

Asymptotic Theory of Filtering for Linear Operator Equations with Discrete Noisy Data*

By C. W. Groetsch and C. R. Vogel

Abstract. We consider Fredholm integral equations of the first kind with continuous kernels in which the data is discretely sampled and contaminated by white noise. A sufficient condition for the convergence of general filtering methods applied to such equations is derived. The condition in essence relates the decay rate of the singular values of the integral operator to the shape of the filter function used in the regularization method. Specific illustrations of the condition are given for Tikhonov regularization, the truncated singular value decomposition, and Landweber iteration.

1. Introduction. Many inverse problems of mathematical physics and problems of remote sensing lead to the solution of a Fredholm integral equation of the first kind,

$$(1) \quad \int_a^b k(s, t)x(t) dt = g(s), \quad c \leq s \leq d,$$

where $k(\cdot, \cdot)$ is a given square integrable kernel and g is a given function called the "data". (See Bertero, De Mol, and Viano [2] for an excellent introductory survey and general formulation of linear inverse problems in mathematical physics.) The major difficulty associated with the solution of such equations arises from the fact that they are *ill-posed*; specifically, the solution x , even if it exists and is unique, generally depends discontinuously on the data g . Since the data in practical circumstances are measured quantities, measurement errors can be magnified by the solution process and can produce large instabilities in the computed solution. In an attempt to cope with these instabilities, an extensive theory of "regularization" has been developed (see, e.g., [1], [6], and [5] for surveys). The idea of regularization methods is to replace the ill-posed problem (1) by a slightly perturbed well-posed problem which depends on a small positive parameter. One then wishes to show that the solutions of these well-posed problems converge to the solution of (1) as the parameter converges to zero in an appropriate fashion.

It is convenient to study Eq. (1) in the abstract form

$$(2) \quad Kx = g$$

where $K: H_1 \rightarrow H_2$ is a given compact linear operator from a real Hilbert space H_1 into a real Hilbert space H_2 (the inner product and norm in H_i ($i = 1, 2$) will be denoted by $(\cdot, \cdot)_i$ and $\|\cdot\|_i$, respectively). We assume that (2) has a solution and we

Received September 5, 1986.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 45L05, 65J10.

*Partially supported by National Science Foundation grants INT 8510037 and DMS 8602000.

seek the minimum norm solution, x . It then follows that

$$(3) \quad x = K^\dagger g = \sum_{i=1}^{\infty} \mu_i^{-1}(g, \psi_i)_2 \varphi_i$$

where K^\dagger is the Moore-Penrose generalized inverse of K and $\{\varphi_i, \psi_i; \mu_i\}$ is a singular system for K , that is, $\{\varphi_i; \lambda_i\}$ and $\{\psi_i; \lambda_i\}$ are orthonormal eigensystems (with $\lambda_i = \mu_i^2 > 0$) for the operators K^*K and KK^* , respectively, where K^* is the adjoint of K (see, e.g., [5]). We assume the singular values are ordered so that $\mu_1 \geq \mu_2 \geq \dots > 0$. It is clear from (3) that noise in the data connected with singular components associated with small singular values can be greatly amplified in computing the solution x . The deleterious effects of this high-frequency noise can be mitigated by introducing “filtered” approximations

$$(4) \quad x_\alpha = \sum_{i=1}^{\infty} w_\alpha(\lambda_i) \mu_i^{-1}(g, \psi_i)_2 \varphi_i \quad (\alpha > 0),$$

