# ASSESSING REGULARISED SOLUTIONS

## M. A. LUKAS[1]

### Abstract

Consider the prototype ill-posed problem of a first kind integral equation $\mathscr{K}u(x) \equiv \int K(x,t)u(t)\,dt = f(x)$ with discrete noisy data $d_i = f(x_i) + \varepsilon_i$, $i = 1,\ldots,n$. Let $u_0$ be the true solution and $u_{n\alpha}$ a regularised solution with regularisation parameter $\alpha$. Under certain assumptions, it is known that if $\alpha \to 0$ but not too quickly as $n \to \infty$, then $u_{n\alpha}$ converges to $u_0$. We examine the dependence of the optimal sequence of $\alpha$ and resulting optimal convergence rate on the smoothness of $f$ or $u_0$, the kernel $K$, the order of regularisation $m$ and the error norm used. Some important implications are made, including the fact that $m$ must be sufficiently high relative to the smoothness of $u_0$ in order to ensure optimal convergence. An optimal filtering criterion is used to determine the order $m = \bar{\theta} + 1/2$ where $\bar{\theta}$ is the maximum smoothness of $u_0$. Two practical methods for estimating the optimal $\alpha$, the unbiased risk estimate and generalised cross validation, are also discussed.

## 1. Introduction

With the growing recognition of the significance of inverse problems in applications, it is important to examine the techniques available for their approximate solution. In this paper we will consider the most prominent of these—the method of regularisation. Although this method can be applied to general inverse problems we focus on the important special case of integral equations of the first kind, which have the form (after scaling onto [0,1])

$$\mathscr{K}u(x) \equiv \int_0^1 K(x,t)u(t)\,dt = f(x) \qquad x \in [0,1]. \tag{1.1}$$

Assume that (1.1) has a unique solution $u_0$.

---

[1]School of Mathematical and Physical Sciences, Murdoch University, Murdoch, W.A. 6150, Australia.

 It is well known that these equations are ill-posed, in the sense that small perturbations in $f$ can lead to large errors in the solution $u_0$. Moreover these equations serve as a good model for general ill-posed inverse problems. Sometimes it is possible (and indeed useful) to determine the degree of ill-posedness of an inverse problem by comparing it with a first kind integral equation.

Under mild conditions on the kernel function $K(x,t)$ in (1.1), the operator $\mathcal{K}: L^2(0,1) \to L^2(0,1)$ has singular values $\omega_i$ (i.e. $\omega_i^2$ are eigenvalues of $\mathcal{K}^*\mathcal{K}$ which form a sequence decreasing to 0. The degree of ill-posedness of (1.1) depends upon the rate of decay of $\omega_i$; the faster the decay the more ill-posed the problem. There are results which relate the smoothness of the kernel $K(x,t)$ to the decay rate of $\omega_i$—basically the smoother the kernel, the faster the decay rate and hence the more ill-posed the problem. See [7] for a discussion of these points.

Below are some examples of first-kind integral equations from applications, which are listed in order of increasing degree of ill-posedness.

*Abel's equation.*

$$Au(x) \equiv \int_x^1 (t^2 - x^2)^{-1/2} u(t)\, dt = f(x).$$

Applications of this equation occur regularly in such areas as metallurgy, seismology and biology. The equation is said to be weakly ill-posed and corresponds to a half differentiation (see [1]).

*Differentiation.*

$$\text{If } \int_0^x u(t)\, dt = f(x) \text{ then } u = f'.$$

*Higher-order differentiation.*

If

$$\int_0^1 \frac{(x-t)_+^{k-1}}{(k-1)!} u(t)\, dt = f(x),$$

where $s_+ = s$ for $s \geq 0$ and $s_+ = 0$ for $s < 0$, then $u = f^{(k)}$. Note that the kernel is $k-2$ times continuously differentiable and it is known that the singular values behave as $\omega_i \approx ci^{-k}$.

*Fujita's equation.*

$$\int_0^1 \frac{\beta t e^{-\beta x t}}{1 - e^{-\beta t}} u(t)\, dt = f(x).$$

This equation arises in the sedimentation analysis of polymers (see [8]). The kernel is analytic and therefore its singular values decay exponentially like $\omega_i \approx ca^i$, $0 < a < 1$.

We shall suppose that the function $f(x)$ in (1.1) is given only as discrete noisy data

$$d_i = f(x_i) + \varepsilon_i \qquad i = 1, \ldots, n. \tag{1.2}$$

The errors $\varepsilon_i$ could be (say) experimental or roundoff errors. We shall assume that $\varepsilon_i$, $i = 1, \ldots, n$, are uncorrelated random variables with mean zero and common variance $\sigma^2$.

Given data (1.2), the method of regularisation defines an approximate solution to (1.1) as the solution to the following problem:

$$\underset{u \in W}{\text{minimise}} \; \frac{1}{n} \sum_{i=1}^{n} (\mathscr{K} u(x_i) - d_i)^2 + \alpha \|u\|_W^2. \tag{1.3}$$

Here $W$ is a Hilbert space of smooth functions, which is often taken to be the Sobolev space of order $m$ defined by

$$W^{m,2}[0, 1] = \{u \in L^2(0,1) : u^{(m)} \text{ exists weakly and } u^{(m)} \in L^2(0,1)\}.$$

Then $m$ is called the order of regularisation. In place of the norm $\|u\|_W$ in (1.3), it is also common to use the seminorm $\|u^{(m)}\|$, where $\|\cdot\|$ denotes the $L^2(0,1)$ norm. However, we will consider the formulation (1.3), because its solution is somewhat simpler to deal with. For a general introduction to regularisation see [9], [18].

The constant $\alpha > 0$ in (1.3) is called the regularisation parameter. Clearly, it controls the tradeoff between respecting the data by minimising the residual sum of squares and smoothing the solution by minimising its norm.

