

ON THE DISCREPANCY PRINCIPLE AND GENERALISED MAXIMUM LIKELIHOOD FOR REGULARISATION

MARK A. LUKAS

Let $f_{n\lambda}$ be the regularised solution of a general, linear operator equation, $Kf_0 = g$, from discrete, noisy data $y_i = g(x_i) + \varepsilon_i$, $i = 1, \dots, n$, where ε_i are uncorrelated random errors with variance σ^2 . In this paper, we consider the two well-known methods – the discrepancy principle and generalised maximum likelihood (GML), for choosing the crucial regularisation parameter λ . We investigate the asymptotic properties as $n \rightarrow \infty$ of the “expected” estimates λ_D and λ_M corresponding to these two methods respectively. It is shown that if f_0 is sufficiently smooth, then λ_D is weakly asymptotically optimal (ao) with respect to the risk and an L^2 norm on the output error. However, λ_D oversmooths for all sufficiently large n and also for all sufficiently small σ^2 . If f_0 is not too smooth relative to the regularisation space W , then λ_D can also be weakly ao with respect to a whole class of loss functions involving stronger norms on the input error. For the GML method, we show that if f_0 is smooth relative to W (for example $f_0 \in W^{\theta,2}$, $\theta > m$, if $W = W^{m,2}$), then λ_M is asymptotically sub-optimal and undersmoothing with respect to all of the loss functions above.

1. INTRODUCTION

We consider the problem of estimating the solution of a linear operator equation (scaled onto $[0, 1]$)

$$(1.1) \quad Kf(x) = g(x), \quad x \in [0, 1],$$

given only discrete, noisy data

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \dots, n.$$

The errors ε_i are modelled as uncorrelated random variables with mean 0 and common variance σ^2 . We assume that $K : L^2(0,1) \rightarrow L^2(0,1)$ is 1–1 and let f_0 be the unique solution of (1.1).

Received 31st January, 1995

Most of the work for this paper was completed while the author was on sabbatical from Murdoch University and visiting the Centre for Mathematical Analysis, The Australian National University, and the Johannes–Kepler–Universität, Linz. He is grateful for the support of these institutions and wishes to thank, in particular, Dr. R.S. Anderssen, Professor N.S. Trudinger and Professor H.W. Engl.

Copyright Clearance Centre, Inc. Serial-fee code: 0004-9729/95 \$A2.00+0.00.

To estimate f_0 , we use the method of regularisation, which defines a regularised solution $f_{n\lambda}$ to be the minimiser of

$$(1.2) \quad n^{-1} \sum_{i=1}^n (Kf(x_i) - y_i)^2 + \lambda \|Pf\|_W^2$$

over a Hilbert space $W \subset L^2(0,1)$. Here $P : W \rightarrow W$ is either the identity or an orthogonal projection of W onto $\text{span}\{\theta_1, \dots, \theta_m\}^\perp$ for some $\theta_i \in W$. This includes the important case where W is the Sobolev space $W^{m,2}[0,1]$ and $\|Pf\|_W^2 = \|f^{(m)}\|_{L^2}^2$, in which m is called the order of regularisation (see [23]). This squared norm term serves to penalise non-smooth approximate solutions, and the regularisation parameter $\lambda > 0$ controls the tradeoff between smoothing the solution and being faithful to the data, a larger parameter giving a smoother solution.

There are two important cases of (1.1) that should be mentioned. One is the class of integral equations of the first kind

$$Kf(x) = \int_0^1 k(x,t)f(t) dt = g(x),$$

examples of which arise in a wide variety of applications. The other case is with $K = I$, the identity, which corresponds to the problem of smoothing the data $y_i = f(x_i) + \varepsilon_i$. In this case, if $\|Pf\|_W^2 = \|f^{(m)}\|_{L^2}^2$, then the regularised solution $f_{n\lambda}$ is known to be just the natural polynomial smoothing spline estimate of degree $2m - 1$.

From computational experience, it is known that $f_{n\lambda}$ can be a very good approximate solution of (1.1) provided a good value of the regularisation parameter λ is used. This is supported by general theoretical results in [2, 11] and [18] which show that if the sequence $\lambda = \lambda(n) \rightarrow 0$, but not too quickly, then $f_{n\lambda}$ converges to f_0 as $n \rightarrow \infty$ in the sense that $E \|f_{n\lambda} - f_0\|_W^2 \rightarrow 0$, where E denotes expectation.

For the regularised solution to be most useful in practice, there should be a reliable method to estimate a good value of λ from the data. Several methods have been proposed, most notably generalised cross-validation (GCV) (Wahba [21]), the discrepancy principle (Morozov [16]) and generalised maximum likelihood (GML) (Anderssen and Bloomfield [1] and Wahba [22]).

In a theoretical assessment of these methods, it is essential to consider their asymptotic behaviour as $n \rightarrow \infty$. In [14], it is shown that the GCV method rates very well in this respect; in fact, the “expected” GCV estimate λ_V is asymptotically optimal with respect to several natural loss functions. It is the purpose of this paper to investigate the asymptotic properties of the corresponding estimates for the discrepancy principle and GML.

To describe these methods, first define the operator $K_n : W \rightarrow \mathbb{R}^n$ by $K_n f_i = Kf(x_i)$, $i = 1, \dots, n$, and define the influence matrix A by the condition that $K_n f_{n\lambda} = Ay$ for any data y . In our framework, the (squared) discrepancy $D(\lambda)$ is defined to be the normalised sum of squared residuals

$$(1.3) \quad D(\lambda) = n^{-1} \|K_n f_{n\lambda} - y\|^2 = n^{-1} \|(I - A)y\|^2,$$

where $\|\cdot\|$ denotes the Euclidean norm, and the discrepancy principle estimate $\hat{\lambda}_D$ is the unique solution of the non-linear equation $D(\lambda) = \sigma^2$. (A unique solution exists in all reasonable problems – see section 3.) Note that the variance σ^2 , or a good estimate of it, must be known.

The discrepancy principle was first proposed and investigated by Morozov [16]. It has been widely used and studied, and several modified versions have been developed (see [8, 17] and their references). However, in all of these works a deterministic rather than the present probabilistic framework is used.

The GML estimate $\hat{\lambda}_M$ is defined to be the minimiser over $\lambda > 0$ of the GML function

$$(1.4) \quad M(\lambda) = \frac{y^T(I - A)y}{[\det^+(I - A)]^{1/(n-m)}},$$

where $\det^+(I - A)$ means the product of the non-zero eigenvalues of $I - A$. Note that, like the GCV estimate, the GML estimate has the advantage of not requiring σ^2 .

A maximum likelihood estimate for choosing λ was first proposed by Anderssen and Bloomfield [1] in the context of numerical differentiation of time series (see also Davies [4]). This was generalised to the above GML estimate by Wahba [22].

Define the “expected” discrepancy principle estimate λ_D to be the solution of $ED(\lambda) = \sigma^2$. (Again a unique solution exists in all reasonable problems – see section 3.) Also define the “expected” GML estimate λ_M to be the minimiser over a certain set $[\lambda_n, \infty)$ (to be defined later) of $EM(\lambda)$. We wish to determine the asymptotic properties of λ_D and λ_M as $n \rightarrow \infty$.

We use the following loss functions (the same as those used in [14]): the risk

$$ER(\lambda) = En^{-1} \|K_n f_{n\lambda} - g\|^2,$$

the corresponding $L^2(F)$ norm of the output error

$$EL(\lambda) = E \|K f_{n\lambda} - g\|_{L^2(F)}^2,$$

where F is an appropriate distribution function, and the family of loss functions $EL_\rho(\lambda)$ (defined in (2.15)) involving stronger norms on the input error $f_{n\lambda} - f_0$. These norms

correspond to spaces W_ρ , where ρ can be thought of as a smoothness index. (In some cases, as shown in [2] and [12], W_ρ is equivalent to the Sobolev space $W^{\rho,2}[0,1]$.)

Define the inefficiency ratio of λ_D with respect to $ER(\lambda)$ to be

$$I_{D/R} = ER(\lambda_D) / [\min_{\lambda \geq \lambda_n} ER(\lambda)],$$

and similarly define $I_{M/R}$ for λ_M . Inefficiency ratios with respect to the other loss functions can be defined in the same way. Estimates of the minimum values of the loss functions appearing in the denominator are known (see [2] and [14]) and are reviewed in section 2.

In the special case of periodic data smoothing, Wahba [20] shows that if f_0 is sufficiently smooth, then λ_D is weakly asymptotically optimal (ao) with respect to the risk, in that $I_{D/R} = O(1)$ as $n \rightarrow \infty$. This result was extended to periodic numerical differentiation and certain convolution integral equations by Davies and Anderssen [7].

In Theorems 3.1 and 3.2, we show that if $f_0 \in W_\beta$, where either $\beta \geq 2$ or $\beta < 2$ and a certain upper bound is achieved, then λ_D is weakly ao in general with respect to both the risk and $EL(\lambda)$. If also $f_0 \notin W_{2+\delta}$, $\delta > 0$, (which can be interpreted as meaning that the order of regularisation m is sufficiently high relative to the smoothness of f_0) then, from Theorem 3.3, λ_D is also weakly ao with respect to $EL_\rho(\lambda)$ for a range of $\rho \geq 0$. However, if $f_0 \in W_\beta$, $\beta > 2$, then λ_D is asymptotically sub-optimal with respect to $EL_\rho(\lambda)$ for any $\rho > 0$ in a certain range.

In [20] and [7], it is also shown that if f_0 is smooth, then λ_D always oversmooths in the special cases above, meaning that for all n sufficiently large, $\lambda_D > \lambda_R$, where λ_R minimises $ER(\lambda)$. In Theorem 3.4, we show that if $f_0 \in W_\beta$, $\beta \geq 3$, then this result is true in general. Theorem 3.5 shows that λ_D also oversmooths for all sufficiently small σ^2 . Numerical results in [3] for the case of data smoothing are consistent with these theoretical results.

