*Y$  and let

$$S = \{x \in D: (Ax - b) \perp Y\}, S' = \{x \in U: \langle A^*y, x \rangle = \langle y, b \rangle \text{ for } y \in Y\}.$$

Let  $P$  be the orthoprojection onto  $X$ .

REMARKS. (1) It is important for the argument below that  $U$  be a hilbert space but, with minor (notational) modifications,  $V$  may be more general. A further generalization of the algorithm to be presented applies when  $U$  is also more general (a uniformly convex banach space) but this requires a more subtle convergence argument and will be discussed elsewhere [12].

(2) If, as is the case in each of the examples adduced above, the transformation  $A$  is continuous, one has  $D = U$  and  $D_* = V^*(= V)$  so the restriction:  $Y \subset D_*$  is automatically satisfied, as is the condition:  $X \subset D$  which appears in the Theorem below.

THEOREM I. *Let the setting be as described above. Then the following are equivalent:*

- (i)  $\hat{x} = Px_*$ ,
- (ii)  $\hat{x}$  is the nearest point in  $X$  to the solution  $x_*$ ,
- (iii)  $\hat{x} \in X \cap S'$ ,
- (iv)  $\hat{x}$  has minimal norm in  $S'$ .

If  $X \subset D$ , then  $S'$  may be replaced by  $S$  in (iii), (iv).

*Proof.* The equivalence of (i), (ii) is standard. Observe that  $x_* \in S \subset S'$  and that  $S'$  is a translate (e.g., by  $x_*$ ) of the closed subspace  $X^\perp (= (A^*Y)^\perp)$  since  $A^*Y$  is dense in  $X$ . Thus,

$$(3) \quad S' = x_* + X^\perp, \quad S = x_* + (X^\perp \cap D).$$

It follows that  $\hat{x} \in X \cap S'$  implies  $\hat{x} = P\hat{x} = Px_*$  so (iii) implies (i); conversely,  $P\hat{x} = \hat{x} = Px_*$  implies  $\hat{x} \in S' \cap X$ . This also shows that (iii) determines a unique point  $x_3$  (i.e.,  $S' \cap X = \{x_3\}$ ); if  $X \subset D$ , then  $x_3 \in D$  and  $X \subset S' = X \cap S$ . Now (3) implies that the minimum of  $\|x\|$  for  $x \in S'$  is attained for  $x = x_3 \in X$  so (iii), (iv) are equivalent. If  $X \subset D$ , this also shows that  $\|x\|$  actually attains a minimum in  $S$  even though  $S$  need not be closed.

3. Finite dimensional subspaces. The cases of practical (com-

putational) importance involve finite dimensional subspaces:  $Y = sp\{y_1, \dots, y_n\}$  so  $X = sp\{z_1, \dots, z_n\}$  with  $z_j = A^*y_j$  for each  $j$ .

It will be apparent that there is no loss of generality in assuming  $Y$  is such that  $\{z_1, \dots, z_n\}$  is linearly independent: i.e., the restriction of  $A^*$  to  $Y$  is nonsingular. It is now computationally easiest to work with the characterization (iii) of the Theorem — which here reduces to the finite set of conditions for  $\hat{x} \in X$ :

$$(4) \quad \langle z_j, \hat{x} \rangle = \langle y_j, b \rangle \quad \text{for } j = 1, \dots, n .$$

Note that if each  $z_j$  is in  $D$  (each  $y_j$  in the domain of  $AA^*$ ) so  $X \subset D$ , then (4) is equivalent to:

$$(4) \quad \langle y_j, A\hat{x} - b \rangle = 0 \quad \text{for } j = 1, \dots, n .$$