An appropriate framework for the solution of the regularisation problem (1.3) is to assume that $W$ is a reproducing kernel Hilbert space (RKHS). That is, $W$ is a Hilbert space of functions such that for each $x \in [0, 1]$ the evaluation functional $W \to R$, $u \to u(x)$ is bounded. Equivalently, there exists a function $R(x, t)$, called the reproducing kernel (RK), such that for all $u \in W$, $x \in [0, 1]$,

$$(R(x, \cdot), u)_W = u(x).$$

It is not hard to show that $R(x, t)$ is unique and symmetric, that is

$$R(x, t) = R(t, x).$$

For example, consider the Sobolev space $W^{m,2}[0, 1]$ with inner product

$$(u, v)_W = \sum_{i=1}^{m} B_i u B_i v + \int_0^1 Tu \, Tv,$$

where $T$ is a linear differential operator of order $m$, e.g. $Tu = u^{(m)}$, and $B_i u = 0$, $i = 1, \ldots, m$, is a set of linear homogeneous boundary conditions, e.g. $B_i u = u^{(i-1)}(0) = 0$. Then $W^{m,2}[0, 1]$ is a RKHS with RK

$$R(x, t) = \sum_{j=1}^{m} \theta_j(x)\theta_j(t) + GG^*(x, t),$$

where $\theta_j$, $j = 1, \ldots, m$, satisfy $T\theta_j = 0$ and $B_i\theta_j = \delta_{ij}$, and $G$ is the Green's function for the boundary-value problem

$$Tu = 0, \qquad B_iu = 0 \quad i = 1, \ldots, m.$$

For instance, if $Tu = u^{(m)}$ and $B_iu = u^{(i-1)}(0)$, then $\theta_j(x) = x^{j-1}/(j-1)!$ and $G(x, y) = (x - y)_+^{m-1}/(m - 1)!$. An important point to note is that as $m$ increases, the RK $R(x, t)$ becomes more smooth.

Under the weak assumption that the functionals $W \to \mathscr{R}$, $u \to \mathscr{K}u(x)$, $x \in [0, 1]$ are bounded, it is known (see [19]) that the regularisation problem (1.3) has the unique solution (called the regularised solution)

$$u_{n\alpha}(t) = \boldsymbol{\eta}(t)(Q_n + \alpha nI)^{-1}\mathbf{d}, \tag{1.4}$$

where

$$\eta_i(t) = RK^*(x_i, t) = \int_0^1 R(x_i, s)K(t, s)\, ds,$$

$$[Q_n]_{ij} = Q(x_i, x_j)$$

and

$$Q(x, y) = KRK^*(x, y) = \int_0^1 \int_0^1 K(x, s)R(s, t)K(y, t)\, ds\, dt. \tag{1.5}$$

Furthermore, because $\mathscr{K}\eta_i(x) = Q(x_i, x)$,

$$\mathscr{K}u_{n\alpha}(x) = \mathbf{Q}(x)(Q_n + \alpha nI)^{-1}\mathbf{d}, \tag{1.6}$$

where $Q_i(x) = Q(x_i, x)$.

It is useful to note that a special case of regularisation is data smoothing. For if $\mathscr{K}u = u$ in (1.3), then the solution $u_{n\alpha}$ is a function smoothing the data $d_i$. If also $\|u\|_W^2$ is replaced with $\|u^{(m)}\|^2$, then $u_{n\alpha}$ is the familiar natural polynomial smoothing spline of degree $2m - 1$.

From the regularisation problem (1.3), it is intuitively clear that there is an optimal $\alpha$. If $\alpha$ is too large then the smoothing term $\alpha\|u\|_W^2$ dominates the residual term, resulting in an overly smooth regularised solution which takes little account of the data. On the other hand, if $\alpha$ is too small then the residual term dominates the smoothing term, and because of the ill-posedness of the problem this results in a very noisy regularised solution.

It is also possible to foresee the fact that $\alpha$ must tend to zero as $n \to \infty$ in order that the regularised solution $u_{n\alpha}$ converge to the true solution $u_0$ of $\mathscr{K}u = f$. This is because as $n$ increases, the residual term in (1.3) reflects an increasing knowledge of the function $f$, and so should be given greater weighting.

The aim of this paper is to assess regularised solutions from a theoretical point of view. In so doing we hope to shed some light on both their potential and limitations as approximate solutions to inverse problems.

In Section 2, we investigate the question of convergence of $u_{n\alpha}$ to the true solution $u_0$ of $\mathscr{K}u = f$. We state some general results (proved in [11]) which show that $u_{n\alpha}$ converges to $u_0$ as $n \to \infty$ in a wide class of norms, provided that $\alpha = \alpha(n) \to 0$ in a certain way. These results are applied to convolution equations with periodic kernels, and we determine how the rate of convergence depends upon the degree of ill-posedness of the problem, the smoothness of $f$ or $u_0$ and the order of regularisation.

In the past, the question of the choice of the order of regularisation has been largely ignored. In Section 3, following [12], we show how an optimal filtering criterion leads, in the convolution case, to the choice of order $m = \bar{\theta} + 1/2$, where $\bar{\theta}$ is the maximum smoothness of $u_0$.

For regularisation to be useful in practice, there must be some method for choosing a good value of $\alpha$, for any given data set. We shall consider two such methods—the unbiased risk estimate and generalised cross validation (GCV). Results are presented (proved in [13]) which support existing numerical evidence that GVC is a very reliable method.

## 2. Convergence of regularised solutions

First we must define the sense in which convergence is meant.

Assume that the function $Q(x, y)$ defined in (1.5) is continuous on $[0, 1] \times [0, 1]$. Then the integral operator $\mathscr{Q}$ with kernel $Q(x, y)$ defined by

$$\mathscr{Q}f(x) = \int_0^1 Q(x, y) f(y) \, dy$$

has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ satisfying $\lambda_i \downarrow 0$ with corresponding orthonormal eigenfunctions $\phi_i \in C[0, 1]$. We shall assume that the $\lambda_i$ decay such that

$$0 < a_1 i^{-2p} \leq \lambda_i \leq a_2 i^{-2p} \qquad p > \tfrac{1}{2}. \tag{2.1}$$

Following [10], define a class of Hilbert spaces $H_\mu$, $0 \leq \mu \leq 1$, by

$$H_\mu = \left\{ f \in L^2(0, 1) \colon \sum_{i=1}^{\infty} (f, \phi_i)^2 / \lambda_i^\mu < \infty \right\},$$

where the inner product $(\cdot, \cdot)$ is the usual $L^2(0, 1)$ inner product. The inner product in $H_\mu$ is

$$(f, g)_{H_\mu} = \sum_{i=1}^{\infty} (f, \phi_i)(g, \phi_i) / \lambda_i^\mu.$$

Equivalently, $H_\mu$ can be defined as

$$H_\mu = \mathscr{Q}^{\mu/2}(L^2(0, 1))$$

with inner product

$$(f,g)_{H_\mu} = (\mathscr{Q}^{-\mu/2}f, \mathscr{Q}^{-\mu/2}g)_{L^2(0,1)},$$

where $\mathscr{Q}^{\mu/2}$ is the $\mu/2$ root of the positive operator $\mathscr{Q}$. Also define another class of Hilbert spaces $W_\mu$, $0 \le \mu \le 1$, by

$$W_\mu = \{u \colon \mathscr{K}u \in H_\mu\}$$

with inner product

$$(u,v)_{W_\mu} = (\mathscr{K}u, \mathscr{K}v)_{H_\mu}.$$

It is known (see [14]) that $H_1$ has the alternative description $H_1 = H \equiv \mathscr{K}(W)$ with inner product

$$(f,g)_{H_1} = (\mathscr{K}^{-1}f, \mathscr{K}^{-1}g)_W.$$

Therefore, from the definition, $W_1 = W$. Also, by definition, $H_0 = L^2(0,1)$. It is clear that if $\mu > \nu$, then $H_\mu \subset H_\nu$ and $\|\cdot\|_{H_\mu}$ is a stronger norm than $\|\cdot\|_{H_\nu}$. Thus we have a class of spaces $H_\mu$, $0 \le \mu \le 1$, ranging from $H_0 = L^2(0,1)$ to $H_1 = H \equiv \mathscr{K}(W)$.