For the "expected" GML estimate λ_M , Wahba [22] shows that if f_0 is smooth in a certain sense and $Pf_0 \neq 0$, then $I_{M/R} \rightarrow \infty$ as $n \rightarrow \infty$. However, her argument depends on several heuristic assumptions about the asymptotic behaviour of certain functions of n and λ . In the case of periodic numerical differentiation of order k , Davies and Anderssen [6] prove that if

$$(1.5) \quad \sum_{\nu=0}^{\infty} \nu^{4(m+k)} |\hat{g}_\nu|^2 < \infty,$$

where \hat{g}_ν is the ν th Fourier coefficient of g , then $I_{M/R} \rightarrow \infty$ as $n \rightarrow \infty$. In this situation, from the discussion before Corollary 5.2 in [14], the condition (1.5) is equivalent to our condition $f_0 \in W_2$.

In Theorems 4.1 and 4.2, we show that in general, if $f_0 \in W_\beta$, with $\beta > 1$ (which means that f_0 is smooth relative to W) and $Pf_0 \neq 0$, then $I_{M/R} \rightarrow \infty$ as $n \rightarrow \infty$; that is, λ_M is asymptotically sub-optimal with respect to the risk. In addition, λ_M always undersmooths in that for all n sufficiently large, $\lambda_M < \lambda_R$, where λ_R minimises $ER(\lambda)$. Theorem 4.3 shows that λ_M is also asymptotically sub-optimal and undersmoothing with respect to $EL(\lambda)$ and any of the loss functions $EL_\rho(\lambda)$, $\rho \geq 0$. Numerical experiments in [22] for data smoothing and in [5] for numerical deconvolution do not conclusively illustrate these asymptotic results, although in [22] $I_{M/R}$ is consistently greater than $I_{V/R}$, the inefficiency of the GCV estimate.

If f_0 is “rough” relative to W , then it is known (see [22]) that λ_M can be ao . However, usually the smoothness of f_0 is unknown. From the above results, one can therefore conclude that the GML estimate has the disadvantage that its asymptotic performance is overly dependent on the unknown smoothness of the solution f_0 .

2. PRELIMINARY RESULTS

It is known (see [9]) that under mild conditions, there is a unique regularised solution $f_{n\lambda}$. Assume that for each $x \in [0, 1]$, the linear functional $f \rightarrow Kf(x)$ is bounded from $W \rightarrow \mathbb{R}$. Then $K_n : W \rightarrow \mathbb{R}^n$, $K_n f_i = Kf(x_i)$, is bounded, and let $K_n^* : \mathbb{R}^n \rightarrow W$ be the adjoint with respect to the inner product $n^{-1}(\cdot, \cdot)$, where (\cdot, \cdot) denotes the Euclidean inner product on \mathbb{R}^n . If $N(K_n) \cap N(P) = 0$, then the Euler-Lagrange equation for the minimisation of (1.2) gives the unique regularised solution

$$f_{n\lambda} = (\lambda P + U_n)^{-1} K_n^* y,$$

where $U_n = K_n^* K_n$.

An alternative form for $f_{n\lambda}$ can be obtained as follows (see [23]). Since for each x the linear functional $f \rightarrow Kf(x)$ is bounded, there exists a representer $\eta_x \in W$ such that for all $f \in W$, $Kf(x) = (f, \eta_x)_W$. Let $\xi_x = P\eta_x$, and denote $\eta_i = \eta_{x_i}$ and $\xi_i = \xi_{x_i}$. Define the matrices T and Σ by $T_{ij} = K\theta_j(x_i)$ and $\Sigma_{ij} = (\xi_i, \xi_j)_W$. If $N(K_n) \cap N(P) = \{0\}$, then T has full rank m and there exists a matrix B such that $BB^T = I_{n-m}$ and $BT = 0_{n-m \times m}$. The regularised solution can then be expressed as

$$(2.1) \quad f_{n\lambda} = \sum_{i=1}^m a_i \theta_i + \xi^T B^T (B \Sigma B^T + n \lambda I)^{-1} B y,$$

where $\mathbf{a} = (a_1, \dots, a_m)^T$ is the unique solution of

$$T \mathbf{a} = \mathbf{y} - (\Sigma + n \lambda I) B^T (B \Sigma B^T + n \lambda I)^{-1} B y.$$

From this it easy to see that the influence matrix A is given by

$$I - A = n\lambda B^T(B\Sigma B^T + n\lambda I)^{-1}B.$$

The discrepancy $D(\lambda)$ and the GML function $M(\lambda)$ can be evaluated using the following spectral decomposition (see [22]). Clearly, there exists an orthogonal matrix U such that

$$B\Sigma B^T = UDU^T,$$

where D is diagonal containing the decreasing eigenvalues $\lambda_{\nu n} = D_{\nu\nu} \geq 0$, $\nu = 1, \dots, n - m$. Setting $W = B^TU$, clearly $W^TW = I_{n-m}$ and

$$(2.2) \quad I - A = n\lambda W(D + n\lambda I)^{-1}W^T.$$

Let $\mathbf{w}_\nu = (w_{1\nu}, \dots, w_{n\nu})^T$ be the ν th column of W and, for any $\mathbf{z} \in \mathbb{R}^n$, define

$$z_{\nu n} = (\mathbf{z}, \mathbf{w}_\nu) = W^T \mathbf{z}_\nu.$$

Then, from (1.3) and (2.2), we have

$$(2.3) \quad D(\lambda) = n\lambda^2 \sum_{\nu=1}^{n-m} y_{\nu n}^2 / (n\lambda + \lambda_{\nu n})^2$$

and from (1.4) and (2.2), we have

$$(2.4) \quad M(\lambda) = \frac{n\lambda \sum_{\nu=1}^{n-m} y_{\nu n}^2 / (n\lambda + \lambda_{\nu n})}{\left[\prod_{\nu=1}^{n-m} n\lambda / (n\lambda + \lambda_{\nu n}) \right]^{1/(n-m)}}.$$

Define functions $\mu_1(\lambda)$ and $\mu_2(\lambda)$ by

$$(2.5) \quad \mu_1(\lambda) = n^{-1} \text{tr} A = n^{-1} \left[m + \sum_{\nu=1}^{n-m} \lambda_{\nu n} / (n\lambda + \lambda_{\nu n}) \right],$$

$$(2.6) \quad \mu_2(\lambda) = n^{-1} \text{tr} A^2 = n^{-1} \left[m + \sum_{\nu=1}^{n-m} [\lambda_{\nu n} / (n\lambda + \lambda_{\nu n})]^2 \right].$$

These functions will play a crucial role in the asymptotic analysis of the discrepancy principle and GML estimates (as they do for the GCV estimate – see [14]).

Let q be the kernel

$$q(\mathbf{x}, \mathbf{y}) = (\eta_{\mathbf{x}}, \eta_{\mathbf{y}})_W = K\eta_{\mathbf{x}}(\mathbf{y}),$$

and let Q_n be the matrix $Q_n = [q(x_i, x_j)]$. In the case where $P = I$, it is known that

$$f_{n\lambda} = \eta(Q_n + n\lambda I)^{-1}y \quad \text{and} \\ A = Q_n(Q_n + n\lambda I)^{-1}.$$

Then the appropriate spectral decomposition is $Q_n = WDW^T$, where $\lambda_{\nu n} = D_{\nu\nu} \geq 0$, $\nu = 1, \dots, n$, and W is orthogonal.

The general, finite spectral decomposition defined above is related to the following infinite spectral decomposition of operators. The latter decomposition is also used to define a family of spaces, which were introduced in [2].

Let \mathcal{F} be the class of distribution functions F on $[0, 1]$ with density bounded away from 0 and ∞ . For any $F \in \mathcal{F}$, let $L^2(F)$ denote the space $L^2(0, 1)$ with inner product

$$(g, h)_{L^2(F)} = \int_0^1 gh \, dF.$$

Clearly, the norms $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^2(F)}$ are equivalent.

Let $K^* : L^2(F) \rightarrow W$ be the adjoint of $K : W \rightarrow L^2(F)$ and define $U = K^*K : W \rightarrow W$. Since $K : W \rightarrow L^2$ is assumed to be 1-1 and compact, U is also 1-1 and compact. Then there is a basis $\{\phi_\nu\}$ for W and eigenvalues $\{\gamma_\nu\}$ of P with respect to U such that $0 \leq \gamma_1 \leq \gamma_2 \leq \dots$, $\gamma_\nu \rightarrow \infty$,

$$(2.7) \quad (\phi_\nu, U\phi_\mu)_W = \delta_{\nu\mu}, \quad (\phi_\nu, P\phi_\mu)_W = \gamma_\nu \delta_{\nu\mu},$$

and for any $f \in W$,

$$(2.8) \quad f = \sum_{\nu=1}^{\infty} (f, U\phi_\nu)_W \phi_\nu.$$

Define an inner product on W by

$$(2.9) \quad (f, v)_\rho = \sum_{\nu=1}^{\infty} (1 + \gamma_\nu)^\rho (f, U\phi_\nu)_W (v, U\phi_\nu)_W,$$

and let W_ρ be the Hilbert space completion under the corresponding norm $\|\cdot\|_\rho$ of the set $\{f \in W : \|f\|_\rho < \infty\}$. It is shown in [14] that $W_1 = W$ with equivalent norms, and $K : W_0 \rightarrow L^2(F)$ is an isometric isomorphism.

The following properties follow easily from the above definitions. From (2.7), for all ν ,

$$(2.10) \quad P\phi_\nu = \gamma_\nu U\phi_\nu.$$

Since $\dim N(P) = m$, we have $\gamma_\nu = 0$ for $\nu = 1, \dots, m$, and $\text{span}\{\phi_\nu : \nu \leq m\} = N(P)$. Therefore, from (2.10), $\gamma_\nu > 0$ for $\nu > m$. Using (2.8), (2.10) and the orthogonality of P , it is not hard to show that

$$(2.11) \quad \|Pf\|_W^2 = \sum_{\nu=m+1}^{\infty} \gamma_\nu (f, U\phi_\nu)_W^2 = \sum_{\nu=m+1}^{\infty} \gamma_\nu (Pf, U\phi_\nu)_W^2.$$

From (2.9), it is clear that for any μ ,

$$(2.12) \quad \|P\phi_\mu\|_\rho^2 \leq \sum_{\nu=1}^{\infty} (1 + \gamma_\nu)^\rho (\phi_\mu, U\phi_\nu)_W^2 = (1 + \gamma_\mu)^\rho.$$

We now list the main assumptions for results in this paper. (These are the same as those in [14].) For convenience we shall use the following notation: for two positive sequences a_n and b_n , denote $a_n \sim b_n$ if and only if $a_n = b_n(1 + o(1))$, denote $a_n \lesssim b_n$ if and only if $a_n = O(b_n)$, and denote $a_n \approx b_n$ if and only if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

ASSUMPTION 1. The random errors ε_i , $i = 1, \dots, n$, satisfy $E\varepsilon_i = 0$ and $E\varepsilon_i\varepsilon_j = \sigma^2\delta_{ij}$.