where w_α is a filter function satisfying $0 \leq w_\alpha(\lambda) \leq 1$ and $w_\alpha(\lambda) \rightarrow 1$ as $\alpha \rightarrow 0$ for each $\lambda > 0$. The approximations x_α given in (4) are idealized approximations in the sense that the exact data function g , rather than a noisy version of g , is used. For these idealized approximations it can be shown that $x_\alpha \rightarrow x$ as $\alpha \rightarrow 0$ (see [5], where the corresponding notation is $R_\alpha(\lambda) = w_\alpha(\lambda)/\lambda$). In the case when only a noisy version \hat{g} of the data is available, the corresponding approximation \hat{x}_α is formed by replacing g with \hat{g} in (4). In the deterministic theory of regularization, the regularization parameter α must then be chosen in such a way that $\alpha \rightarrow 0$ and $\hat{x}_\alpha \rightarrow x$ as the error level in the data converges to zero. In this note we take a time series approach to general regularization. That is, we assume the data are discretely sampled and contaminated with white noise. The goal is then to relate the regularization parameter α to the data rate n (the number of data points sampled) in such a way that $\alpha \rightarrow 0$ as $n \rightarrow \infty$ and \hat{x}_α converges in expectation to x .

The most commonly used regularization method is Tikhonov regularization. In this method the approximation x_α is taken to be the minimizer of the Tikhonov functional

$$\|Ku - g\|^2 + \alpha \|Tu\|^2 \quad (\alpha > 0)$$

where T is a regularizing operator (in the simplest case, $T = I$, and in general, T involves derivatives) whose function is to damp out instabilities in the computed solutions. The norms in the Tikhonov functional above are L^2 -norms and the minimizer of the functional x_α satisfies the Euler equation

$$(K^*K + \alpha T^*T)x_\alpha = K^*g,$$

where the adjoints are L^2 -adjoints. If we assume that $\|Tx\|$ is a norm, then

$$(x, y)_1 = (Tx, Ty)$$

defines an inner product, and it is easy to see that the Euler equation above is equivalent to

$$(K^*K + \alpha I)x_\alpha = K^*g,$$

where the adjoints now are computed with respect to the inner product $(\cdot, \cdot)_1$ (see e.g. [5], [6]). The Tikhonov approximations now take the form

$$x_\alpha = (K^*K + \alpha I)^{-1}K^*g,$$

which corresponds to the filter function $w_\alpha(\lambda) = \lambda/(\lambda + \alpha)$, where α is a positive parameter.

Another widely used method is the truncated singular value expansion

$$x_m = \sum_{i=1}^m \mu_i^{-1}(g, \psi_i)_2 \varphi_i.$$

In this case, the regularization parameter is a positive integer $m = [1/\alpha]$ and the filter function satisfies

$$w_m(\lambda) = \begin{cases} 0, & \lambda < \lambda_{m+1}, \\ 1, & \lambda \geq \lambda_m. \end{cases}$$

Iterative methods, such as Landweber iteration,

$$x^{(m)} = (I - K^*K)x^{(m-1)} + K^*g, \quad x^{(0)} = 0$$

(we assume here that the problem has been scaled so that $\|K\| \leq 1$), where

$$w_m(\lambda) = 1 - (1 - \lambda)^m, \quad m = 0, 1, 2, \dots,$$

also fit into this scheme. (See also Anderssen and Prenter [1] for the relationship between Landweber iteration and Wiener filtering.) Note that in each of these examples the filter function is a monotonically increasing function of λ for each fixed α (or m).

Our interest is in the convergence of general filtering methods when the data are discretely sampled and noisy. In the case of Tikhonov regularization for kernels with eigenvalues having polynomial decay rates, results of this type were originally given by Wahba [13] and have been sharpened by Lukas [7] and, for convolution equations, by Davies and Anderssen [3]. See also the technical report [9] by Nychka and Cox. Vogel [12] has analyzed the case of truncated singular value expansions.

2. Discretely Sampled Noisy Data. In practical problems the available data consists of discretely sampled values which are contaminated by error. Suppose that the sampling points are $s_1 < s_2 < \dots < s_n$ and the measured data are represented by a vector \mathbf{d} , where

$$d_i = g(s_i) + \varepsilon_i, \quad 1 \leq i \leq n,$$

and the error $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is discrete white noise, that is,

$$E(\varepsilon_i) = 0 \quad \text{and} \quad E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2, & i = j \\ 0, & i \neq j \end{cases} \quad (1 \leq i, j \leq n),$$

where E is the expectation functional. Let Y_n represent the Hilbert space consisting of \mathbb{R}^n endowed with the inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n u_i v_i / n$$