From the definition of $W_\mu$ and since $\mathscr{K}u_0 = f$, we have

$$\|u_{n\alpha} - u_0\|_{W_\mu} = \|\mathscr{K}u_{n\alpha} - f\|_{H_\mu}.$$

Hence we have an immediate connection between convergence in the domain space $W_\mu$ and in the range space $H_\mu = \mathscr{K}(W_\mu)$. When it is obvious from the function, we will simply write $\|\cdot\|_\mu$ for both $\|\cdot\|_{H_\mu}$ and $\|\cdot\|_{W_\mu}$.

Because we assume that the errors $\varepsilon_i$ are random variables, then from (1.4), $\mu_{n\alpha}$ is a random function. Therefore, convergence of $u_{n\alpha}$ will be taken to mean with respect to the expected squared error $E\|u_{n\alpha} - u_0\|_\mu^2$ and it is this that must be estimated.

To describe the convergence results, we shall use the following notation. Given two positive sequences $a_n, b_n$, denote $a_n \precsim b_n$ if there is a constant $c$ such that $a_n \le c b_n$ for all $n$. Write $a_n \approx b_n$ if there are constants $c_1, c_2 > 0$ such that $c_1 b_n \le a_n \le c_2 b_n$ for all $n$. We shall also use the asymptotic notation $a_n \sim b_n$, which means $a_n/b_n \to 1$.

We shall need the following quadrature assumption. Assume there exists $\nu$, $0 < \nu < 1 - (1/4p)$, and sequence $k_n \to 0$ such that for all $f, g \in H$

$$\left| \int_0^1 fg - \frac{1}{n} \sum_{i=1}^n f(x_i)g(x_i) \right| \le k_n \|f\|_\nu \|g\|_\nu. \tag{2.2}$$

The following theorem is proved in [11]. See also [19], [17] and [2].

THEOREM 2.1. *With assumption* (2.1) *and* (2.2), *let* $f \in H_s$, *where* $s \ge \max\{\nu, \mu\}$ *and* $\mu < 2 - \nu - (1/2p)$. *Suppose that* $\alpha \to 0$ *as* $n \to \infty$ *in such a way that*

$$k_n \alpha^{-\nu - (1/4p)} \to 0$$

*and if $\mu > \nu$, $s > \nu + 2$*

$$k_n \alpha^{-\nu/2 - \mu/2 - (1/4p)} \to 0.$$

*Then for $\mu \le s \le \mu + 2$,*

$$\alpha^2 \|f\|_\mu^2 + \frac{\sigma^2}{n} \alpha^{-\mu - (1/2p)} \precsim E\|u_{n\alpha} - u_0\|_\mu^2 \precsim \alpha^{s-\mu} \|f\|_s^2 + \frac{\sigma^2}{n} \alpha^{-\mu - (1/2p)}$$

*and for $s \ge \mu + 2$,*

$$E\|u_{n\alpha} - u_0\|_\mu^2 \approx \alpha^2 \|f\|_{\mu+2}^2 + \frac{\sigma^2}{n} \alpha^{-\mu - (1/2p)}.$$

The proof of Theorem 2.1 begins by decomposing the expected mean square error into the bias squared plus the variance as follows:

$$\begin{aligned}
E\|u_{n\alpha} - u_0\|_\mu^2 &= E\|Eu_{n\alpha} - u_0 + u_{n\alpha} - Eu_{n\alpha}\|_\mu^2 \\
&= \|Eu_{n\alpha} - u_0\|_\mu^2 + E\|u_{n\alpha} - Eu_{n\alpha}\|_\mu^2 \\
&\quad + 2E(Eu_{n\alpha} - u_0, u_{n\alpha} - Eu_{n\alpha})_\mu \\
&= \|Eu_{n\alpha} - u_0\|_\mu^2 + E\|u_{n\alpha} - Eu_{n\alpha}\|_\mu^2
\end{aligned}$$

since from (1.4), for any $v \in W_\mu$,

$$\begin{aligned}
E(v, u_{n\alpha} - Eu_{n\alpha})_\mu &= \sum_{i,j=1}^n (v, \eta_i)_\mu (Q_n + \alpha nI)_{ij}^{-1} E(d_j - f(x_j)) \\
&= 0.
\end{aligned}$$

The bias squared and variance are then estimated separately, giving rise to the two terms in the bounds of Theorem 2.1.

If it is assumed that the Fourier coefficients of $f$ decay in a certain regular way, then the bounds in Theorem 2.1 can be replaced by precise estimates.

THEOREM 2.2. *Suppose that assumptions (2.1) and (2.2) hold and $f$ satisfies $(f, \phi_i)^2 \approx \lambda_i^r$, which implies that $f \in H_s$ for all $s < \bar{s} = r - 1/2p$. If for some $\varepsilon > 0$, $\alpha \to 0$ as $n \to \infty$ in such a way that*

$$k_n \alpha^{-\nu - 1/4p - \varepsilon} \to 0 \quad \text{when } \mu \le \nu \text{ or when } \mu > \nu, \ \bar{s} \le \nu + 2$$

*and*

$$k_n \alpha^{-\nu/2 - \mu/2 - (1/4p) - \varepsilon/2} \to 0 \quad \text{when } \mu > \nu, \ \bar{s} > \nu + 2,$$

*then*

$$E\|u_{n\alpha} - u_0\|_\mu^2 \approx \begin{cases} \alpha^{\bar{s}-\mu} + \dfrac{\sigma^2}{n} \alpha^{-\mu - (1/2p)} & \mu < \bar{s} < \mu + 2 \\[2ex] \alpha^2 ln\dfrac{1}{\alpha} + \dfrac{\sigma^2}{n} \alpha^{-\mu - (1/2p)} & \bar{s} = \mu + 2 \\[2ex] \alpha^2 + \dfrac{\sigma^2}{n} \alpha^{-\mu - (1/2p)} & \bar{s} > \mu + 2. \end{cases}$$

This result is proved in [11].