ASSUMPTION 2.

- (a) The operator $K : W \rightarrow L^2$ is 1-1, bounded and compact, and $K(W)$ is dense in L^2 .
- (b) $P : W \rightarrow W$ is an orthogonal projection with $m = \dim N(P) < \infty$.
- (c) There exists $r > 1$ such that $\gamma_\nu \approx \nu^r$ for $\nu > m$.

ASSUMPTION 3.

- (a) For each $x \in [0, 1]$, the functional $W \rightarrow \mathbb{R}$, $f \rightarrow Kf(x)$ is bounded.
- (b) For all n sufficiently large, $N(K_n) \cap N(P) = \{0\}$.

ASSUMPTION 4. For the kernel $q(x, y) = (\eta_x, \eta_y)_W$, there exists \bar{q} such that $q(x, x) \leq \bar{q}$, for all $x \in [0, 1]$. (This is equivalent to q being uniformly bounded, since $|q(x, y)| \leq q^{1/2}(x, x)q^{1/2}(y, y)$.)

It is not hard to show that (see [14])

$$(2.13) \quad n^{-1}Q_n = K_n K_n^* \quad \text{and} \quad n^{-1}\Sigma = K_n P K_n^*.$$

Using the first equality, it is clear from Assumption 4 that for all $\mathbf{z} \in \mathbb{R}^n$,

$$(2.14) \quad \|K_n^* \mathbf{z}\|_W^2 = n^{-1}(K_n K_n^* \mathbf{z}, \mathbf{z}) \leq \bar{q} n^{-1} \|\mathbf{z}\|^2.$$

By definition of U and U_n , clearly, for all $f, v \in W$,

$$((U - U_n)f, v)_W = (Kf, Kv)_{L^2(F)} - n^{-1} \sum_{i=1}^n Kf(x_i)Kv(x_i).$$

ASSUMPTION 5. There exists $F \in \mathcal{F}$, $s \in (0, 1 - 1/\tau)$, $\{\rho_1, \dots, \rho_J\} \subseteq [0, s]$ and a sequence $d_n \rightarrow 0$ such that, for all $f, v \in W$,

$$|((U - U_n)f, v)_W| \leq d_n \sum_{j=1}^J \|f\|_{\rho_j} \|v\|_{s-\rho_j}.$$

Define the general loss function $EL_\rho(\lambda)$ by

$$(2.15) \quad EL_\rho(\lambda) = E \|f_{n\lambda} - f_0\|_\rho^2.$$

Note that for $\rho = 0$, from the isometric isomorphism $K : W_0 \rightarrow L^2(F)$,

$$EL_0(\lambda) = EL(\lambda) \equiv E \|K f_{n\lambda} - g\|_{L^2(F)}^2.$$

Using Assumption 1, $EL_\rho(\lambda)$ can be decomposed as

$$EL_\rho(\lambda) = \|B_{n\lambda}f_0\|_\rho^2 + V_{n\lambda},$$

where $\|B_{n\lambda}f_0\|_\rho^2$ is the squared bias

$$\|B_{n\lambda}f_0\|_\rho^2 = \|E f_{n\lambda} - f_0\|_\rho^2 = \left\| [(\lambda P + U_n)^{-1} U_n - I] f_0 \right\|_\rho^2$$

and $V_{n\lambda}$ is the variance

$$V_{n\lambda} = E \|f_{n\lambda} - E f_{n\lambda}\|_\rho^2 = E \left\| (\lambda P + U_n)^{-1} K_n^* \varepsilon \right\|_\rho^2.$$

Under Assumptions 1, 2, 3 and 5, estimates of the squared bias and variance are derived in [2] (see also [11] in the case where $P = I$). The results are stated in Propositions 3.1 and 3.2 in [14]. If $f_0 \in W_\beta$, $\beta \geq \max\{\rho, s\}$, then as $n \rightarrow \infty$, the bias estimate has the form

$$\begin{aligned} \min\{1, \lambda^2\} \|f_0\|_\rho^2 &\lesssim \|B_{n\lambda}f_0\|_\rho^2 \lesssim \min\{1, \lambda^{\beta-\rho}\} \|f_0\|_\beta^2 \quad \text{if } \rho \leq \beta < \rho + 2, \\ \|B_{n\lambda}f_0\|_\rho^2 &\approx \min\{1, \lambda^2\} \|f_0\|_{\rho+2}^2 \quad \text{if } \beta \geq \rho + 2, \end{aligned}$$

and the variance estimate has the form

$$V_{n\lambda} \sim \sigma^2 n^{-1} C(\lambda, \rho),$$

where

$$C(\lambda, \rho) = \sum_{\nu=1}^{\infty} (1 + \gamma_\nu)^\rho (1 + \lambda \gamma_\nu)^{-2}.$$

The estimates are uniform in $\lambda \in [\lambda_n, \infty)$, where λ_n is a certain sequence which $\rightarrow 0$. It is known that if $-1/r < \rho < 2 - 1/r$, then for all $\lambda > 0$, $C(\lambda, \rho) \approx D(\lambda; -(\rho + 1/r), -2)$, where

$$D(\lambda; a, b) = \begin{cases} \lambda^a, & \lambda \leq 1, \\ \lambda^b, & \lambda > 1. \end{cases}$$

Define S to be the set of functions for which the upper bound on the squared bias is achieved in the sense that

$$\|B_{n\lambda}f_0\|_\rho^2 \approx \min\{1, \lambda^{\beta-\rho}\} \|f_0\|_\beta^2$$

for $\rho \leq \beta < \rho + 2$. It is shown in [11] that there is such a set. If λ_ρ minimises $EL_\rho(\lambda)$ and either $f_0 \in W_\beta$, $\beta \geq \rho + 2$, or $f_0 \in S \cap W_\beta$, $\rho < \beta < \rho + 2$, then from Corollary 3.1 in [14], $\lambda_\rho \approx \lambda_\rho^*$ and $EL_\rho(\lambda_\rho) \approx EL_\rho(\lambda_\rho^*) \rightarrow 0$ as $n \rightarrow \infty$, where

$$(2.16) \quad \lambda_\rho^* = \begin{cases} (\sigma^2 n^{-1})^{\rho/(\beta r + 1)}, & \rho < \beta < \rho + 2, \\ (\sigma^2 n^{-1})^{\rho/(2r + \rho r + 1)}, & \beta \geq \rho + 2. \end{cases}$$

Theorems 4.1 and 4.3 in [14] provide estimates of the functions $\mu_1(\lambda)$ and $\mu_2(\lambda)$ defined in (2.5) and (2.6) respectively. The results are that under Assumptions 2–5, if $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ in such a way that $d_n^2 \lambda_n^{-(s+1)} \rightarrow 0$, then

$$(2.17) \quad \mu_1(\lambda) \sim n^{-1} \sum_{\nu=1}^{\infty} (1 + \lambda \gamma_\nu)^{-1} \approx n^{-1} D(\lambda; -1/r, -1),$$

$$(2.18) \quad \mu_2(\lambda) \sim n^{-1} C(\lambda, 0) \approx n^{-1} D(\lambda; -1/r, -2),$$

both uniformly in $\lambda \in [\lambda_n, \infty)$.

We shall also need estimates of the risk loss function $ER(\lambda)$. Using Assumption 1, this can be decomposed as

$$(2.19) \quad ER(\lambda) = b^2(\lambda) + v(\lambda),$$

where $b^2(\lambda)$ is the squared bias

$$(2.20) \quad b^2(\lambda) = n^{-1} \|(I - A)g\|^2 = n\lambda^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / (n\lambda + \lambda_{\nu n})^2$$

and $v(\lambda)$ is the variance

$$(2.21) \quad v(\lambda) = n^{-1} E \|A\varepsilon\|^2 = \sigma^2 \mu_2(\lambda).$$

Clearly, (2.18) gives an estimate of $v(\lambda)$. Upper and lower bounds for $b^2(\lambda)$ are derived in Theorem 4.5 in [14]. These are of the same order as those for $\|B_{n\lambda}f_0\|_0^2 = \|EKf_{n\lambda} - g\|_{L^2(F)}^2$. If λ_R minimises $ER(\lambda)$ and either $f_0 \in W_\beta$, $\beta \geq 2$, or $f_0 \in S \cap W_\beta$, $s \leq \beta < 2$, then from Corollary 4.1 in [14], $\lambda_R \approx \lambda^*$ and $ER(\lambda_R) \approx ER(\lambda^*) \rightarrow 0$ as $n \rightarrow \infty$, where

$$(2.22) \quad \lambda^* = \lambda_0^* = \begin{cases} (\sigma^2 n^{-1})^{r/(\beta r + 1)}, & 0 < \beta < 2, \\ (\sigma^2 n^{-1})^{r/(2r + 1)}, & \beta \geq 2. \end{cases}$$

3. ASYMPTOTIC PROPERTIES OF λ_D

From (2.3), the discrepancy $D(\lambda)$ is strictly increasing with λ from $D(0) = 0$, and $D(\lambda) \rightarrow n^{-1} \sum_{\nu=1}^{n-m} y_{\nu n}^2$ as $\lambda \rightarrow \infty$. Hence, if $\sigma^2 < n^{-1} \sum_{\nu=1}^{n-m} y_{\nu n}^2$ (otherwise the signal would be totally dominated by noise), then the equation $D(\lambda) = \sigma^2$ defines a unique solution $\hat{\lambda}_D$, the discrepancy principle estimate.