and define the operator $K_n: L^2[a, b] \rightarrow Y_n$ by

$$(K_n x)_i = (Kx)(s_i) = \int_a^b k(s_i, t) x(t) dt.$$

The L^2 -inner product and norm will be denoted (\cdot, \cdot) and $\|\cdot\|$, respectively. Note that the adjoint of the discrete-valued operator K_n is the operator $K_n^*: Y_n \rightarrow L^2[a, b]$ given by

$$K_n^* \mathbf{u} = \frac{1}{n} \sum_{i=1}^n k(s_i, \cdot) u_i.$$

The discrete, error-contaminated model for (1) may then be written

$$\mathbf{d} = K_n x + \epsilon.$$

If $\{\varphi_{in}, \psi_{in}, \mu_{in}\}_{i=1}^n$ is a singular system for K_n , then

$$K_n^\dagger \mathbf{v} = \sum_{i=1}^n \mu_{in}^{-1} \langle \mathbf{v}, \psi_{in} \rangle \varphi_{in}$$

for each $\mathbf{v} \in Y_n$. According to a result of Nashed and Wahba [8], if the kernel $k(\cdot, \cdot)$ is continuous and the union of the sampling points $\bigcup_{n=1}^\infty \{s_i\}_{i=1}^n$ is dense in $[c, d]$, then

$$(5) \quad \|K_n^\dagger \mathbf{g} - K^\dagger g\| \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where $\mathbf{g} = (g(s_1), \dots, g(s_n))^T$. We shall study the convergence of the general filtered discrete approximations

$$x_{n,\alpha} = \sum_{i=1}^n w_\alpha(\lambda_{in}) \mu_{in}^{-1} \langle \mathbf{d}, \psi_{in} \rangle \varphi_{in}$$

with contaminated data vector $\mathbf{d} = K_n x + \epsilon$. The analysis will depend upon the decay rate λ_i of the eigenvalues and the shape of the filter function w_α . Under appropriate assumptions we show that the condition

$$\sum_{i=1}^n w_\alpha(\lambda_i)^2 / \lambda_i = o(n) \quad \text{as } \alpha \rightarrow 0, n \rightarrow \infty$$

is sufficient to guarantee the convergence of the approximations $x_{n,\alpha}$ to $x = K^\dagger g$, where λ_i are assumed decay rates for the eigenvalue λ_{in} (see below). Sufficient conditions for convergence of particular regularization methods (i.e., particular w_α) can then be given for various assumed decay rates.

Basic to our analysis is that the eigenvalues and coefficients of the “discrete” operator $K_n^* K_n$ satisfy:

$$\lambda_{in} \rightarrow \lambda_i \quad \text{and} \quad (x, \varphi_{in}) \rightarrow (x, \varphi_i) \quad \text{as } n \rightarrow \infty$$

(see Vogel [12] for a heuristic argument and Spence [10] for a more rigorous treatment). Specifically, we shall assume that

$$(6) \quad \begin{aligned} \lambda_{in} &= \beta_{in} \lambda_i, & 0 < a \leq \beta_{in} \leq b < \infty, \\ (x, \varphi_{in}) &= \gamma_{in} (x, \varphi_i), & |\gamma_{in}| \leq c < \infty, \end{aligned}$$

where $\beta_{in} \rightarrow 1$ and $\gamma_{in} \rightarrow 1$ as $n \rightarrow \infty$ ($\beta_{in} = 0$ for $i > n$). For the case of polynomial decay, the first condition corresponds to the assumptions of Wahba [13] (see also Lukas [7] and Davies and Anderssen [3]). We shall not assume a specific form for the decay rates λ_i or the coefficients (x, φ_i) . The convergence condition for $x_{n,\alpha}$ is that the decay rate of the eigenvalues is related to the filter function and data rate as expressed in Theorem 2 below.