Suppose that $\bar{\alpha} = \bar{\alpha}(n)$ minimises $E\|u_{n\alpha} - u_0\|_\mu^2$. That is, $\bar{\alpha}$ is the optimal value of $\alpha$ with respect to the $\mu$ norm. By minimising the estimates in Theorem 2.2 we have the following.

COROLLARY 2.1. *Assume that $\bar{\alpha} \to 0$ as in Theorem 2.2. Then*

$$\bar{\alpha} \approx \begin{cases} \left(\dfrac{\sigma^2}{n}\right)^{2p/(2p\bar{s}+1)} & \mu < \bar{s} < \mu + 2 \\[3mm] \left(\dfrac{\sigma^2}{n}\right)^{2p/(4p+2p\mu+1)} & \bar{s} > \mu + 2 \end{cases} \tag{2.3}$$

*and*

$$e_{n\bar{\alpha}} \equiv E\|u_{n\bar{\alpha}} - u_0\|_\mu^2 \approx \begin{cases} \left(\dfrac{\sigma^2}{n}\right)^{2p(\bar{s}-\mu)/(2p\bar{s}+1)} & \mu < \bar{s} < \mu + 2 \\[3mm] \left(\dfrac{\sigma^2}{n}\right)^{4p/(4p+2p\mu+1)} & \bar{s} > \mu + 2. \end{cases} \tag{2.4}$$

We now want to examine how these optimal estimates depend on the parameters involved. First note that $\mu$ and $\bar{s}$ (or $r$) are not absolute parameters. Since $\lambda_i \approx i^{-2p}$, the norm $\|f\|_\mu$, $f \in H_\mu$, is equivalent to that defined as

$$\left(\sum_{i=1}^{\infty} (f, \phi_i)^2 i^{2p\mu}\right)^{1/2}.$$

Therefore the norm is described in absolute terms by $2p\mu$. Similarly the function class $H_s$ containing $f$ is described in absolute terms by $2ps$.

We shall discuss the optimal convergence rates (2.3) and (2.4) with particular reference to the special case of a convolution equation

$$\int_0^1 K(x - t)u(t)\,dt = f(x)$$

in which the function $K$ is periodic of period 1. In this case, expand $K$ in a Fourier series

$$K(x - t) = \sum_{j=-\infty}^{\infty} \hat{K}_j e^{2\pi i j(x-t)},$$

where $\hat{K}_j$ is the Fourier coefficient

$$\hat{K}_j = \int_0^1 K(y)e^{2\pi i j y}\,dy = \overline{\hat{K}}_{-j}.$$

We shall assume that $|\hat{K}_j| \approx j^{-k}$, $k > 1$, and so the Fourier series is uniformly convergent.

For this convolution case we will take $W$ to be the space of periodic functions

$$W(m) \equiv \{u \in W^{m,2}[0,1]: u^{(i)}(0) = u^{(i)}(1), i = 0, \ldots, m-1\}$$

with inner product

$$(u,v)_{W(m)} = \int_0^1 uv + \int_0^1 u^{(m)}v^{(m)},$$

or equivalently

$$W(m) = \{u \in L^2(0,1): \sum_{j=-\infty}^{\infty} (2\pi j)^{2m}|\hat{u}_j|^2 < \infty\}$$

with

$$(u,v)_{W(m)} = (u,v) + \sum_{j=-\infty}^{\infty} (2\pi j)^{2m}\hat{u}_j\bar{\hat{v}}_j.$$

The reproducing kernel for $W(m)$ is known to be (see [3])

$$R(x,y) = R(x-y) = 1 + \sum_{\substack{j=-\infty \\ j\neq 0}}^{\infty} (2\pi j)^{-2m}e^{2\pi ij(x-y)}$$

$$= 1 + \frac{(-1)^{m-1}}{(2m)!}\overline{B}_{2m}(x-y),$$

where $\overline{B}_{2m}$ is the periodic extension of the Bernoulli polynomial $B_{2m}$ of degree $2m$ on $[0,1]$. These polynomials are defined inductively by

$$B_0 = 1$$

$$B'_{n+1} = (n+1)B_n, \qquad \int_0^1 B_{n+1}(t)\,dt = 0 \quad \text{for } n \geq 0,$$

so that the next two are

$$B_1(t) = t - \tfrac{1}{2}$$
$$B_2(t) = t^2 - t + \tfrac{1}{6}.$$

Some straightforward calculation shows that the kernel $Q(x,y)$ in (1.5) is

$$Q(x,y) = Q(x-y) = |\hat{K}_0|^2 + \sum_{j\neq 0} |\hat{K}_j|^2 (2\pi j)^{-2m}e^{2\pi ij(x-y)}.$$

From this expansion of $Q(x,y)$ it is clear that the eigenvalues of $\mathcal{Q}$ are

$$\lambda_j = \begin{cases} |\hat{K}_0|^2 & j = 0 \\ |\hat{K}_j|^2 (2\pi j)^{-2m} & j \neq 0 \end{cases}$$

with corresponding orthonormal eigenfunctions

$$\phi_j(x) = \exp(2\pi ijx).$$

Since $|\hat{K}_j| \approx j^{-k}$, then $\lambda_j \approx j^{-2k-2m}$, and so the parameter $p = k + m$. The space $H_\mu$ can be written as

$$H_\mu = \left\{ f \in L^2(0,1) \colon \sum_{j \neq 0} |\hat{f}_j|^2 j^{2p\mu} < \infty \right\}$$

and $W_\mu$ is the completion of

$$\left\{ u \in L^2(0,1) \colon \sum_{j \neq 0} |\hat{u}_j|^2 j^{2p\mu-2k} < \infty \right\}$$

under the inner product

$$(u,v)_{W_\mu} = (\mathscr{K}u, \mathscr{K}v)_{H_\mu}.$$

But this space $H_\mu$ is simply the periodic Sobolev space (or Besov space) of fractional order of smoothness $p\mu$ which we will denote by $W(p\mu)$. So $H_\mu = W(p\mu)$ and similarly $W_\mu = W(p\mu{-}k)$, which, if $k > p\mu$, is of negative order and therefore a space of distributions. Note that as required $W_1 = W(m) = W$ and $H_0 = W(0) = L^2(0,1)$.