Taking the expectation of $D(\lambda)$ and using Assumption 1 gives

$$\begin{aligned} ED(\lambda) &= b^2(\lambda) + \sigma^2 n^{-1} \text{tr}(I - A)^2 \\ &= b^2(\lambda) + \sigma^2 - \sigma^2 [2\mu_1(\lambda) - \mu_2(\lambda)], \end{aligned}$$

where b^2 , μ_1 and μ_2 are defined in (2.20), (2.5) and (2.6) respectively. Therefore, the equation $ED(\lambda) = \sigma^2$ is equivalent to

$$(3.1) \quad b^2(\lambda) = \sigma^2 [2\mu_1(\lambda) - \mu_2(\lambda)].$$

Clearly, $b^2(\lambda)$ is strictly increasing from $b^2(0) = 0$, and $b^2(\lambda) \rightarrow n^{-1} \sum_{\nu=1}^{n-m} g_{\nu n}^2$ as $\lambda \rightarrow \infty$. In addition, it is easy to show that

$$2\mu_1(\lambda) - \mu_2(\lambda) = 1 - n^{-1} \sum_{\nu=1}^{n-m} [n\lambda / (n\lambda + \lambda_{\nu n})]^2,$$

which is positive and strictly decreasing from $2\mu_1(0) - \mu_2(0) = 1$, and $2\mu_1(\lambda) - \mu_2(\lambda) \rightarrow mn^{-1}$ as $\lambda \rightarrow \infty$. Hence, if

$$(3.2) \quad \sigma^2 mn^{-1} < n^{-1} \sum_{\nu=1}^{n-m} g_{\nu n}^2,$$

then (3.1) has a unique solution λ_D , the ‘‘expected’’ discrepancy principle estimate.

Lemma 4.4 in [14] shows that under Assumptions 2 and 5, if $f_0 \in W_\beta$, $\beta \geq s$, then $n^{-1} \sum_{\nu=1}^{n-m} g_{\nu n}^2 \sim \|\bar{P}g\|_{L^2(F)}^2$ as $n \rightarrow \infty$, where $\bar{P} : L^2(F) \rightarrow L^2(F)$ is the orthogonal projection onto $\text{span}\{K\theta_i\}^\perp$. Therefore, no matter how large σ^2 is, (3.2) is satisfied for all sufficiently large n .

The following theorem describes the asymptotic behaviour of $\lambda_D = \lambda_D(n)$ as $n \rightarrow \infty$.

THEOREM 3.1. *Suppose that Assumptions 1-5 hold, $f_0 \in W_\beta$, $\beta \geq s$, and $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ in such a way that $n^{-1}\lambda_n^{-1/\tau} \rightarrow 0$ and*

$$\begin{aligned} d_n^2 \lambda_n^{-(s+1)} &\rightarrow 0, & \text{if } s \leq \beta < 3 - 1/\tau, \\ d_n^2 \lambda_n^{-(s+1/\tau+\beta-2)} &\rightarrow 0, & \text{if } \beta \geq 3 - 1/\tau. \end{aligned}$$

Let λ^* be as defined in (2.22), and assume that $\lambda^* \geq \lambda_n$ and $\lambda_D \geq \lambda_n$. If $\beta < 2$, then as $n \rightarrow \infty$,

$$\lambda^* \lesssim \lambda_D \lesssim (\sigma^2 n^{-1})^{\tau/(2r+1)},$$

and if also $f_0 \in S$, then $\lambda_D \approx \lambda^*$. If $\beta \geq 2$, then $\lambda_D \approx \lambda^*$ as $n \rightarrow \infty$.

PROOF: From (2.17) and (2.18), if $\lambda \geq \lambda_n$, then as $n \rightarrow \infty$,

$$\begin{aligned} \sigma^2 [2\mu_1(\lambda) - \mu_2(\lambda)] &\sim \sigma^2 n^{-1} \left[2 \sum_{\nu=1}^{\infty} (1 + \lambda\gamma_\nu)^{-1} - \sum_{\nu=1}^{\infty} (1 + \lambda\gamma_\nu)^{-2} \right] \\ &= \sigma^2 n^{-1} \left[\sum_{\nu=1}^{\infty} (1 + \lambda\gamma_\nu)^{-1} + \lambda \sum_{\nu=1}^{\infty} \gamma_\nu (1 + \lambda\gamma_\nu)^{-2} \right] \\ &\approx \sigma^2 n^{-1} [D(\lambda; -1/\tau, -1) + \lambda D(\lambda; -(1 + 1/\tau), -2)] \\ (3.3) \quad &\approx \sigma^2 n^{-1} D(\lambda; -1/\tau, -1). \end{aligned}$$

Using the lower bound on $b^2(\lambda)$ in Theorem 4.5 in [14], there exists a constant $c > 0$, independent of n and λ , such that $\lambda_D \leq \bar{\lambda}_D$, where $\bar{\lambda}_D = \bar{\lambda}_D(n)$ is the unique solution of

$$(3.4) \quad c \min\{1, \lambda^2\} = \sigma^2 [2\mu_1(\lambda) - \mu_2(\lambda)].$$

This implies that for all n sufficiently large, $\bar{\lambda}_D < 1$. Otherwise, there would be a subsequence $\bar{\lambda}'_D$ of $\bar{\lambda}_D$ such that $\bar{\lambda}'_D \geq 1$, but from (3.3) and (3.4), $c \approx \sigma^2 n^{-1} \bar{\lambda}'_D^{-1}$ and so $\bar{\lambda}'_D \rightarrow 0$, a contradiction. Hence, solving for $\bar{\lambda}_D$ from (3.3) and (3.4), we get $\lambda_D \leq \bar{\lambda}_D$, where

$$\bar{\lambda}_D \approx (\sigma^2 n^{-1})^{2r/(2r+1)}.$$

Again from Theorem 4.5 in [14], since $\lambda_D < 1$, we have as $n \rightarrow \infty$,

$$b^2(\lambda_D) \lesssim \lambda_D^\beta \|f_0\|_\beta^2, \quad \text{if } \beta < 2,$$

$$b^2(\lambda_D) \approx \lambda_D^2 \|f_0\|_2^2, \quad \text{if } \beta \geq 2,$$

and \approx holds for $\beta < 2$ if also $f_0 \in S$. The required results for λ_D follow easily from these estimates, (3.3) and the defining equation (3.1). \square

Using Theorem 3.1, we obtain the following results about the asymptotic inefficiency of λ_D .

THEOREM 3.2. *Suppose that Assumptions 1–5 hold, $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ as in Theorem 3.1, and $\lambda^* \geq \lambda_n$ and $\lambda_D \geq \lambda_n$. Let λ_R and λ_L minimise $ER(\lambda)$ and $EL(\lambda)$ respectively over $\lambda \geq \lambda_n$. If $f_0 \in W_\beta$, $\beta \geq 2$, or $f_0 \in S \cap W_\beta$, $s \leq \beta < 2$, then as $n \rightarrow \infty$,*

$$\lambda_D \approx \lambda_R \approx \lambda_L \approx \lambda^*,$$

$$\frac{ER(\lambda_D)}{ER(\lambda_R)} = O(1) \quad \text{and} \quad \frac{EL(\lambda_D)}{EL(\lambda_L)} = O(1).$$

PROOF: It is easy to check that λ_n satisfies the conditions in Corollaries 3.1 (with $\rho = 0$) and 4.1 in [14]. From these and Theorem 3.1, we have

$$\lambda_D \approx \lambda^* \approx \lambda_R \approx \lambda_L,$$

$$ER(\lambda_D) \approx ER(\lambda^*) \approx ER(\lambda_R),$$

$$EL(\lambda_D) \approx EL(\lambda^*) \approx EL(\lambda_L),$$

and the result follows. \square

THEOREM 3.3. *Suppose that Assumptions 1–5 hold, $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ as in Theorem 3.1, and $\lambda^* \geq \lambda_n$ and $\lambda_D \geq \lambda_n$. Let λ_ρ minimise $EL_\rho(\lambda)$ over $\lambda \geq \lambda_n$. If $f_0 \notin W_{2+\delta}$, $\delta > 0$, and either $f_0 \in W_\beta$, $\beta = 2$, or $f_0 \in S \cap W_\beta$, $s \leq \beta < 2$, then for any $0 \leq \rho < \min\{\beta, 2 - s - 1/r\}$, $\lambda_D \approx \lambda_\rho$ and*

$$\frac{EL_\rho(\lambda_D)}{EL_\rho(\lambda_\rho)} = O(1).$$

However, if $f_0 \in W_\beta$, $\beta > 2$, then for any $0 < \rho < 2 - s - 1/r$, $\lambda_D/\lambda_\rho \rightarrow 0$ as $n \rightarrow \infty$ and

$$\frac{EL_\rho(\lambda_D)}{EL_\rho(\lambda_\rho)} \rightarrow \infty.$$

(Note that $2 - s - 1/\tau > 1$.)

PROOF: Clearly λ_n satisfies the conditions of Corollary 3.1 in [14]. For the first part, from this corollary and Theorem 3.1, we have

$$\lambda_D \approx (\sigma^2 n^{-1})^{\tau/(\beta\tau+1)} \approx \lambda_\rho,$$

$$EL_\rho(\lambda_D) \approx EL_\rho(\lambda_\rho),$$

and the result follows. For the second part, from Theorem 3.1 and Corollary 3.1 in [14], we have

$$\frac{\lambda_D}{\lambda_\rho} \approx \frac{\lambda^*}{\lambda_\rho^*} = \begin{cases} (\sigma^2 n^{-1})^{\tau/(2\tau+1)-\tau/(2\tau+\rho\tau+1)}, & 0 < \rho < \beta - 2, \\ (\sigma^2 n^{-1})^{\tau/(2\tau+1)-\tau/(\beta\tau+1)}, & \rho \geq \beta - 2, \end{cases}$$

$$\rightarrow 0.$$

Moreover, from Propositions 3.1 and 3.2 in [14], $EL_\rho(\lambda_D) \approx V_{n\lambda_D}$ and so

$$\frac{EL_\rho(\lambda_D)}{EL_\rho(\lambda_\rho)} \approx \begin{cases} (\sigma^2 n^{-1})^{\tau(2-\rho)/(2\tau+1)-2\tau/(2\tau+\rho\tau+1)}, & 0 < \rho < \beta - 2, \\ (\sigma^2 n^{-1})^{\tau(2-\rho)/(2\tau+1)-\tau(\beta-\rho)/(\beta\tau+1)}, & \rho \geq \beta - 2, \end{cases}$$

and by checking the exponents this $\rightarrow \infty$. □

These results show that in many cases λ_D is weakly asymptotically optimal with respect to the given loss functions. Theorem 3.2 extends results of Wahba [20] in the case of periodic data smoothing, and Davies and Anderssen [7] in the case of periodic numerical differentiation, both of which only use the risk criterion.