We begin our analysis by introducing the expected square error in the filtered discrete approximations:

$$T_n(\alpha) := E \|x_{n,\alpha} - K_n^\dagger \mathbf{g}\|^2.$$

The first step is to show that $T_n(\alpha)$ can be decomposed into deterministic and random parts.

LEMMA. We have $T_n(\alpha) = D_n(\alpha) + R_n(\alpha)$, where

$$D_n(\alpha) = \sum_{i=1}^n [1 - w_\alpha(\lambda_{i_n})]^2 \langle \mathbf{g}, \psi_{i_n} \rangle^2 / \lambda_{i_n}$$

and

$$R_n(\alpha) = \frac{\sigma^2}{n} \sum_{i=1}^n w_\alpha(\lambda_{i_n})^2 / \lambda_{i_n}.$$

Proof. Let $\psi_{i_n} = (\psi_{i_n}^{(1)}, \dots, \psi_{i_n}^{(n)})^T$; then by the assumptions on the noise,

$$E(\langle \boldsymbol{\varepsilon}, \psi_{i_n} \rangle) = \sum_{j=1}^n \psi_{i_n}^{(j)} E(\varepsilon_j) / n = 0$$

and

$$E(\langle \boldsymbol{\varepsilon}, \psi_{i_n} \rangle^2) = \sum_{j=1}^n \sum_{k=1}^n \psi_{i_n}^{(j)} \psi_{i_n}^{(k)} E(\varepsilon_j \varepsilon_k) / n^2 = \sigma^2 / n.$$

Note that

$$\begin{aligned} K_n^\dagger \mathbf{g} - x_{n,\alpha} &= \sum_{i=1}^n [1 - w_\alpha(\lambda_{i_n})] \mu_{i_n}^{-1} \langle \mathbf{g}, \psi_{i_n} \rangle \varphi_{i_n} \\ &\quad - \sum_{i=1}^n w_\alpha(\lambda_{i_n}) \mu_{i_n}^{-1} \langle \boldsymbol{\varepsilon}, \psi_{i_n} \rangle \varphi_{i_n}, \end{aligned}$$

hence by the orthonormality of the eigenfunctions and the results on expectations noted above,

$$\begin{aligned} T_n(\alpha) &= E \left(\|x_{n,\alpha} - K_n^\dagger \mathbf{g}\|^2 \right) \\ &= \sum_{i=1}^n [1 - w_\alpha(\lambda_{i_n})]^2 \langle \mathbf{g}, \psi_{i_n} \rangle^2 / \lambda_{i_n} + \frac{\sigma^2}{n} \sum_{i=1}^n w_\alpha(\lambda_{i_n})^2 / \lambda_{i_n} \\ &= D_n(\alpha) + R_n(\alpha). \quad \square \end{aligned}$$

THEOREM 1. If $\alpha \rightarrow 0$ as $n \rightarrow \infty$, then $D_n(\alpha) \rightarrow 0$.

Proof. Since

$$\langle \mathbf{g}, \psi_{i_n} \rangle^2 / \lambda_{i_n} = \langle K_n x, \psi_{i_n} \rangle^2 / \lambda_{i_n} = (x, K_n^* \psi_{i_n})^2 / \lambda_{i_n} = (x, \varphi_{i_n}),$$

we have by (6)

$$D_n(\alpha) = \sum_{i=1}^n [1 - w_\alpha(\lambda_{i_n})]^2 (x, \varphi_{i_n})^2 \leq c^2 \sum_{i=1}^{\infty} d_{n_i}(x, \varphi_i)^2,$$

where $d_{ni} = [1 - w_\alpha(\lambda_{in})]^2$ for $i \leq n$ and $d_{ni} = 0$ for $i > n$. However, by (6) and the assumptions on the filter function w_α , we have for each i ,

$$1 = \lim_{n \rightarrow \infty} w_\alpha(a\lambda_i) \leq \lim_{n \rightarrow \infty} w_\alpha(\lambda_{in}) \leq \lim_{n \rightarrow \infty} w_\alpha(b\lambda_i) = 1$$

and hence $d_{ni} \rightarrow 0$ for each i as $n \rightarrow \infty$. Also,

$$d_{ni}(x, \varphi_i)^2 \leq (x, \varphi_i)^2 \quad \text{and} \quad \sum_{i=1}^n (x, \varphi_i)^2 \leq \|x\|^2 < \infty$$

and therefore $D_n(\alpha) \rightarrow 0$ as $n \rightarrow \infty$, by the dominated convergence theorem. \square