Let $\bar{\theta} = p\bar{s} - k$. Then $f \in H_s$, $s < \bar{s}$, if and only if $f \in W(\theta + k)$, $\theta < \bar{\theta}$, which is equivalent to $u_0 \in W(\theta)$, $\theta < \bar{\theta}$. Since the $W_\mu$ norm and the $W(\tau)$ norm for $\tau = p\mu - k$ are equivalent, the estimate (2.4) can be written as

$$E\|u_{n\bar{\alpha}} - u_0\|^2_{W(\tau)} \approx \begin{cases} \left(\dfrac{\sigma^2}{n}\right)^{(2\bar{\theta}-2\tau)/(2\bar{\theta}+2k+1)} & \tau < \bar{\theta} < \tau + 2m + 2k, \\[4mm] \left(\dfrac{\sigma^2}{n}\right)^{4(m+k)/(4m+6k+2\tau+1)} & \bar{\theta} > \tau + 2m + 2k. \end{cases}$$

$$(2.5)$$

We shall now examine how the above convergence rates for $\bar{\alpha}$ and $e_{n\bar{\alpha}}$ depend on each of the following aspects.

**Function $f$ or solution $u_0$.** Consider the behaviour of the convergence rates (2.3) and (2.4) as we increase $2p\bar{s} = 2pr - 1$, thereby restricting $f$ and $u_0$. Clearly $\bar{\alpha}$ decreases more slowly as $2p\bar{s}$ increases until $2p\bar{s} = 2p\mu + 4p$, from which point its convergence rate is independent of $2p\bar{s}$. On the other hand, $e_{n\bar{\alpha}}$ decreases more quickly as $2p\bar{s}$ increases until $2p\bar{s} = 2p\mu + 4p$. This is not surprising since the function $f$ (or $u_0$) is better behaved as $2p\bar{s}$ increases, but at the point $2p\bar{s} = 2p\mu + 4p$ the approximation process becomes saturated.

For the convolution example above, increasing $2p\bar{s} = 2\bar{\theta} + 2k$ is equivalent to increasing the order of smoothness of $f \in W(\theta + k)$ and $u_0 \in W(\theta)$, $\theta < \bar{\theta}$. As $\bar{\theta}$ increases, the convergence rates (2.3) and (2.5) exhibit the same behaviour as above with the saturation point at $\bar{\theta} = \tau + 2m + 2k$. It is useful to consider the special case of data smoothing by periodic splines of degree $2m - 1$. This

is the case with $\mathscr{K}$ equal to the identity, so that $k = 0$. If also $\tau = 0$ (so $W(\tau) = L^2(0,1)$), then the estimate (2.5) becomes

$$E\|u_{n\bar{\alpha}} - u_0\|^2_{L^2(0,1)} \approx \begin{cases} \left(\dfrac{\sigma^2}{n}\right)^{2\bar{\theta}/(2\bar{\theta}+1)} & 0 < \bar{\theta} < 2m, \\[2mm] \left(\dfrac{\sigma^2}{n}\right)^{4m/(4m+1)} & \bar{\theta} > 2m. \end{cases}$$

This is consistent with the optimal convergence rates using periodic splines found in [3].

**Kernel $K$.** Next, consider the dependence of the optimal convergence rates on the kernel $K$ and therefore on the degree of ill-posedness of the problem. If the kernel $K$ is assumed to be smoother, then the same will be true for the kernel $Q$. Therefore we can expect the eigenvalues of $Q$ to decay more quickly, which means $p$ is increased. As $p$ increases (but assuming $2p\bar{s}$ fixed) the exponent of $\sigma^2/n$ for $\bar{\alpha}$ in (2.3) which equals

$$\begin{cases} \dfrac{2p}{4p + 2p\mu + 1} & p < \dfrac{2p\bar{s} - 2p\mu}{4} \\[2mm] \dfrac{2p}{2p\bar{s} + 1} & p > \dfrac{2p\bar{s} - 2p\mu}{4} \end{cases}$$

increases. This means the optimal regularisation parameter converges to 0 more quickly. The exponent of $\sigma^2/n$ for $e_{n\bar{\alpha}}$ in (2.4), which equals

$$\begin{cases} \dfrac{4p}{4p + 2p\mu + 1} & p < \dfrac{2p\bar{s} - 2p\mu}{4} \\[2mm] \dfrac{2p\bar{s} - 2p\mu}{2p\bar{s} + 1} & p > \dfrac{2p\bar{s} - 2p\mu}{4}, \end{cases}$$

increases for $p < (2p\bar{s} - 2p\mu)/4$ and is constant for $p > (2p\bar{s} - 2p\mu)/4$.

This observation gives little hint of the practical difficulty of solving severely ill-posed problems by regularisation. The main reason for this difficulty is that, because of the very fast decay of $\lambda_i$, only a very few $\lambda_i$ and their corresponding $\phi_i$ can be computed accurately. Hence the computed regularised solution is constrained to lie in a subspace of $W$ with low dimension, much less than $n$. This is why it is difficult to recover even a bimodal solution of Fujita's equation. See [20], [10] and [16] for discussion of these points.

**Order of regularisation $m$.** As $m$ increases, the reproducing kernel $R$, and therefore the kernel $Q$, become smoother. Hence $p$ increases, giving the same effect as for the kernel $K$. Therefore we can optimise the rate of convergence by ensuring that
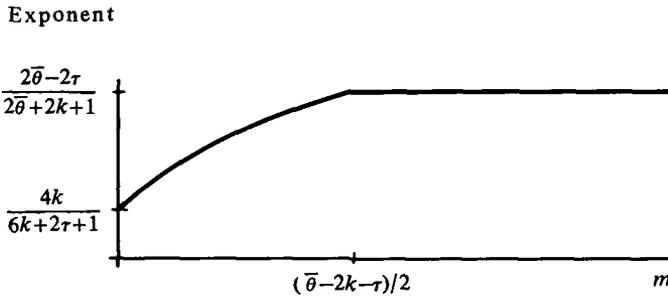
$$p \geq (2p\bar{s} - 2p\mu)/4.$$

Exponent



FIGURE 1

This is equivalent to $\bar{s} \le 2 + \mu$ which is always the case if $\bar{s} \le 2$. This can be interpreted as meaning that $f$ should not be too smooth relative to $Q$; in fact no smoother than that associated with $f \in \mathscr{C}(L^2(0,1))$.

In the convolution case, the graph of the exponent of $\sigma^2/n$ in (2.5) against $m$ is displayed in Figure 1. Therefore the rate of convergence is optimised for $m > (\bar{\theta} - 2k - \tau)/2$, which is true for all norms $W(\tau)$, $\tau \ge 0$, if $m > (\bar{\theta} - 2k)/2$.