In these papers, it is also shown that if f_0 is smooth, then for all n sufficiently large, λ_D always oversmooths with respect to the risk; that is $\lambda_D > \lambda_R$, where λ_R minimises $ER(\lambda)$. Numerical experiments in [3] for data smoothing are consistent with this result. The proofs in these cases rely on the known growth rate of the eigenvalues of the m th derivative operator.

In the present situation, the eigenvalues γ_ν are only bounded above and below as $c_1\nu^\tau \leq \gamma_\nu \leq c_2\nu^\tau$ for some constants $c_1, c_2 > 0$. Nevertheless, the next theorem shows that in general, if f_0 is smooth relative to W , then for all n sufficiently large, λ_D always oversmooths with respect to the risk.

We shall use the following definition and lemma. A symmetric kernel $h(x, y)$ is said to have the smoothness of the Green's function of a self-adjoint linear differential operator of order $2p$ if

- (a) for each $\ell = 0, \dots, 2p - 2$, the derivative $(\partial^\ell/\partial x^\ell)h(x, y)$ exists and is continuous at all $(x, y) \in [0, 1] \times [0, 1]$, and for $\ell = 2p - 1$ and $2p$, $(\partial^\ell/\partial x^\ell)h(x, y)$ exists and is continuous for $x \neq y$, and

- (b) $\lim_{x \downarrow y} (\partial^{2p-1} / \partial x^{2p-1})h(x, y)$ and $\lim_{x \uparrow y} (\partial^{2p-1} / \partial x^{2p-1})h(x, y)$ exist and are bounded for all $y \in [0, 1]$.

LEMMA 3.1. *Suppose that Assumptions 2, 3 and 5 hold, and the kernel $q(x, y)$ has the smoothness properties of the Green’s function of a self-adjoint linear differential operator of order $2p$. Also suppose that $\lambda = \lambda(n) \rightarrow 0$ as $n \rightarrow \infty$ such that $d_n^2 \lambda^{-(s+1/r)} \rightarrow 0$ and $h_n^2 \lambda^{-1} \rightarrow 0$, where $h_n = \max\{x_{i+1} - x_i\}$ (assuming without loss of generality that $\{x_i\}$ are in ascending order). If $f_0 \in W_3$, then as $n \rightarrow \infty$,*

$$n^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / [\lambda_{\nu n} (n\lambda + \lambda_{\nu n})^2] \lesssim \|f_0\|_3^2.$$

PROOF: From the definition of $g_{\nu n}$, $\lambda_{\nu n}$ and Σ , we have

$$\begin{aligned} n^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / [\lambda_{\nu n} (n\lambda + \lambda_{\nu n})^2] &= n^2 \mathbf{g}^T B^T (B\Sigma B^T + n\lambda)^{-1} (B\Sigma B^T)^{-1} (B\Sigma B^T + n\lambda I)^{-1} B\mathbf{g} \\ (3.5) \quad &= \left\| n\xi^T B^T (B\Sigma B^T)^{-1} (B\Sigma B^T + n\lambda I)^{-1} B\mathbf{g} \right\|_W^2. \end{aligned}$$

Now let $f^n \in W$ denote the unique function that minimises $\|Pf\|_W^2$ over W subject to $K_n f = \mathbf{g}$. It is known (see [9, 10]) that

$$(3.6) \quad Pf^n = \xi^T B^T (B\Sigma B^T)^{-1} B\mathbf{g}.$$

Introduce any data $\mathbf{y} = \mathbf{g} + \varepsilon$, where ε satisfies Assumption 1, and let $f_{n\lambda}$ be the corresponding regularised solution. Then

$$\begin{aligned} Pf^n - EPf_{n\lambda} &= \xi^T B^T \left[(B\Sigma B^T)^{-1} - (B\Sigma B^T + n\lambda I)^{-1} \right] B\mathbf{g} \\ (3.7) \quad &= n\lambda \xi^T B^T (B\Sigma B^T)^{-1} (B\Sigma B^T + n\lambda I)^{-1} B\mathbf{g}. \end{aligned}$$

Combining (3.5) and (3.7), we get

$$n^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / [\lambda_{\nu n} (n\lambda + \lambda_{\nu n})^2] = \lambda^{-2} \|Pf^n - EPf_{n\lambda}\|_W^2.$$

The right hand side can be bounded by adding and subtracting Pf_0 as follows:

$$\begin{aligned} \lambda^{-1} \|Pf^n - EPf_{n\lambda}\|_W &\leq \lambda^{-1} \|Pf^n - Pf_0\|_W + \lambda^{-1} \|Pf_0 - EPf_{n\lambda}\|_W \\ (3.8) \quad &\leq \lambda^{-1} \|f^n - f_0\|_W + \lambda^{-1} \|B_{n\lambda} f_0\|_W. \end{aligned}$$

For the first term on the right hand side, we use a theorem in [15] (see also [19] in the case where $P = I$) which implies that

$$\lambda^{-1} \|f^n - f_0\|_W \lesssim \lambda^{-1} h_n^p,$$

and this $\rightarrow 0$ as $n \rightarrow \infty$ by assumption. From Proposition 3.1 in [14], since the W norm is equivalent to the W_1 norm, the second term on the right hand side of (3.8) is bounded by

$$\lambda^{-1} \|B_{n\lambda} f_0\|_W \lesssim \|f_0\|_3,$$

and the result follows. □

Note that in the important case where K is an integral operator with kernel $k(x, t)$ and W has a reproducing kernel $r(t, s)$, it is known (see [19]) that

$$q(x, y) = \int_0^1 \int_0^1 k(x, t)r(t, s)k(y, s) dsdt.$$

Thus the smoothness of $q(x, y)$ will depend on the smoothness of $k(x, t)$ and $r(t, s)$.

THEOREM 3.4. *Suppose that Assumptions 1, 2, 3 and 5 hold, and $q(x, y)$ has the smoothness properties of the Green’s function of a self-adjoint linear differential operator of order $2p$. Also suppose that $d_n^2 n^{r(s+1)/(2r+1)} \rightarrow 0$ and $h_n^p n^{r/(2r+1)} \rightarrow 0$, where $h_n = \max\{x_{i+1} - x_i\}$, and for all sufficiently large n , the risk $ER(\lambda)$ has a unique local and global minimum over $\lambda > 0$ at $\lambda_R = \lambda_R(n)$. If $f_0 \in W_3$, then for all n sufficiently large, $\lambda_D > \lambda_R$.*

PROOF: From (2.19) and (2.21), the risk equals $ER(\lambda) = b^2(\lambda) + \sigma^2 \mu_2(\lambda)$. Using (2.20), the derivative of $b^2(\lambda)$ can be written as

$$\frac{db^2(\lambda)}{d\lambda} = 2\lambda^{-1}b^2(\lambda) - 2\lambda^2 G_{n\lambda},$$

where

$$G_{n\lambda} = n^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / (n\lambda + \lambda_{\nu n})^3.$$

Hence, from (3.1) and (2.6), we have

$$\begin{aligned} \frac{dER(\lambda_D)}{d\lambda} &= 2\lambda_D^{-1} \sigma^2 [2\mu_1(\lambda_D) - \mu_2(\lambda_D)] - 2\lambda_D^2 G_{n\lambda_D} - 2\sigma^2 \sum_{\nu=1}^{n-m} \lambda_{\nu n}^2 / (n\lambda_D + \lambda_{\nu n})^3 \\ (3.9) \quad &> 4\lambda_D^{-1} \sigma^2 [\mu_1(\lambda_D) - \mu_2(\lambda_D)] - 2\lambda_D^2 G_{n\lambda_D}. \end{aligned}$$

From the proof of Theorem 4.3 in [14] and Theorem 3.1 above, as $n \rightarrow \infty$,

$$\begin{aligned}
 4\lambda_D^{-1}\sigma^2[\mu_1(\lambda_D) - \mu_2(\lambda_D)] &= 4\sigma^2 \sum_{\nu=1}^{n-m} \lambda_{\nu n}/(n\lambda_D + \lambda_{\nu n})^2 \\
 &\sim 4\sigma^2 n^{-1} \sum_{\nu=m+1}^{\infty} \gamma_{\nu}(1 + \lambda_D\gamma_{\nu})^{-2} \\
 &\approx 4\sigma^2 n^{-1} \lambda_D^{-(1+1/r)} \\
 &\approx 4\sigma^2 n^{-1} (\sigma^2 n^{-1})^{-(r+1)/(2r+1)} \\
 &= 4(\sigma^2 n^{-1})^{\tau/(2r+1)}.
 \end{aligned}$$

On the other hand, from Theorem 3.1 and Lemma 3.1,

$$\begin{aligned}
 2\lambda_D^2 G_{n\lambda} &\leq 2\lambda_D^2 n^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / [\lambda_{\nu n}(n\lambda_D + \lambda_{\nu n})^2] \\
 &\lesssim 2(\sigma^2 n^{-1})^{2r/(2r+1)} \|f_0\|_3^2.
 \end{aligned}$$

Therefore, from (3.9), $(d/d\lambda)ER(\lambda_D)$ is positive for all sufficiently large n . The result follows since $ER(\lambda)$ is assumed to have a unique local and global minimum for all sufficiently large n . □

It is worth noting that, as shown in [13], $ER(\lambda)$ has a unique local and global minimum if the finite sequence $ng_{\nu n}^2/\lambda_{\nu n}^2$, $\nu = 1, \dots, n - m$, is non-increasing. A heuristic argument (see [13]) can be used to indicate that this condition should hold for a large subset of functions $f_0 \in W_3$, and therefore the uniqueness assumption in Theorem 3.2 is not as restrictive as it might seem.