THEOREM 2. *If $\alpha \rightarrow 0$ as $n \rightarrow \infty$ in such a way that*

$$\sum_{i=1}^n w_\alpha(\lambda_i)^2 / \lambda_i = o(n) \quad \text{as } n \rightarrow \infty,$$

then $R_n(\alpha) \rightarrow 0$.

Proof. From the monotonicity of the filter function and (6) we have

$$\frac{w_\alpha(\lambda_{in})^2}{\lambda_{in}} \leq \frac{w_\alpha(b\lambda_i)^2}{a\lambda_i}$$

and hence

$$\frac{n}{\sigma^2} R_n(\alpha) = \sum_{i=1}^n w_\alpha(\lambda_{in})^2 / \lambda_{in} \leq \frac{1}{a} \sum_{i=1}^n w_\alpha(b\lambda_i)^2 / \lambda_i.$$

By scaling the original equation by $b^{-1/2}$, it follows that

$$\sum_{i=1}^n w_\alpha(\lambda_i)^2 / \lambda_i = o(n)$$

is sufficient to give $R_n(\alpha) \rightarrow 0$ as $n \rightarrow \infty$. \square

Using the result (5) of Nashed and Wahba and the theorems above we have

COROLLARY. *If $\alpha \rightarrow 0$ and $\sum_{i=1}^\infty w_\alpha(\lambda_i)^2 / \lambda_i = o(n)$, as $n \rightarrow \infty$, then $E(\|K^\dagger g - x_{n,\alpha}\|) \rightarrow 0$ as $n \rightarrow \infty$.*

3. Some Illustrations. We now interpret the results above for some particular regularization methods and eigenvalue decay rates. We consider first the case of “polynomial decay”, that is, $\lambda_i \approx i^{-p}$ for some $p > 0$, meaning $c_1 i^{-p} \leq \lambda_{in} \leq c_2 i^{-p}$ for some positive constants c_1 and c_2 . By the argument in Theorem 2 we see it is enough to consider $\lambda_i = i^{-p}$.

For the case of Tikhonov regularization we have $w_\alpha(\lambda) = \lambda / (\lambda + \alpha)$ and hence

$$\sum_{i=1}^n w_\alpha(\lambda_i)^2 / \lambda_i \leq \int_0^\infty t^{-p} / (t^{-p} + \alpha)^2 dt = \alpha^{-(p+1)/p} lp$$

where $lp = \int_0^\infty u^p / (1 + u^p)^2 du < \infty$ for $p > 1$. Therefore, a sufficient condition for convergence is $\alpha^{-1} = o(n^{p/(p+1)})$. This condition is satisfied when the regularization parameter is chosen by generalized cross validation (see [13]) or by the discrepancy principle (see [3]).

For the truncated singular value decomposition we have

$$\sum_{i=1}^{\infty} w_m(\lambda_i)^2/\lambda_i = \sum_{i=1}^m w_m(\lambda_i)^2/\lambda_i \leq \int_0^m t^p dt = m^{p+1}/(p+1),$$

and hence we obtain the sufficient condition $m = o(n^{1/(p+1)})$ (see Vogel [12]) for the convergence of this method.

In the case of Landweber iteration

$$\begin{aligned} \sum_{i=1}^{\infty} w_m(\lambda_i)^2/\lambda_i &\leq \int_1^{\infty} t^p [1 - (1 - t^{-p})^m]^2 dt \\ &\leq \int_0^1 [1 - u^m]^2 / (1 - u)^{2+1/p} du \leq m^2 lp/p, \end{aligned}$$

where $lp = \int_0^1 (1 - u)^{-1/p} du < \infty$ for $p > 1$. Hence a sufficient condition for convergence is $m = o(n^{1/2})$.