**Norm** $\|\cdot\|_\mu$. As $2p\mu$ increases in (2.4), $e_{n\bar{\alpha}}$ decreases more slowly. This is not surprising because the norm $\|\cdot\|_\mu$ will be stronger. The same occurs in (2.5) as $\tau$ increases.

In (2.3) however, it is clear that as $2p\mu$ increases, $\bar{\alpha}$ decreases more slowly until $2p\mu = 2p\bar{s} - 4p$, from which point the decay rate is independent of $2p\mu$. Therefore if $2p\bar{s} \le 4p$ or $\bar{s} \le 2$, then the optimal rate of convergence for the regularisation parameter does not depend on the norm used. This is important in practice because if some method gives a good estimate of $\bar{\alpha}_0$ which minimises say $E\|\mathscr{K}u_{n\alpha} - f\|^2$ (that is $\mu = 0$) then that estimate should also be close to $\bar{\alpha}_1$ which minimises $E\|u_{n\alpha} - u_0\|_W^2$ (i.e. $\mu = 1$).

Note that this condition $\bar{s} \le 2$ is the same as that above which ensured the optimal rate of convergence for $e_{n\bar{\alpha}}$. In the convolution case, it is equivalent to $m \ge (\bar{\theta} - k)/2$.

Lastly in this section we shall verify the assumptions made in Corollary 2.1, in the convolution case. The following result is a special case of Theorem 2.5 in [11].

THEOREM 2.3. *For the convolution case, assumption (2.2) holds with $\nu = \gamma/(m+k)$ and $k_n \le n^{-\gamma}$, for any $\gamma$ such that $1/2 < \gamma < m + k - 1/4$.*

THEOREM 2.4. *For the convolution case, $\bar{\alpha} \to 0$ as in Theorem 2.2 if $m + k > 3/4$ and*

$$\begin{cases} (m + k = 1/4)(2\bar{\theta} + 2k - 1) > 1/2 & \tau < \bar{\theta} < \tau + 2m + 2k, \\ (m + k - 1/4)(4m + 6k + 2\tau - 1) > 1/2 & \bar{\theta} > \tau + 2m + 2k. \end{cases}$$

PROOF. This is simply a matter of checking the conditions using Theorem 2.3.

## 3. Order of regularisation

In Section 2 it was shown that in the convolution case, the rate of convergence of $u_{n\alpha}$ to $u_0$ with respect to all $W(\tau)$ norms, $\tau \geq 0$, is optimised if $m \geq (\bar{\theta} - 2k)/2$. This means that one should always choose an order of regularisation which is sufficiently high. A similar finding was made in a different context by Natterer [15].

However, within the range $m \geq (\bar{\theta} - 2k)/2$ there should be an optimal order in some sense. We describe here the results in [12], in which an optimal filtering criterion is used to derive an optimal order of regularisation.

Let $\bar{\lambda}_i \geq$, $i = 1, \ldots, n$, be the decreasing eigenvalues of the positive semi-definite matrix $(1/n)Q_n$ and let $\bar{\phi}_i$ be the corresponding eigenvectors normalised such that

$$\langle \bar{\phi}_i, \bar{\phi}_j \rangle \equiv \frac{1}{n} \sum_{k=1}^{n} \bar{\phi}_{ik} \bar{\phi}_{jk} = \delta_{ij}.$$

Clearly $\bar{\lambda}_i$ and $\bar{\phi}_i$ are approximations to $\lambda_i$ and $\phi_i$ respectively. From (1.6) it is not hard to show that

$$\mathcal{K} u_{n\alpha}(x_j) = \sum_{i=1}^{n} \bar{\lambda}_i / (\bar{\lambda}_i + \alpha) \langle \mathbf{d}, \bar{\phi}_i \rangle \bar{\phi}_{ij}. \tag{3.1}$$

Note that $\bar{\lambda}_i / (\bar{\lambda}_i + \alpha)$ is a decreasing sequence and $\langle \mathbf{d}, \bar{\phi}_i \rangle = \langle \mathbf{f}, \bar{\phi}_i \rangle + \langle \varepsilon, \bar{\phi}_i \rangle$ where $\mathbf{f}_i = f(x_i)$. Because most of the noise $\varepsilon$ will be contained in the "high frequency" components it is clear from (3.1) that regularisation achieves a filtering of the noise from the data. The sequence $\bar{\lambda}_i / (\bar{\lambda}_i + \alpha)$, $i = 1, \ldots, n$, is called the regularisation filter.

Define the optimal filter $l_i^*, i = 1, \ldots, n$, to be the sequence $l_i$ which minimises

$$E \frac{1}{n} \left| \sum_{i=1}^{n} l_i \langle \mathbf{d}, \bar{\phi}_i \rangle \bar{\phi}_i - \mathbf{f} \right|^2,$$

where $| \cdot |$ is the usual Euclidean vector norm. It is not difficult to show that

$$l_i^* = \frac{\langle \mathbf{f}, \bar{\phi}_i \rangle^2}{\langle \mathbf{f}, \bar{\phi}_i \rangle^2 + \sigma^2/n} \qquad i = 1, \ldots, n. \tag{3.2}$$

Of course the filter cannot be used in practice to construct an approximate solution because $f$ is not known.

However suppose that the regularisation filter were equal to the optimal filter for all $n$, i.e.

$$\frac{\bar{\lambda}_i}{\bar{\lambda}_i + \alpha} = \frac{\langle \mathbf{f}, \bar{\phi}_i \rangle^2}{\langle \mathbf{f}, \bar{\phi}_i \rangle^2 + \sigma^2/n}.$$

Under certain conditions, this implies (see [12]) that

$$(f, \phi_i)^2 = C\lambda_i,$$

where

$$C = \|f\|^2 \Big/ \int_0^1 Q(x, x)\, dx.$$

That is $(f, \phi_i)^2 \approx \lambda_i^r$ with $r = 1$, which gives the optimal filtering condition

$$2p = 2p\bar{s} + 1. \tag{3.3}$$

In the convolution case, since $p = m + k$ and $p\bar{s} = \bar{\theta} + k$, the condition $2p = 2p\bar{s} + 1$ is equivalent to

$$m = \bar{\theta} + \tfrac{1}{2}. \tag{3.4}$$

Recall that $\bar{\theta}$ is the maximum smoothness assumed for $u_0$, i.e. $u_0 \in W(\theta)$ for all $\theta < \bar{\theta}$. Therefore if one is prepared to assume some degree of smoothness for $u_0$, then the choice (3.4) of order of regularisation is very simple. For example if we assume that $u_0 \in W(2)$ but $u_0 \notin W(3)$, then of integer orders $m$, either $m = 2$ or $m = 3$ would be optimal.