Lastly in this section, we consider the way in which λ_D depends on the variance σ^2 . From the proof of Theorem 3.2, one can construct a value of N such that λ_D oversmooths for all $n \geq N$, and clearly this value of N decreases as the variance σ^2 decreases. This implies that for a fixed n , λ_D oversmooths for all sufficiently small σ^2 . In the next theorem, we prove this fact directly. This behaviour was observed in numerical experiments reported in [3] for the case of data smoothing.

THEOREM 3.5. *For fixed n , if $ER(\lambda)$ has a unique local and global minimum over $\lambda > 0$ at $\lambda = \lambda_R$, then for all σ^2 sufficiently small, $\lambda_D > \lambda_R$.*

PROOF: From (3.1) and the behaviour of $b^2(\lambda)$ and $2\mu_1(\lambda) - \mu_2(\lambda)$, it is clear that $\lambda_D \downarrow 0$ as $\sigma^2 \downarrow 0$. In addition, for all sufficiently small λ ,

$$\begin{aligned}
 b^2(\lambda) &\sim n\lambda^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / \lambda_{\nu n}^2, \\
 \sigma^2[2\mu_1(\lambda) - \mu_2(\lambda)] &\sim \sigma^2,
 \end{aligned}$$

and therefore $\lambda_D \sim c\sigma$ as $\sigma \rightarrow 0$, where

$$c = \left[n \sum_{\nu=1}^{n-m} g_{\nu n}^2 / \lambda_{\nu n}^2 \right]^{-1/2}$$

Using this in the derivative of $ER(\lambda)$, we get

$$\begin{aligned} \frac{dER(\lambda_D)}{d\lambda} &= 2\lambda_D^{-1}\sigma^2[2\mu_1(\lambda_D) - \mu_2(\lambda_D)] - 2\lambda_D^2 G_{n\lambda_D} - 2\sigma^2 \sum_{\nu=1}^{n-m} \lambda_{\nu n}^2 / (n\lambda_D + \lambda_{\nu n})^3 \\ &\sim \sigma 2c^{-1} - \sigma^2 \left(2c^2 n^2 \sum_{\nu=1}^{n-m} g_{\nu n}^2 / \lambda_{\nu n}^3 + 2 \sum_{\nu=1}^{n-m} 1 / \lambda_{\nu n} \right) \\ &> 0 \end{aligned}$$

for all sufficiently small σ , and the result follows. □

4. ASYMPTOTIC PROPERTIES OF THE GML ESTIMATE

In this section, we extend the asymptotic analysis of GML carried out by Wahba [22]. Define the functions $\bar{\mu}_1(\lambda)$ and $\bar{\mu}_2(\lambda)$ by:

$$\begin{aligned} \bar{\mu}_1(\lambda) &= (n - m)^{-1}(n\mu_1(\lambda) - m) = (n - m)^{-1} \sum_{\nu=1}^{n-m} \lambda_{\nu n} / (n\lambda + \lambda_{\nu n}) \\ \bar{\mu}_2(\lambda) &= (n - m)^{-1}(n\mu_2(\lambda) - m) = (n - m)^{-1} \sum_{\nu=1}^{n-m} [\lambda_{\nu n} / (n\lambda + \lambda_{\nu n})]^2. \end{aligned}$$

Clearly $0 < \bar{\mu}_1(\lambda) < 1$ and $0 < \bar{\mu}_2(\lambda) < 1$ for $\lambda > 0$. Also $\bar{\mu}_1(\lambda)$ and $\bar{\mu}_2(\lambda)$ satisfy the same estimates as $\mu_1(\lambda)$ and $\mu_2(\lambda)$ in (2.17) and (2.18) respectively, uniformly in $\lambda \in [\lambda_n, \infty)$. Taking expectation of the GML function $M(\lambda)$ defined by (2.4), and using Assumption 1, we have

$$(4.1) \quad n^{-1}EM(\lambda) = [\lambda G(\lambda) + \bar{\sigma}^2(1 - \bar{\mu}_1(\lambda))] / d(\lambda),$$

where

$$\begin{aligned} G(\lambda) &= \sum_{\nu=1}^{n-m} g_{\nu n}^2 / (n\lambda + \lambda_{\nu n}), \\ d(\lambda) &= \left[\prod_{\nu=1}^{n-m} n\lambda / (n\lambda + \lambda_{\nu n}) \right]^{1/(n-m)}, \end{aligned}$$

and $\bar{\sigma}^2 = \sigma^2(n - m)/n$. We wish to estimate the minimiser λ_M of $EM(\lambda)$.

First consider the case where $f_0 \in \text{span}\{\theta_i\}$. Then $g_{\nu n} = 0$, $\nu = 1, \dots, n - m$, and so (4.1) can be written as

$$(4.2) \quad [\sigma^2(n - m)]^{-1} EM(\lambda) = [d(\lambda)]^{-1} (n - m)^{-1} \sum_{\nu=1}^{n-m} n\lambda / (n\lambda + \lambda_{\nu n}).$$

Arguing as in [22], the right hand side of (4.2), being the ratio of an arithmetic to a geometric mean, is bounded below by 1, and clearly, it approaches 1 as $\lambda \rightarrow \infty$. Therefore, assuming that the $\lambda_{\nu n}$ are not all equal, $EM(\lambda)$ is minimised at $\lambda_M = \infty$.

Now consider the general case where $f_0 \notin \text{span}\{\theta_i\}$. Define $G_1(\lambda)$ by

$$G_1(\lambda) = \sum_{\nu=1}^{n-m} \lambda_{\nu n} g_{\nu n}^2 / (n\lambda + \lambda_{\nu n})^2.$$

Differentiating (4.1) and using the equalities

$$\begin{aligned} d'(\lambda) &= d(\lambda)\bar{\mu}_1(\lambda)/\lambda, \\ \bar{\mu}_1(\lambda) + \lambda\bar{\mu}'_1(\lambda) &= \bar{\mu}_2(\lambda), \\ G(\lambda) + \lambda G'(\lambda) &= G_1(\lambda), \end{aligned}$$

gives

$$(4.3) \quad n^{-1} \frac{d EM(\lambda)}{d\lambda} = [\lambda G_1(\lambda) - \lambda\bar{\mu}_1(\lambda)G(\lambda) - \bar{\sigma}^2\bar{\mu}_2(\lambda) + \bar{\sigma}^2\bar{\mu}'_1(\lambda)] / [\lambda d(\lambda)].$$

By setting the derivative equal to zero and making certain assumptions about the asymptotic behaviour of $g_{\nu n}^2$, $\bar{\mu}_1(\lambda)$ and $\bar{\mu}_2(\lambda)$, Wahba [22] obtains an asymptotic decay rate for λ_M . We shall obtain the corresponding result without these assumptions. First we prove the following lemmas.

LEMMA 4.1. *Suppose that Assumptions 1, 2, 3 and 5 hold and let $f_0 \in W_\beta$, $\beta > 1$. If $\lambda = \lambda(n) \rightarrow 0$ as $n \rightarrow \infty$ in such a way that*

$$\begin{aligned} d_n^2 \lambda^{-(s+1/r)} &\rightarrow 0, & \text{if } 1 < \beta < 2, \\ d_n^2 \lambda^{-(s+1/r+\beta-2)} &\rightarrow 0, & \text{if } \beta \geq 2, \end{aligned}$$

then $G(\lambda) \sim \|Pf_0\|_W^2$ and $G_1(\lambda) \sim \|Pf_0\|_W^2$.

PROOF: As shown in [22], for any n and λ ,

$$(4.4) \quad G_1(\lambda) \leq G(\lambda) \leq \sum_{\nu=1}^{n-m} g_{\nu n}^2 / \lambda_{\nu n} = \|Pf^n\|_W^2 \leq \|Pf_0\|_W^2,$$

where f^n is the unique function in W that minimises $\|Pf\|_W^2$ subject to $Kf(x_i) = g(x_i)$, $i = 1, \dots, n$. The equality in (4.4) holds because from (3.6),

$$\|Pf^n\|_W^2 = \mathbf{g}^T B^T (B \Sigma B^T)^{-1} B \mathbf{g}.$$

Now, from the definition of $\lambda_{\nu n}$ and $g_{\nu n}$, and (2.1), we have

$$G_1(\lambda) = \left\| \xi^T B^T (B \Sigma B^T + n \lambda I)^{-1} B \mathbf{g} \right\|_W^2 = \|PEf_{n\lambda}\|_W^2.$$

Since the W norm is equivalent to the W_1 norm and $\beta > 1$, Proposition 3.1 in [14] implies that as $n \rightarrow \infty$,

$$\begin{aligned} \|PEf_{n\lambda}\|_W &\geq \|Pf_0\|_W - \|PEf_{n\lambda} - Pf_0\|_W \\ &= \|Pf_0\|_W (1 + o(1)). \end{aligned}$$

Using this together with (4.4), gives the required results. □

LEMMA 4.2. *Under Assumptions 2-5, if $f_0 \in W_\beta$, $\beta \geq s$, and $g = Kf_0$, then as $n \rightarrow \infty$,*

$$n^{-2} \sum_{\nu=1}^{n-m} \lambda_{\nu n} g_{\nu n}^2 \sim \|PK^* \bar{P}g\|_W^2,$$

where $\bar{P} : L^2(F) \rightarrow L^2(F)$ is the orthogonal projection onto $\text{span}\{K\theta_i\}^\perp$.