Expanding  $\hat{x} \in X$  with respect to the basis  $\{z_1, \dots, z_n\}$  (i.e., setting  $\hat{x} = \xi_1 z_1 + \dots + \xi_n z_n$ ) the conditions (4) provide  $n$  linear equations for the  $n$  unknown coefficients:

$$(5) \quad \begin{array}{r} \gamma_{1,1}\xi_1 + \dots + \gamma_{1,n}\xi_n = \beta_1 \\ \vdots \\ \gamma_{n,1}\xi_1 + \dots + \gamma_{n,n}\xi_n = \beta_n \end{array}$$

where

$$(6) \quad \begin{array}{l} \gamma_{j,k} = \langle z_j, z_k \rangle \quad (j, k = 1, \dots, n) , \\ \beta_j = \langle z_j, b \rangle \quad (j = 1, \dots, n) . \end{array}$$

Thus the matrix  $G = ((\gamma_{j,k}))$  is the Gramian matrix for the basis  $\{z_1, \dots, z_n\}$  of  $X$  and is clearly nonsingular.

From the nature of the formulation (5), (6) it is clear that the point  $\hat{x}$  determined thereby will be stable under small perturbations of the data and computational imprecisions even though the original transformation  $A$  may be singular. Indeed, it is this consideration which makes the method a useful approach to singular problems.

4. **Solution of singular problems.** In applying the discussion above to the computational solution of a (possibly singular) problem of the form (1), we consider the following general setting:

- a pair of hilbert spaces  $U, V$ ,
- a densely defined linear map  $A: D \subset U \rightarrow V$ ,
- an element  $b \in R(A)$  so  $b = Ax_*$  for some  $x_* \in D$ ,
- a sequence  $\{y_1, y_2, \dots\}$  in the domain of  $A^*$ .

For  $j = 1 = 1, 2, \dots$  we let  $z_j = A^*y_j$  and for each  $n = 1, 2, \dots$  we

let  $Y_n = sp\{y_1, \dots, y_n\}$ ,  $X_n = sp\{z_1, \dots, z_n\}$  and define  $x_n$  as the (unique) solution in  $X_n$  of the conditions (4).

**THEOREM II.** *Let the setting be as above and suppose that  $A$  is closed and that  $A, \{y_1, y_2, \dots\}$  are such that  $N(A) \subset \overline{sp}\{z_1, z_2, \dots\}^\perp$  so:*

$$(7) \quad \langle z_j, x \rangle = 0 (j = 1, 2, \dots) \implies x \in N(A).$$

*Then:*

- (i) for each  $n = 1, 2, \dots$  the conditions (4) determine a unique  $x_n \in X_n$ ,
- (ii) the sequence  $(x_1, x_2, \dots)$  so defined converges to a limit,  $\hat{x}$ ,
- (iii) the limit  $\hat{x}$  is in  $D$  and satisfies (1)  $A\hat{x} = b$ . If  $N(A)$  is nontrivial,  $\hat{x}$  will be the unique solution of (1) having minimum norm.

**REMARK.** Observe that if  $A$  is not only closed but *continuous*, then the condition:  $\{y_j\}$  in  $D_*$  is automatic and (7) is equivalent to requiring that  $sp\{N(A^*), y_1, y_2, \dots\}$  be dense in  $V$  (e.g., one might take  $\{y_1, y_2, \dots\}$  to be any (orthonormal) basis for  $V$ ) so  $sp\{z_1, z_2, \dots\}$  is dense in  $\overline{R}(A^*)$ .

*Proof* (of Theorem II). The unique determination of each  $x_n$  by (4) follows from Theorem I and the discussion of the finite dimensional case. Let  $Y$  be the closure in  $V$  of  $\bigcup_n Y_n$  (so  $Y = \overline{sp}\{y_1, y_2, \dots\}$ ) and let  $X$  be the closure in  $U$  of  $\bigcup_n X_n$  (so  $X = sp\{z_1, z_2, \dots\}$ ). It follows from Theorem I that there is a unique point  $\hat{x} = Px_*$  ( $P$  is the ortho-projection onto  $X$ ) satisfying:  $\langle A^*y, x \rangle = \langle y, b \rangle$  for all  $y \in Y$  and so, in particular, satisfying (4) for  $j = 1, 2, \dots$ . Setting  $x = (\hat{x} - x_*)$ , one has