Note that $m = \bar{\theta} + 1/2$ is within the range $m \geq (\bar{\theta} - 2k)/2$ yielding the optimal rate of convergence for $u_{n\alpha}$ to $u_0$. When $m = \bar{\theta} + 1/2$, then the exponent in (2.3) is

$$\frac{2p}{2p\bar{s} + 1} = \frac{2m + 2k}{2\bar{\theta} + 2k + 1} = \frac{2\bar{\theta} + 1 + 2k}{2\bar{\theta} + 2k + 1} = 1,$$

so that $\bar{\alpha} \approx \sigma^2/n$. Numerical experiments reported in [12] support the claim that these choices of $m$ and $\alpha$ are optimal.

If nothing is known about the smoothness of $u_0$, it is still possible to estimate an order of regularisation by such methods as the unbiased risk estimate and generalised cross validation. These will be discussed in the next section.

Related results on the optimal choice of $m$ exist (see [4]) in the case of continuous regularisation (i.e. with $\|\mathcal{K}u - f\|^2$ in place of $(1/n)\sum(\mathcal{K}u(x_i) - d_i)^2$ in (1.3)) and deterministic error. It should also be noted that the optimal filter

(3.2) is analogous to the optimal Wiener filter applied to the deconvolution of a stationary stochastic process contaminated by noise (see Chapter V of [18]).

## 4. Choice of regularisation parameter

In practice, the choice of regularisation parameter is crucial to obtaining a good regularised solution. We shall consider two methods which have been found to be very effective in giving a good choice of the parameter. Both of these methods can be defined in terms of the matrix

$$A(\alpha) = Q_n(Q_n + n\alpha I)^{-1},$$

which is called the influence matrix because

$$\mathscr{K} u_\alpha(x_i) = A(\alpha)\,d_i \qquad i = 1, \ldots, n.$$

Consider the mean square error in the range space

$$R(\alpha) \equiv \frac{1}{n} \sum_{i=1}^{n} (\mathscr{K} u_{n\alpha}(x_i) - f(x_i))^2.$$

It is well known and easy to verify, using the assumed properties of $\varepsilon$, that the expectation of $R(\alpha)$ is

$$ER(\alpha) = (1/n)|(I - A(\alpha))\mathbf{f}|^2 + (\sigma^2/n)\text{Tr}\,A^2(\alpha),$$

where Tr denotes the trace of the matrix. The minimiser of $ER(\alpha)$ would of course be a good choice for $\alpha$, but it cannot be found because $f$ is not known. However there are methods which provide a good estimate of this minimiser. From the discussion in Section 2, and since asymptotically $R(\alpha) \sim \|\mathscr{K} u_{n\alpha} - f\|^2$, these methods also provide good estimates of the minimisers $\bar{\alpha}_0$ and $\bar{\alpha}_1$, as long as $\bar{s} \leq 2$.

**Unbiased risk estimate.** This estimate is recommended when the variance $\sigma^2$ of the errors is known. Define

$$\hat{R}(\alpha) = (1/n)|(I - A(\alpha))\,\mathbf{d}|^2 - (\sigma^2/n)\text{Tr}(I - A(\alpha))^2 + (\sigma^2/n)\text{Tr}\,A^2(\alpha).$$

It is easy to show that $\hat{R}(\alpha)$ is an unbiased estimate of $ER(\alpha)$, that is, $E\hat{R}(\alpha) = ER(\alpha)$. Therefore, choose $\alpha$ to be the minimiser of $\hat{R}(\alpha)$. This estimate was proposed in this context by Craven and Wahba [3] and Lukas [10] and has been found to be very reliable in numerical experiments.

Clearly $\hat{R} = \hat{R}(m, \alpha)$ is also a function of the order of regularisation $m$. By minimising $\hat{R}(m, \alpha)$ with respect to both $m$ and $\alpha$, a good estimate of the optimal order of regularisation is also obtained.

**Generalised cross validation (GCV).** This estimate is recommended when $\sigma^2$ is unknown.

Let $u_\alpha^{(k)}$ be the regularised solution found by ignoring the $k$th data value. That is $u_\alpha^{(k)}$ minimises

$$\frac{1}{n}\sum_{i\neq k}(\mathscr{K}u(x_i) - d_i)^2 + \alpha\|u\|_W^2.$$

If $\tilde{\alpha}$ is a good choice of $\alpha$, then for each $k$, $\mathscr{K}u_{\tilde{\alpha}}^{(k)}(x_k)$ should be closer to $d_k$ on average than $\mathscr{K}u_\alpha^{(k)}(x_k)$ for other values of $\alpha$. That is $\mathscr{K}u_{\tilde{\alpha}}^{(k)}(x_k)$ should be better able to predict $d_k$. Define the total prediction error, called the GCV function, $V(\alpha)$, by

$$V(\alpha) = \frac{1}{n}\sum_{k=1}^{n}(\mathscr{K}u_\alpha^{(k)}(x_k) - d_k)^2 w_k^2. \tag{4.1}$$

The weights are defined by

$$w_k = \frac{1 - a_{kk}}{(1/n)\sum(1 - a_{kk})} = \frac{1 - a_{kk}}{(1/n)\mathrm{Tr}(I - A)}$$

where $a_{kk}$ is the $k$th diagonal element of $A = A(\alpha)$. Then the GCV choice for $\alpha$ is the minimiser $\hat{\alpha}_V$ of $V(\alpha)$.

The expression (4.1) for $V(\alpha)$ however is not suitable for practical computation. Instead one uses the equivalent practical expression

$$V(\alpha) = \frac{\frac{1}{n}\sum_{k=1}^{n}(\mathscr{K}u_\alpha(x_k) - d_k)^2}{[\frac{1}{n}\sum_{k=1}^{n}(1 - a_{kk})]^2} = \frac{(1/n)|(I - A)\,\mathbf{d}|^2}{[(1/n)\mathrm{Tr}(I - A)]^2}. \tag{4.2}$$

This method of generalised cross validation was proposed by Wahba [19] (see also [3]).

The weights in (4.1) and equivalently the denominator in (4.2) are defined so that $EV(\alpha)$ behaves like $ER(\alpha)$ in a neighbourhood of the minimum of $ER(\alpha)$. In fact we have the following results.

THEOREM 4.1. *Under assumptions* (2.1) *and* (2.2), *suppose that* $\alpha \to 0$ *as* $n \to \infty$ *as in Theorem 2.1. If furthermore* $n\alpha^{1/2p} \to \infty$, *then*

$$EV(\alpha) \sim ER(\alpha) + \sigma^2.$$

This result was foreshadowed in [19] and [3], and proved rigorously in [13].

