OPTIMALITY IN NUMERICAL DIFFERENTIATION

A R Davies

1. INTRODUCTION

Consider the problem of estimating the mth derivative of a data function \( g(x) \), given only \( N \) sampled values

\[
y_n = g(x_n) + \varepsilon_n, \quad n = 0, \ldots, N-1,
\]

where \( \varepsilon_n \) are uncorrelated random errors with mean zero and common variance \( \sigma^2 \) (possibly unknown). For simplicity consider equally-spaced sampling points \( x_n = n/N \) on the interval \([0,1)\). Let \( m \) be a strictly positive integer, and denote the mth derivative by \( f(x) = g^{(m)}(x) \), which is to be estimated on the interval \( 0 \leq x \leq 1 \).

If \( K \) denotes an integral operator such that \( Kf = g \), then a stabilized derivative can be constructed using pth-order Tikhonov regularization:

\[
\min_{f \in F_p} \left\{ \frac{1}{N} \sum_{n=0}^{N-1} \left[ (Kf)_{x_n} - y_n \right]^2 + \lambda \|f\|_p^2 \right\},
\]

where \( F_p \) is a suitably chosen Hilbert space with norm \( \| \cdot \|_p \) parametrized by the order of regularization \( p > 0 \), and the constant \( \lambda > 0 \) is the regularization parameter. Let \( f_{N;\alpha} \) denote the minimizer of (2), where \( \alpha \) is the parameter pair \( \hat{\alpha} = (p,\lambda) \).

In theory we may define an absolutely optimal parameter set \( \hat{\alpha} \) as that which minimizes (with respect to \( \hat{\alpha} \)) the error

\[
\| f_{N;\alpha} - f \|_F
\]

where \( \| \cdot \|_F \) denotes the strongest norm consistent with the smoothness of the exact derivative \( f \). In the data space there will exist a norm \( \| \cdot \|_G \) such that (3) and

\[
\| Kf_{N;\alpha} - g \|_G
\]
are equal; however, in the absolute sense the G-norm is necessarily stronger than the F-norm since $K$ is a smoothing operator. Thus, if we seek an $\alpha$ which is optimal with respect to the $L^2$-error in the space of the derivative, it is not sufficient to minimize the $L^2$-error in the data space. This is one reason why absolute optimality is difficult to achieve in practice. Another reason is that the smoothness of $f$ is usually unknown.

It is not our intention in this report to address the quest for an absolutely optimal parameter set $\alpha$; indeed, the author knows of no practical method which can achieve this. Instead we discuss the weaker concept of $D$-optimality and the even weaker concept of $S$-optimality, which we define below.

First consider the predictive mean-square signal error

$$S(\alpha) = \frac{1}{N} \sum_{n=0}^{N-1} \left[ (Kf_{N,\alpha})(x_n) - g(x_n) \right]^2.$$  
(5)

The minimizer of $S(\alpha)$ is estimated quite closely by several practical statistical methods, at least when $p$, the order of regularization, is fixed. For the present let $p$ be fixed. We emphasize this by writing $\lambda$ where previously we have used $\alpha$. Let $\lambda_0$ denote the minimizer of $S(\lambda)$ with respect to $\lambda$. Following [1] we say that a value of $\lambda$ is

(i) strongly $S$-optimal if $S(\lambda)/S(\lambda_0) = 1 + \mathcal{O}(N^{-a})$, $a > a_0 > 0$, as $N \to \infty$;  
(6a)

(ii) $S$-optimal if $S(\lambda)/S(\lambda_0) = 1 + \mathcal{O}(1)$ as $N \to \infty$, where $\mathcal{O}(1) \to 0$ as $N \to \infty$;  
(6b)

(iii) weakly $S$-optimal if $S(\lambda)/S(\lambda_0) = \mathcal{O}(1)$ as $N \to \infty$;  
(6c)

(iv) $S$-suboptimal if $S(\lambda)/S(\lambda_0) \to \infty$ as $N \to \infty$.  
(6d)

Since $f_{N,\lambda}$ is determined from the data $y_{N,\lambda}$, $S(\lambda)$ is a random variable. Its analysis as such is difficult, and it is easier to study $E S(\lambda)$ instead, where $E$ denotes expectation with respect to the error distribution. With this in mind let $\lambda_0^*$ denote the minimizer of $E S(\lambda)$. We maintain the definitions (6) if
$S(\lambda)$ is replaced by $ES(\lambda)$ and $\lambda_0$ by $\lambda_0$. 