PROOF: Let $\bar{P}_n : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the orthogonal projection of \mathbb{R}^n onto $\text{span}\{K_n \theta_i\}^\perp$. From the definition of the matrix B , it is clear that the matrix of \bar{P}_n is $B^T B$. Then, from the definition of $\lambda_{\nu n}$ and $g_{\nu n}$, and (2.13), we have

$$\begin{aligned} n^{-2} \sum_{\nu=1}^{n-m} \lambda_{\nu n} g_{\nu n}^2 &= n^{-2} (\mathbf{g}, B^T B \Sigma B^T B \mathbf{g}) \\ &= n^{-1} (\mathbf{g}, \bar{P}_n K_n P K_n^* \bar{P}_n \mathbf{g}) \\ &= \|PK_n^* \bar{P}_n \mathbf{g}\|_W^2. \end{aligned}$$

Now $\bar{P}_n \mathbf{g} = K_n \bar{f}_0^n$, where

$$(4.5) \quad \bar{f}_0^n = f_0 - \sum_{i=1}^m c_i^n \theta_i$$

and the c_i^n are defined by the equations $(\bar{P}_n \mathbf{g}, K_n \theta_i) = 0$, $i = 1, \dots, m$, as in the proof of Lemma 4.4 in [14]. Similarly, $\bar{P}g = K \bar{f}_0$, where

$$(4.6) \quad \bar{f}_0 = f_0 - \sum_{i=1}^m c_i \theta_i$$

and the c_i are defined by the equations $(\bar{P}g, K\theta_i)_{L^2(F)} = 0, i = 1, \dots, m$. Therefore, adding and subtracting $PU_n\bar{f}_0$, we get

$$\begin{aligned}
 \left\| \left[n^{-2} \sum_{\nu=1}^{n-m} \lambda_{\nu n} g_{\nu n}^2 \right]^{1/2} - \|PK^*\bar{P}g\|_W \right\| &\leq \|PU_n\bar{f}_0^n - PU\bar{f}_0\|_W \\
 (4.7) \qquad \qquad \qquad &\leq \|P(U - U_n)\bar{f}_0\|_W + \|PU_n(\bar{f}_0 - \bar{f}_0^n)\|_W.
 \end{aligned}$$

The first term on the right hand side of (4.7) converges to 0 as $n \rightarrow \infty$, since, from (2.11), (2.10), Assumption 5 and (2.12),

$$\begin{aligned}
 \|P(U - U_n)\bar{f}_0\|_W^2 &= \sum_{\nu=m+1}^{\infty} \gamma_{\nu} ((U - U_n)\bar{f}_0, U\phi_{\nu})_W^2 \\
 &\leq d_n^2 \sum_{\nu=m+1}^{\infty} \gamma_{\nu}^{-1} \left[\sum_{j=1}^J \|\bar{f}_0\|_{\rho_j} \|P\phi_{\nu}\|_{s-\rho_j} \right]^2 \\
 &\leq d_n^2 J^2 \|\bar{f}_0\|_s^2 \sum_{\nu=m+1}^{\infty} \gamma_{\nu}^{-1} (1 + \gamma_{\nu})^s,
 \end{aligned}$$

and the sum converges because $s < 1 - 1/r$. For the second term on the right hand side of (4.7), first we have from (2.14),

$$\begin{aligned}
 \|PU_n(\bar{f}_0 - \bar{f}_0^n)\|_W^2 &\leq \|K_n^*(I_n\bar{P}g - \bar{P}_n\mathbf{g})\|_W^2 \\
 &\leq \bar{q}n^{-1} \|I_n\bar{P}g - \bar{P}_n\mathbf{g}\|^2 \\
 &= \bar{q}n^{-1} \left\| \sum_{i=1}^m (c_i - c_i^n) K_n\theta_i \right\|^2 \\
 &\leq \bar{q} \sum_{i=1}^m (c_i - c_i^n)^2 \sum_{i=1}^m n^{-1} \|K_n\theta_i\|^2,
 \end{aligned}$$

where $(I_n\bar{P}g)_j = \bar{P}g(x_j), j = 1, \dots, n$. From Assumption 5,

$$n^{-1} \|K_n\theta_i\|^2 \rightarrow \|K\theta_i\|_{L^2(F)}^2,$$

and from the proof of Lemma 4.4 in [14],

$$\sum_{i=1}^m (c_i - c_i^n)^2 \rightarrow 0.$$

Thus, the right hand side of (4.7) converges to 0, and the result follows. □

THEOREM 4.1. *Suppose that Assumptions 1-5 hold, and let $f_0 \in W_\beta$, $\beta > 1$, with $Pf_0 \neq 0$. Let $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ in such a way that $n^{-1}\lambda_n^{-1/r} \rightarrow 0$ and*

$$\begin{aligned} d_n^2 \lambda_n^{-(s+1)} &\rightarrow 0, && \text{if } 1 < \beta < 3 - 1/r, \\ d_n^2 \lambda_n^{-(s+1/r+\beta-2)} &\rightarrow 0, && \text{if } \beta \geq 3 - 1/r. \end{aligned}$$

Define

$$\bar{\lambda} = \left(\bar{\sigma}^2 n^{-1} / \|Pf_0\|_W^2 \right)^{r/(r+1)}$$

and assume that $\bar{\lambda} \geq \lambda_n$. Then there exists a sequence $\lambda_M = \lambda_M(n)$ of minimisers of $EM(\lambda)$ over $\lambda \geq \lambda_n$ satisfying $\lambda_M \approx \bar{\lambda}$ as $n \rightarrow \infty$.

PROOF: First we show that for all n sufficiently large, $EM(\lambda)$ has a minimiser $\lambda_M = \lambda_M(n)$ over $[\lambda_n, \infty)$, and $\lambda_M \rightarrow 0$ as $n \rightarrow \infty$. Suppose that $\lambda \geq a$, for some $a > 0$. Then, from (2.17) and (2.18), we have as $n \rightarrow \infty$,

$$(4.8) \quad \bar{\mu}_1(\lambda) \approx n^{-1}\lambda^{-1}, \quad \bar{\mu}_2(\lambda) \approx n^{-1}\lambda^{-2}.$$

Also, from Lemma 4.5 in [14] and Lemma 4.2,

$$\begin{aligned} G_1(\lambda) &\geq (\lambda + \bar{q})^{-2} n^{-2} \sum_{\nu=1}^{n-m} \lambda_{\nu n} g_{\nu n}^2 \\ &\sim (\lambda + \bar{q})^{-2} \|PK^* \bar{P}g\|_W^2 \\ (4.9) \quad &\geq \lambda^{-2} a^2 (a + \bar{q})^{-2} \|PK^* \bar{P}g\|_W^2 \end{aligned}$$

and, from Lemma 4.4 in [14],

$$\begin{aligned} G(\lambda) &\leq \lambda^{-1} n^{-1} \sum_{\nu=1}^{n-m} g_{\nu n}^2 \\ (4.10) \quad &\sim \lambda^{-1} \|\bar{P}g\|_{L^2(F)}^2. \end{aligned}$$

Combining (4.3), (4.8), (4.9) and (4.10), we get, for some positive constants c_1, c_2 and c_3 ,

$$\begin{aligned} n^{-1} \lambda d(\lambda) \frac{d EM(\lambda)}{d\lambda} &\geq \lambda^{-2} \left[\lambda \left(a^2 (a + \bar{q})^{-2} \|PK^* \bar{P}g\|_W^2 - c_1 n^{-1} \|\bar{P}g\|_{L^2(F)}^2 \right) \right. \\ &\quad \left. - \bar{\sigma}^2 (c_2 n^{-1} - c_3 n^{-2}) \right]. \end{aligned}$$

Clearly, there exists N such that for all $n \geq N$ and $\lambda \geq a$, the right hand side is > 0 . Hence, for all $n \geq N$, $EM(\lambda)$ has a minimiser $\lambda_M = \lambda_M(n)$ over $[\lambda_n, \infty)$, and furthermore $\lambda_M \rightarrow 0$ as $n \rightarrow \infty$. Now, setting the right hand side of (4.3) equal to 0 at $\lambda = \lambda_M$, gives the equation

$$(4.11) \quad \lambda_M G_1(\lambda_M) - \lambda_M \bar{\mu}_1(\lambda_M) G(\lambda_M) = \bar{\sigma}^2 \bar{\mu}_2(\lambda_M) - \bar{\sigma}^2 \bar{\mu}_1^2(\lambda_M).$$

Since $\lambda_M \rightarrow 0$ as $n \rightarrow \infty$, from (2.17) and (2.18) we have $\bar{\mu}_1(\lambda_M) \rightarrow 0$, $\bar{\mu}_1^2(\lambda_M)/\bar{\mu}_2(\lambda_M) \rightarrow 0$ and $\bar{\mu}_2(\lambda_M) \approx n^{-1} \lambda_M^{-1/r}$. Using this and Lemma 4.1, equation (4.11) gives

$$\lambda_M \|Pf_0\|_W^2 \sim \bar{\sigma}^2 \bar{\mu}_2(\lambda_M) \approx \bar{\sigma}^2 n^{-1} \lambda_M^{-1/r},$$

which implies that $\lambda_M \approx \bar{\lambda}$. Since $\bar{\lambda} \geq \lambda_n$ by assumption, this completes the proof. \square

Now we consider the asymptotic inefficiency of the GML estimate. If $Pf_0 = 0$, then as shown in [22], $I_{M/R} = 1$ for all n . This follows because, if $Pf_0 = 0$, then $b^2(\lambda) = 0$, and so from (2.19) and (2.20), $ER(\lambda) = \sigma^2 \mu_2(\lambda)$, which clearly is minimised at $\lambda_R = \infty = \lambda_M$.

However, in the usual case where $Pf_0 \neq 0$, the following theorem shows that if f_0 is smooth relative to W , then λ_M is asymptotically sub-optimal and undersmoothing with respect to $ER(\lambda)$. This result was conjectured in [22].