Suppose that $\alpha_V$ minimises $EV(\alpha)$ and $\alpha_0$ minimises $ER(\alpha)$. Because $n\alpha_0^{1/2p} \to \infty$ (see [13]), Theorem 4.1 shows that the graph of $EV(\alpha)$ tracks that of $ER(\alpha)$ in a neighbourhood of $\alpha_0$ as depicted in Figure 2. Furthermore, for sufficiently large $n$, the value of $ER(\alpha_V)$ will only be a fraction greater than $ER(\alpha_0)$. This comes from the following.
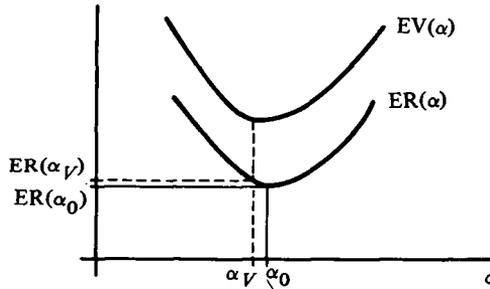
FIGURE 2

THEOERM 4.2. *Under assumptions* (2.1) *and* (2.2), *suppose that* $\alpha_V$ *and* $\alpha_0$ *both tend to* 0 *as in Theorem* 2.1. *If furthermore* $n\alpha_V^{1/2p} \to \infty$, *then the "inefficiency ratio"* $ER(\alpha_V)/ER(\alpha_0)$ *satisfies*

$$\lim_{n \to \infty} \frac{ER(\alpha_V)}{ER(\alpha_0)} = 1.$$

Again, this result was foreshadowed in [19] and [3], and proved rigorously in [13]. See also [5], [6] for this result in the convolution case.

Suppose that $\hat{\alpha}_V$ minimises $V(\alpha)$ and $\hat{\alpha}_0$ minimises $R(\alpha)$. That is $\hat{\alpha}_V$ is the actual GCV estimate and $\hat{\alpha}_0$ is the optimal regularisation parameter for the given data set. Numerical experiments in [21] indicate that with realistic noise levels (say $\sigma \leq 0.1$ and $\int_0^1 |f(t)|\,dt \approx 1$) and with $n \geq 30$ (approximately), the inefficiency ratio $R(\hat{\alpha}_V)/R(\hat{\alpha}_0)$ is seldom greater than 2, over many independent data sets. This demonstrates that the GCV estimate is generally very reliable.

If however, the noise level is higher or the number of data points smaller, then the GCV estimate may not be reliable. In this case, it has often been observed that there are several local minima of $V(\alpha)$ vying for the global minimum. Therefore one should perform the minimisation carefully. Furthermore it is possible that the regularised solution corresponding to the global minimum is not as good as the regularised solution corresponding to a comparable local minimum. In this case, it is advisable to do some experimentation with the regularised solution.

The GCV function is also a function of the order of regularisation $m$, so that $V = V(m, \alpha)$. If $V(m, \alpha)$ is minimised with respect to both $m$ and $\alpha$, then a good estimate of the optimal order is obtained. In [12], it is shown that asymptotically, the minimiser $m_V$ of $EV(m, \alpha)$ behaves like the optimal filtering order, that is, $\theta + (1/2)$ in the convolution case.

# Acknowledgement

# References

[1]   R. S. Anderssen, "Application and numerical solution of Abel-type integral equations", *MRC Technical Summary Report No. 1987*, University of Wisconsin-Madison (1977).

[2]   D. D. Cox, "Approximation of method of regularisation estimators", *University of Wisconsin-Madison, Department of Statistics, Technical Report No. 723* (1983).

[3]   P. Craven and G. Wahba, "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalised cross-validation", *Numer. Math.* **31** (1979) 377–403.

[4]   J. Cullum, "The effective choice of the smoothing norm in regularisation", *Math. Comp.* **33** (1979) 149–170.

[5]   A. R. Davies and R. S. Anderssen, "Optimisation in the regularisation of ill-posed problems", *J. Austral. Math. Soc. Ser. B* **28** (1986) 114–133.

[6]   A. R. Davies and R. S. Anderssen, "Improved estimates of statistical regularisation parameters in Fourier differentiation and smoothing", *Numer. Math.* **48** (1986) 671–697.

[7]   F. R. de Hoog, "Review of Fredholm equations of the first kind", in *The application and numerical solution of integral equations* (eds. R. S. Anderssen, F. R. de Hoog and M. A. Lukas), (Sijthoff & Noordhoff, 1980).

[8]   M. Gehatia and D. R. Wiff, "Solution of Fujita's equation for equilibrium sedimentation by applying Tikhonov's regularising functions", *J. Polymer Science: Part A-2* **8** (1970) 2039–2050.

[9]   M. A. Lukas, "Regularisation", in *The application and numerical solution of integral equations* (eds. R. S. Anderssen, F. R. de Hoog and M. A. Lukas), (Sijthoff & Noordhoff, 1980).

[10]  M. A. Lukas, "Regularisation of linear operator equations", Ph.D. Thesis, Australian National University, Canberra (1981).

[11]  M. A. Lukas, "Convergence rates for regularised solutions", to appear in Math. Comp. (1988).

[12]  M. A. Lukas, "Optimal filtering and the order of regularisation", in preparation.

[13]  M. A. Lukas, "Asymptotic behaviour of practical estimates of the regularisation parameter", in preparation.

[14]  M. Z. Nashed and G. Wahba, "Generalised inverses in reproducing kernel spaces: an introduction to regularisation of linear operator equations", *SIAM J. Math. Anal.* **5** (1974) 974–987.

[15]  F. Natterer, "Error bounds for Tikhonov regularisation in Hilbert scales", *Applicable Anal.* **18** (1984) 29–37.

[16]  G. N. Newsam, "Measures of information in linear ill-posed problems", Australian National University, Centre for Mathematical Analysis, Research Report No. 28 (1984).

[17]  J. Rice and M. Rosenblatt, "Smoothing splines: regression, derivatives and deconvolution", *Ann. Statist.* **11** (1983) 141–156.

[18]  A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, 1977).

[19]  G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy", *SIAM J. Numer. Anal.* **14** (1977) 651–667.
[20]  G. Wahba, "Ill-posed problems: numerical and statistical methods for mildy, moderately and severely ill-posed problems with noisy data", *University of Wisconsin-Madison, Department of Statistics, Technical Report No. 595* (1980).
[21]  G. Wahba, "A comparison of GCV and GML for choosing the smoothing parameter in the generalised spline smoothing problem", *Ann. Statist.* **13** (1985) 1378–1402.