The *predictive mean-square derivative error* is defined by

$$D(\alpha) = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \sum_{\alpha} (x_n, f(x_n)) - f(x_n) \right]^2.$$  \hspace{1cm} (7)

When $p$ is fixed we can define a D-optimal value of $\lambda$ in the same way as (6), simply replacing $S$ by $D$. Unfortunately, for the same value of $p$, a value of $\lambda$ which is S-optimal is not D-optimal, nor is a D-optimal value of $\lambda$ absolutely optimal unless the natural smoothness of $f$ is no greater than $L^2$.

Certain practical methods for determining $\lambda$ are known to yield values which vary from S-suboptimal to strongly S-optimal. (This depends on the method and the regularity of the data; see §2 below). For example, when the noise variance $\sigma^2$ is known, the unbiased risk method always yields strongly S-optimal values of $\lambda$; when $\sigma^2$ is unknown, the unbiased risk method cannot be used, but the method of cross-validation can yield S-optimal (sometimes strongly S-optimal) values of $\lambda$. It is therefore natural to pose the question: *can practical methods be devised for determining values of $\lambda$ which are D-optimal (preferably strongly D-optimal)?* This report presents some background information which may be of value in answering this question.

2. S-OPTIMALITY.

**Data regularity** Let $\nu = m + p$, and $\beta > 0$ be a constant such that

$$\sum_{q=0}^{\infty} 2^\nu |\hat{g}_q|^2 < \infty, \quad \omega_q = 2\pi q,$$ \hspace{1cm} (8)

where $\hat{g}_q$ are the Fourier coefficients of the underlying data function $g$, i.e.

$$\hat{g}_q = \hat{g}(\omega_q), \text{ where } \hat{g}(\omega) = \frac{1}{2\pi} \int_{0}^{1} g(x)e^{-i\omega x}dx.$$ \hspace{1cm} (9)

Clearly, $r = \beta \nu$ is a measure of the smoothness of $g$. If there is a natural
limit on this smoothness, i.e. \( r \leq r_0 \), say, then as \( p \) is increased the value of \( \beta \) in (8) is reduced so that \( \beta \nu \leq r_0 \). On the other hand, when \( p \) is fixed, \( \beta \) is a measure of the smoothness of \( g \).

As has been shown by Lukas [2], the asymptotic properties of \( \lambda_0^* \) and \( ES(\lambda_0^*) \) depend quite crucially on whether \( 0 < \beta < 2 \), or whether \( \beta > 2 \). Let us consider the case of Fourier differentiation where \( F_p \) in (2) is the space of trigonometric polynomials of degree at most \( \frac{1}{2} N \), and \( \| f \|_p = \| f (p) \|_{L^2} \). The associated operator \( K \) is defined in [1]. We state the following without proof:

**Theorem 1.** Under the regularity assumption (8), the value of \( \lambda \) which minimizes \( ES(\lambda) \) may be written

(i) when \( \beta > 2 
\hat{\lambda}_0^* = \left[ \frac{\kappa_{\nu} \cdot \sigma^2}{4\nu} \cdot \frac{1}{\| g \|_{\nu}^2} \right]^{2\nu/(4\nu+1)} [1 + o(1)] \text{ as } N \to \infty, \tag{10} \]

where \( \kappa_{\nu} \) is the constant

\[
\kappa_{\nu} = \frac{1}{\pi} \int_0^\infty \frac{dt}{(1 + e^{2\nu t})^2}, \tag{11} \]

\[
\| g \|_\nu = \frac{1}{\pi} \int_0^\infty e^{2\nu r} \left| \hat{g}(\omega) \right|^2 d\omega, \quad r > 0, \tag{12} \]

and \( o(1) \to 0 \) as \( N \to \infty \). The minimizing value of \( ES(\lambda) \) as \( N \to \infty \) is

\[
ES(\hat{\lambda}_0^*) = (4\nu + 1) \left[ \frac{\kappa_{\nu} \cdot \sigma^2}{4\nu} \cdot \frac{1}{\| g \|_{\nu}^{4\nu/(4\nu+1)}} \right]^{1/(4\nu+1)} [1 + o(1)]. \tag{13} \]

(ii) When \( 0 < \beta < 2 \), we have instead

\[
\hat{\lambda}_0^* > \left[ \frac{\kappa_{\nu} \cdot \sigma^2}{\beta \nu} \cdot \frac{1}{\| g \|_{\beta \nu}} \right]^{2\nu/(2\beta \nu+1)} [1 + o(1)] \text{ as } N \to \infty, \tag{14} \]
and
\[ ES(\lambda_0) \leq (2\beta \nu + 1) \left[ \frac{K_\nu}{2^\nu} \cdot \frac{\sigma^2}{N} \right] \| g \|_{\nu}^{1/(2\beta \nu + 1)} [1 + o(1)]. \] (15)

**Corollary.** If $\beta \geq 3$, the $o(1)$ terms in (10) and (13) become $0\left(\frac{-2\nu}{4\nu+1}\right)$.

(For a proof see [1].)