$$\langle z_j, x \rangle = \langle A^*y_j, \hat{x} \rangle - \langle y_j, Ax_* \rangle = 0 \quad (j = 1, 2, \dots)$$

so, from (7),  $(\hat{x} - x_*) = x \in N(A) \subset D$ . Since  $x_* \in D$ , this means  $\hat{x} \in D$  and  $A\hat{x} = Ax_* = b$ . The characterization of  $\hat{x}$  when  $N(A) \neq \{0\}$  follows from the characterization (iv) of Theorem I. From the characterization (ii) in Theorem I, it follows that  $x_n$  is the nearest point in  $X_n$  to  $\hat{x}$  (we may now take  $x_* = \hat{x}$ ), so, as  $\bigcup_n X_n$  is dense in  $N(A)^\perp$ , one must have  $x_n \rightarrow \hat{x}$ .

**5. Discussion.** A variety of methods have been applied to the computational solution of ill-posed problems. We mention, in particular, the papers of Krasnoselskii [5] and Kryanev [6], Tikhonov 'regularization' method [13], [3], the work of Nashed and Wahba [8], and the methods of Carasso [1] and of Lattes and Lions [7] for

backward evolution equations. In addition, even in the absence of theoretical justification there has been extensive use, in computational practice, of 'standard' approaches — e.g., choosing the approximant  $x_n$  to minimize  $\|Ax - b\|$  over a subspace  $X_n$  or Galerkin's method — applied to the solution of ill-posed problems.

We observe here, in comparison with Galerkin's method, that the presently proposed algorithm has the inconvenience that the approximating subspaces  $X_n$  are not specified directly but are computed by applying  $A^*$  to specifiable subspaces  $Y_n$ . One might, of course, attempt to reverse this by specifying  $\{z_1, \dots\}$  and then solving:  $A^*y_j = z_j$  to obtain each  $y_j$ ; unfortunately, if the original problem is singular ( $A^{-1}$  unbounded) then each of these equations also involves solution of a singular problem and, in any case, one would be restricted by the necessity of taking  $z_j \in R(A^*)$ .

Some additional freedom can be introduced by varying the inner product. If the original  $\langle \cdot, \cdot \rangle$  is replaced by a new inner product  $\langle \cdot, \cdot \rangle' = \langle \cdot, B \cdot \rangle$ , defined in terms of a given positive operator  $B$ , then the new adjoint  $A_B^*$  is just  $B^{-1}A^*$  so, now,  $A^*y_j = Bz_j$ . Observe that, if  $A$  itself is positive (a point of relevance of the typical imposition of monotonicity conditions in connection with the use of Galerkin's method), then one may take  $B = A^*(=A)$  and so  $y_j = z_j$  for each  $j$ . It is easily seen that in this case the present method and Galerkin's method coincide. Needless to say, the  $x_n$  computed using  $\langle \cdot, \cdot \rangle'$  is now the best approximant to  $x_*$  in the sense of the new norm  $\|\cdot\|'$  and one obtains convergence in  $U$  from this only if  $\|\cdot\|'$  dominates the original norm, i.e., if  $B$  has a bounded inverse. Note that even if  $A$  is not positive one may replace (1) by

$$(8) \quad (A^*A)x = b_*$$

with  $b_* = A^*b$  (provided  $b \in D_*$ ), which will be equivalent to (1) if  $R(A)$  is dense so  $A^*$  is injective. One could then work, instead, with the positive operator  $A^*A$ .