One can also consider the case of "exponential decay": $\lambda_i \approx \exp(-pi)$. For this decay rate, and Tikhonov's method, we have

$$\sum_{i=1}^{\infty} w_{\alpha}(\lambda_i)^2/\lambda_i \leq \int_0^{\infty} \frac{e^{pt}}{(1 + \alpha e^{pt})^2} dt = \frac{1}{\alpha p} \int_{\alpha}^{\infty} \frac{du}{(1 + u)^2} = \frac{1}{p\alpha(1 + \alpha)},$$

and hence $\alpha^{-1} = o(n^{\nu})$ for some ν with $0 < \nu \leq 1$ is sufficient to guarantee that $R_n(\alpha) \rightarrow 0$ as $n \rightarrow \infty$.

By similar arguments it is easy to show that for exponential decay the condition $e^m = o(n^{1/p})$ is sufficient for convergence of the truncated singular value decomposition and $m = o(n^{1/2})$ is sufficient for convergence of Landweber iteration. These results for exponential decay apply in particular to the problem of harmonic continuation in the unit disk in which the eigenvalues are $\lambda_i = r^i$ where $0 < r < 1$ (see Franklin [4]).

Acknowledgments. This work was completed at the Centre for Mathematical Analysis at the Australian National University. The authors wish to thank the Centre, and particularly Dr. Bob Anderssen, for their kind hospitality. They also thank Dr. Anderssen for a number of helpful comments on this paper.

Department of Mathematical Sciences
University of Cincinnati
Cincinnati, Ohio 45221

Department of Mathematical Sciences
Montana State University
Bozeman, Montana 59717

1. R. S. ANDERSSSEN & P. M. PRENTER, "A formal comparison of methods for the numerical solution of first kind integral equations," *J. Austral. Math. Soc. Ser. B*, v. 22, 1981, pp. 488–500.
2. M. BERTERO, C. DE MOL & G. A. VIANO, "The stability of inverse problems," in *Inverse Scattering Problems in Optics* (H. P. Baltes, ed.), Topics in Current Physics, vol. 20, Springer-Verlag, New York, 1980, pp. 161–214.
3. A. R. DAVIES & R. S. ANDERSSSEN, *Optimization in the Regularization of Ill-Posed Problems*, Centre for Mathematical Analysis Report 30-85, Australian National University, Canberra, 1985.
4. J. N. FRANKLIN, "On Tikhonov's method for ill-posed problems," *Math. Comp.*, v. 28, 1974, pp. 889–907.
5. C. W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, London, 1984.

6. M. A. LUKAS, "Regularization," in *The Application and Numerical Solution of Integral Equations* (R. S. Anderssen et al., eds.), Sijthoff and Noordhoff, Alphen aan den Rijn, the Netherlands, 1980, pp. 152–182.
7. M. A. LUKAS, *Convergence Rates for Regularized Solutions*, Report 5084, Colorado State University, 1984.
8. M. Z. NASHED & G. WAHBA, "Convergence rates of approximate least squares solutions of linear integral and operator equations of the first kind," *Math. Comp.*, v. 28, 1974, pp. 69–80.
9. D. W. NYCHKA & D. D. COX, *Convergence Rates for Regularized Solutions of Integral Equations from Discrete Noisy Data*, Technical Report No. 752, Department of Statistics, University of Wisconsin-Madison, 1984.
10. A. SPENCE, "Error bounds and estimates for eigenvalues of integral equations," *Numer. Math.*, v. 29, 1978, pp. 133–147.
11. A. N. TIKHONOV & V. Y. ARSENIN, *Solutions of Ill-posed Problems* (translated from the Russian), Wiley, New York, 1977.
12. C. R. VOGEL, "Optimal choice of truncation level for the truncated SVD solution of linear first kind integral equations when data are noisy," *SIAM J. Numer. Anal.*, v. 23, 1986, pp. 109–117.
13. G. WAHBA, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, v. 14, 1977, pp. 651–667.