THEOREM 4.2. *Suppose that Assumptions 1–5 hold and let $f_0 \in W_\beta$, $\beta > 1$, with $Pf_0 \neq 0$. Suppose that $\bar{\lambda}$ and λ_n are defined and related as in Theorem 4.1, and $\lambda^* \geq \lambda_n$, where λ^* is defined in (2.22). Let λ_R minimise $ER(\lambda)$ over $\lambda \geq \lambda_n$. If λ_M is the sequence of minimisers of $EM(\lambda)$ over $\lambda \geq \lambda_n$, defined by Theorem 4.1, then as $n \rightarrow \infty$, $\lambda_M/\lambda_R \rightarrow 0$ and*

$$\frac{ER(\lambda_M)}{ER(\lambda_R)} \gtrsim \begin{cases} (\sigma^2 n^{-1})^{\tau/(\tau+1) - \beta\tau/(\beta\tau+1)}, & 1 < \beta < 2, \\ (\sigma^2 n^{-1})^{\tau/(\tau+1) - 2\tau/(2\tau+1)}, & \beta \geq 2, \end{cases} \rightarrow \infty.$$

PROOF: As in the proof of Theorem 5.1 in [14], $\lambda_R \rightarrow 0$ as $n \rightarrow \infty$, and hence, from Theorems 4.4 and 4.5 in [14], $\lambda_R \gtrsim \lambda^*$. Therefore, from Theorem 4.1 and since $\beta > 1$, we have $\lambda_M/\lambda_R \lesssim \bar{\lambda}/\lambda^* \rightarrow 0$. In addition, (2.18) and Theorem 4.1 imply that

$$ER(\lambda_M) \geq \sigma^2 \mu_2(\lambda_M) \approx \sigma^2 n^{-1} \bar{\lambda}^{-1/r} \approx (\sigma^2 n^{-1})^{\tau/(\tau+1)}.$$

On the other hand, from Corollary 4.1 in [14],

$$ER(\lambda_R) \leq ER(\lambda^*) \lesssim \begin{cases} (\sigma^2 n^{-1})^{\beta\tau/(\beta\tau+1)}, & 1 < \beta < 2, \\ (\sigma^2 n^{-1})^{2\tau/(2\tau+1)}, & \beta \geq 2. \end{cases}$$

The result follows easily by dividing the bounds and checking the exponents. □

Under the same conditions, the GML estimate λ_M is also asymptotically sub-optimal and undersmoothing with respect to $EL(\lambda)$ or any of the stronger loss functions $EL_\rho(\lambda)$, as shown below.

THEOREM 4.3. *Suppose that Assumptions 1–5 hold, and let $f_0 \in W_\beta$, $\beta > 1$, with $Pf_0 \neq 0$. Suppose that $\bar{\lambda}$ and λ_n are defined and related as in Theorem 4.1, and $\lambda_\rho^* \geq \lambda_n$, where λ_ρ^* is defined in (2.16) and $0 \leq \rho < 2 - s - 1/\tau$. Let λ_ρ minimise $EL_\rho(\lambda)$ over $\lambda \geq \lambda_n$. If λ_M is the sequence of minimisers of $EM(\lambda)$ over $\lambda \geq \lambda_n$, defined by Theorem 4.1, then as $n \rightarrow \infty$, $\lambda_M/\lambda_\rho \rightarrow 0$ and*

$$\frac{EL_\rho(\lambda_M)}{EL_\rho(\lambda_\rho)} \gtrsim \begin{cases} (\sigma^2 n^{-1})^{\tau(1-\rho)/(\tau+1) - \tau(\beta-\rho)/(\beta\tau+1)}, & \rho < \beta < \rho + 2, \\ (\sigma^2 n^{-1})^{\tau(1-\rho)/(\tau+1) - 2\tau/(2\tau+\rho\tau+1)}, & \beta \geq \rho + 2, \end{cases}$$

$\rightarrow \infty$.

PROOF: Using Proposition 3.1 and Corollary 3.1 in [14], we have

$$\min\{1, \lambda_\rho^2\} \|f_0\|_\rho^2 \lesssim EL_\rho(\lambda_\rho) \lesssim EL_\rho(\lambda_\rho^*) \rightarrow 0,$$

and so $\lambda_\rho \rightarrow 0$ as $n \rightarrow \infty$. Hence, from Propositions 3.1 and 3.2 in [14], $\lambda_\rho \gtrsim \lambda_\rho^*$. Therefore, from Theorem 4.1 and since $\beta > 1$, we get $\lambda_M/\lambda_\rho \lesssim \bar{\lambda}/\lambda_\rho^* \rightarrow 0$. In addition, Proposition 3.2 in [14] and Theorem 4.1 imply that

$$EL_\rho(\lambda_M) \gtrsim \sigma^2 n^{-1} C(\lambda_M, \rho) \approx \sigma^2 n^{-1} \bar{\lambda}^{-(\rho+1/\tau)} \approx (\sigma^2 n^{-1})^{\tau(1-\rho)/(\tau+1)}.$$

Since $EL_\rho(\lambda_\rho) \leq EL_\rho(\lambda_\rho^*)$, the result follows from Corollary 3.1 in [14]. □

Consider the case of periodic numerical differentiation of order k using regularisation of order m . From the discussion before Corollary 5.2 in [14], the condition $f_0 \in W_\beta$, $\beta > 1$, in Theorems 4.2 and 4.3 is equivalent to $f_0 \in W(\theta)$, $\theta > m$, where

$$W(\theta) = \left\{ f \in L^2 : \sum_{\nu=-\infty}^{\infty} (2\pi\nu)^{2\theta} |\hat{f}_\nu|^2 < \infty \text{ and } \hat{f}_0 = 0 \right\}$$

$$= \left\{ f \in W^{\theta,2} : f^{(j)}(0) = f^{(j)}(1), j = 0, \dots, \theta - 1, \text{ and } \int_0^1 f = 0 \right\} \text{ if } \theta \in \mathbb{Z}^+.$$

Therefore in this case, these two theorems generalise a result in [6] which uses the risk criterion and the condition $f_0 \in W(2m + k) = W_2$.

The asymptotic properties of λ_M are quite different if f_0 is “rough” relative to W . In [22], Wahba shows that if $g_{\nu n}^2 = b\lambda_{\nu n}$, $\nu = 1, \dots, n - m$, then $\lambda_M = \lambda_R = \sigma^2/(nb)$ and so $I_{M/R} = 1$. This suggests that if as $n \rightarrow \infty$,

$$(4.12) \quad \sum_{\nu=1}^{n-m} g_{\nu n}^2/\lambda_{\nu n} \approx n,$$

then GML should perform quite well. Note that from (4.4), the condition (4.12) implies that $f_0 \notin W$.

If the smoothness of f_0 is known in terms of say $\bar{\theta} = \sup\{\theta : f_0 \in W^{\theta,2}\}$ (where $\bar{\theta}$ may be fractional), then it may be possible to choose the space W (or the order of regularisation) so that (4.12) is satisfied. This point is discussed further in [12]. However, if little or nothing is known about the smoothness of f_0 (which is the usual case), then (4.12) will be of no assistance.

REFERENCES

- [1] R.S. Anderssen and P. Bloomfield, 'Numerical differentiation procedures for non-exact data', *Numer. Math.* **22** (1974), 157–182.
- [2] D.D. Cox, 'Approximation of regularization estimators', *Ann. Statist.* **16** (1988), 694–712.
- [3] P. Craven and G. Wahba, 'Smoothing noisy data with spline functions', *Numer. Math.* **31** (1979), 377–403.
- [4] A.R. Davies, 'On the maximum likelihood regularization of Fredholm convolution equations of the first kind', in *Treatment of integral equations by numerical methods*, (C.T.H. Baker and G.F. Miller, Editors) (Academic Press, London, 1982), pp. 95–105.
- [5] A.R. Davies, M. Iqbal, K. Maleknejad and T.C. Redshaw, 'A comparison of statistical regularization and Fourier extrapolation methods for numerical deconvolution', in *Numerical treatment of inverse problems in differential and integral equations*, (P. Deuffhard and E. Hairer, Editors) (Birkhäuser, Berlin, 1983), pp. 320–334.
- [6] A.R. Davies and R.S. Anderssen, 'Improved estimates of statistical regularization parameters in Fourier differentiation and smoothing', *Numer. Math.* **48** (1986), 671–697.
- [7] A.R. Davies and R.S. Anderssen, 'Optimization in the regularization of ill-posed problems', *J. Austral. Math. Soc. Ser. B* **28** (1986), 114–133.
- [8] H.W. Engl and H. Gfrerer, 'A posteriori parameter choice for general regularization methods for solving linear ill-posed problems', *Appl. Numer. Math.* **4** (1988), 395–417.
- [9] G.S. Kimeldorf and G. Wahba, 'Some results on Tchebycheffian spline functions', *J. Math. Anal. Appl.* **33** (1971), 82–95.
- [10] M.A. Lukas, 'Regularization', in *The application and numerical solution of integral equations*, (R.S. Anderssen, F.R. de Hoog and M.A. Lukas, Editors) (Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980), pp. 151–182.
- [11] M.A. Lukas, 'Convergence rates for regularized solutions', *Math. Comp.* **51** (1988), 107–131.
- [12] M.A. Lukas, 'Assessing regularised solutions', *J. Austral. Math. Soc. Ser. B* **30** (1988), 24–42..
- [13] M.A. Lukas, 'Methods for choosing the regularization parameter', in *Inverse problems in partial differential equations*, (A.K. Pani and R.S. Anderssen, Editors), Proceedings of the Centre for Mathematics and its Applications (Australian National University, Canberra, ACT, 1992), pp. 89–110.

- [14] M.A. Lukas, 'Asymptotic optimality of generalized cross-validation for choosing the regularization parameter', *Numer. Math.* **66** (1993), 41–66.
- [15] M.A. Lukas, 'Convergence rates for moment collocation solutions of linear operator equations', *Numer. Funct. Anal. Optim.* (to appear).
- [16] V.A. Morozov, 'On the solution of functional equations by the method of regularization', *Soviet Math. Dokl.* **7** (1966), 414–417.
- [17] V.A. Morozov, *Methods for solving incorrectly posed problems* (Springer-Verlag, Berlin, Heidelberg, New York, 1984).
- [18] D.W. Nychka and D.D. Cox, 'Convergence rates for regularized solutions of integral equations from discrete noisy data', *Ann. Statist.* **17** (1989), 556–572.
- [19] G. Wahba, 'Convergence rates of certain approximate solutions to Fredholm integral equations of the first kind', *J. Approx. Theory* **7** (1973), 167–185.
- [20] G. Wahba, 'Smoothing noisy data by spline functions', *Numer. Math.* **24** (1975), 383–393.
- [21] G. Wahba, 'Practical approximate solutions to linear operator equations when the data are noisy', *SIAM J. Numer. Anal.* **14** (1977), 651–667.
- [22] G. Wahba, 'A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem', *Ann. Statist.* **13** (1985), 1378–1402.
- [23] G. Wahba, *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics **59** (SIAM, Philadelphia, 1990).

School of Mathematical and Physical Sciences
Murdoch University
Murdoch WA 6150
Australia