It will be noted that Theorem II proves convergence but gives no *direct* information as to the rate of convergence. All that can be said about this is given by the characterization of  $x_*$  as the best approximant in  $X_n$  to the solution  $x_*$ . If one considers an example with  $\{z_1, z_2, \dots\}$  an orthonormal basis of  $U$  and  $x_* = \xi_1 z_1 + \xi_2 z_2 + \dots$  selected appropriately (defining  $b$ ), it is clear that convergence can be arbitrarily slow. To do better, one must know that, for the particular  $\{y_1, y_2, \dots\}$  considered and under whatever 'regularity' properties and *a priori* bounds as may apply to the specific situation considered, the solution  $\hat{x}$  will be approximable in  $X_n$  with satisfactory accuracy.

An approach which may be used, on occasion, to provide estimates of convergence rates is to observe that one often has regularity conditions in the form of a stronger existence assertion: not only is  $b \in R(A)$  but  $b \in R(AC)$ , for some operator  $C$ , so  $x_*$  has the form:  $Cu_*$  and (1) is replaced by

$$(9) \quad (AC)u = b.$$

Applying the same algorithm to solving this gives  $\tilde{z}_j = (AC)^*y_j = C^*z_j$  and (4) gives

$$\begin{aligned} \langle y_j, b \rangle &= \langle \tilde{z}_j, \hat{u} \rangle = \langle C^*z_j, \hat{u} \rangle \\ &= \langle z_j, C\hat{u} \rangle \end{aligned}$$

so  $x_n = Cu_n$  for each  $n$ . If  $C$  is compact (it is typically the embedding operator of one Sobolev space in another), then the convergence of the sequence guaranteed by Theorem II may provide convergence at a usefully estimable (e.g., in terms of  $\|\hat{u}\|$ ) rate of the sequence  $\{x_n\}$ . A related approach is used in [3]. Estimation of the convergence rates for some specific applications will appear elsewhere.

**6. An application.** Let  $W$  be the space of functions  $w = w(t, \omega)$  for  $0 \leq t \leq T$  and  $\omega \in \Omega$  which satisfy the diffusion equation:

$$(10) \quad w_t = \nabla \cdot p \nabla w$$

( $p = p(t, \omega) > 0$ ) and for which

$$(11) \quad z = w(0, \cdot) \in L_2(\Omega), \quad x = w|_{\Sigma} \in L_2(\Sigma)$$

(here,  $\Omega$  is assumed bounded in  $R^n$  with  $\Sigma = (0, T) \times \partial\Omega$  smooth). Letting  $v = w(T, \cdot) \in L_2(\Omega)$ , one has  $v = Ax + Sz$  with  $A: U = L_2(\Sigma) \rightarrow V = L_2(\Omega)$  and  $S: V \rightarrow V$  both bounded linear transformations;  $S = S_T$  is the solution operator for the pure initial value problem (i.e., for homogeneous Dirichlet conditions) and  $A$  gives the effect of the boundary data for homogeneous initial conditions.

Given  $z, v$  one can view the situation as a control problem and, in particular, for any  $y \in V$  one might seek a null-control: a set of Dirichlet data  $x \in U$  for which the solution of (10), (11) vanishes at  $T$  ( $v = 0$ ). This amounts to solving the equation:  $Ax = b$  ( $b = -Sz$ ). It is known (cf, e.g., [10] or [11]) that  $R(S) \subset R(A)$  so a solution exists. Indeed, the map: [ $y \mapsto$  nullcontrol  $x$  of minimum norm] is a bounded linear map:  $V \rightarrow U$ , although not directly computationally accessible while  $A, S$  are easily available in standard computational practice.

Given  $y \in V^*(V)$ , let  $u$  be the solution of the well-posed problem

$$(12) \quad -u_t = \nabla \cdot p \nabla u, \quad u|_{\Sigma} = 0, \quad u(T, \cdot) = y.$$

Then

$$\begin{aligned} \int_{\Omega} (Ax)y &= \int_0^T \int_{\Omega} (wu)_t \\ &= \int_0^T \int_{\Omega} [\nabla \cdot p \nabla w]u - w[\nabla \cdot p \nabla u] \\ &= \int_{\Sigma} p[-w_{\nu}u + wu_{\nu}] = \int_{\Sigma} pxu_{\nu} \end{aligned}$$

Thus  $A^*y = pu_{\nu}$ , where  $u_{\nu} \in L_2(\Sigma)$  is the normal derivative of the solution of (12).

To compute the approximant  $\hat{x}_n$  to the desired boundary null-control  $\hat{x}$ , one would start with  $\{y_1, \dots, y_n\}$ , solve (12) numerically using each  $y_j$  as data, and compute each  $z_j = A^*y_j = p\partial u_j/\partial\nu$ . The Gramian  $G = ((\langle z_j, z_k \rangle))$  can now be obtained by numerical integration over  $\Sigma$ . To improve the conditioning of the eventual numerical solution of (5), it may be worthwhile to orthonormalize (approximately)  $\{z_1, \dots, z_n\}$ , reflect the relations of the new basis of  $X_n$  to the old in re-selecting new  $\{\tilde{y}_1, \dots, \tilde{y}_n\}$  and then, rather than simply accepting the orthonormalized  $\{z_j\}$ , re-computing an (almost) orthonormal basis  $\{\tilde{z}_1, \dots, \tilde{z}_n\}$  from  $\{\tilde{y}_1, \dots, \tilde{y}_n\}$ . The term  $B = -Sz$  is similarly obtained by numerical solution of (10) and (11) with homogeneous boundary data and then the  $\{\beta_j\}$  computed by numerical integration over  $\Omega$ . The resulting system (5) can now be solved to obtain the coefficients  $\{\xi_j\}$  and so  $\hat{x}_n$ . Note that the *computed*  $\hat{x}_n$  approximates the *true*  $\hat{x}_n$  as closely as desired provided the numerical solution of (10), (11) and of (12), the numerical integrations and the solution of (5) are each with sufficient accuracy and, in turn, this will be a good approximant to the solution  $\hat{x}$  if  $n$  is large enough. An actual computational experiment is described in [2].

#### REFERENCES

1. B. C. Buzbee and A. Carasso, *On the numerical computation of parabolic problems for preceding times*, Math. Comp., **27** (1973), 237-266.
2. W. C. Chewning and T. I. Seidman, *A convergent scheme for boundary control of the heat equation*, SIAM J. Control, to appear.
3. Joel Franklin, *On Tikhonov's method for ill-posed problems*, Math. of Comp., **28** (1974), 889-907.
4. J. Hadamard, *Lectures on Cauchy's Problem in Linear P.D.E.* (sec. 21), 1923, reprinted by Dover Publ., N.Y. (1952).
5. M. A. Krasnoselskii, *Solutions of equations involving adjoint operators by successive approximations*, Usp. Mat. Nauk, **15** (1960).
6. A. V. Kryanev, *The solution of incorrectly posed problems by methods of successive approximations*, Soviet Math. Doklady, **14** (1973), 673-676.

7. R. Lattes and J. L. Lions, *The Method of Quasi-Reversibility, Applications to Partial Differential Equations*, Amer. Elsevier, N. Y., 1969.
8. M. Z. Nashed and G. Wahba, *Regularization and approximation of linear operator equations in reproducing kernel spaces*, Bull. Amer. Math. Soc., **80** (1974).
9. W. V. Petryshyn, *On the approximation solvability of equations involving A-proper mappings*, Bull. Amer. Math. Soc., **81** (1975), 223-312.
10. D. L. Russell, *A unified boundary controllability theory for hyperbolic and parabolic partial differential equations*, Studies in Appl. Math., **52** (1973), 189-211.
11. T. I. Seidman, *Observation and prediction for the heat equation, III*, J. Differential Equations, to appear.
12. ———, *Solution of singular equations, II: nonlinear problems*, to appear.
13. A. N. Tikhonov, *Solution of incorrectly formulated problem and the regularization method.*, (trans.) Sov. Math. Dokl., **4** (1963), 1036-1038.

Received April 17, 1975.

UNIVERSITY OF MARYLAND, BALTIMORE COUNTY