From (13) and (15) we see that $ES(\lambda_0)$ is a function of $p$. As $N \to \infty$ we can seek to optimize $p$ by maximizing the exponents of $(\sigma^2/N)$ in (13) and (15).

This demands that
\[ p > \frac{1}{2}(\beta \nu - 2m) \] (16)

and in particular that $p$ must be sufficiently large to bring $\beta$ into the range $0 < \beta < 2$. Of course this is an asymptotic result, and in practice it need not necessarily be consistent with the minimization of $S(\alpha)$ when $N$ is finite.

**Practical methods.** In practice $S(\alpha)$ cannot be computed directly without knowing the exact data function and so alternative means of estimating its minimizer (with respect to $\alpha$) is needed. Practical methods fall into two classes:

Class I. Those which require a knowledge of the noise variance $\sigma^2$.

Class II. Those which do not.

Most methods can be implemented in terms of an influence matrix $A(\alpha)$. If $g_{N;\alpha}$ denotes the $N$-vector whose elements are the values of the function $(Kf_{N;\alpha})(x)$ at $x = x_n$, then the influence matrix is such that
\[ g_{N;\alpha} = \hat{A}(\alpha)\gamma. \] (17)

For Fourier differentiation, $A(\alpha)$ is the circulant matrix whose eigenvalues constitute the discrete regularization filter imposed by the regularization process [1]. Class I methods invariably involve the mean-square error or discrepancy:
or modifications thereof. It may be shown that

\[
T(\alpha; y_N) = \frac{1}{N} \sum_{n=0}^{N-1} \left( (Kf_{N,\alpha})_n - y_n \right)^2 = \frac{1}{N} \| (I - A(\alpha)) y_N \|_2^2, \tag{18}
\]

or modifications thereof. It may be shown that

\[
\mathbb{E}T(\alpha; y_N) = \mathbb{E}S(\alpha) + \sigma^2 \cdot 0\left[ N^{1/(2\nu)} \right]^{-1} \tag{19}
\]

provided \( N^{1/(2\nu)} \rightarrow \infty \) as \( N \rightarrow \infty \). Notable among Class I methods are the

discrepancy principles of Arcangeli [3], Morozov [4], and their generalizations [5,6]; the Bayesian method of Turchin [7,8]; the unbiased risk method [9]; and the Bayesian method of Turchin-Klein [10]. Class II methods involve the

optimization of certain functions in lieu of \( S(\alpha) \). Among these are cross-validation [11] and maximum likelihood [12]. We summarize in the adjoining table the practical implementation and classification (with respect to \( S_\alpha \)-optimality) of the main methods in each class.

We see from the table that only unbiased risk minimization, cross-validation, and maximum-likelihood enable the optimization of \( p \) as well as \( \lambda \). Of the Class I methods tabulated, unbiased risk minimization is clearly the most powerful. This is because of the identity

\[
\mathbb{E}R(\alpha) = \mathbb{E}S(\alpha) \tag{20}
\]

which immediately yields strong \( S_\alpha \)-optimality in the expectational sense with \( \alpha = \infty \). Of the Class II methods tabulated, cross-validation is the most powerful.

3. D-OPTIMALITY.

Let \( p \) be fixed, and let \( \lambda^{(m)}_0 \) denote the minimizer of \( ED(\lambda) \). To obtain similar results for \( \lambda^{(m)}_0 \) and \( ED(\lambda^{(m)}_0) \) to those given for \( \lambda_0 \) and \( ES(\lambda_0) \) in Theorem 1, when the data regularity condition (8) is replaced by

\[
\sum_{q=0}^{\infty} \omega_{q+1}^2 (\beta q + m) \left| \frac{1}{\nu} \|^2 < \infty, \tag{21}
\]
<table>
<thead>
<tr>
<th>Method</th>
<th>Practical Implementation</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morozov discrepancy [4]</td>
<td>Fix p. Solve (T1) for λ: [ \frac{1}{N} | (I - A(\lambda)) y_N | ^2 = \sigma^2 ] (T1)</td>
<td>( \beta &lt; 2 )</td>
</tr>
<tr>
<td>Turchin [7,8]</td>
<td>Fix p. Solve (T2) for λ: [ \frac{\sigma^2}{N} \text{Tr}(\lambda) + \frac{1}{N} | (I - A(\lambda)) y_N | ^2 = \sigma^2 ] (T2)</td>
<td>( \beta &lt; 2 )</td>
</tr>
<tr>
<td>Unbiased risk minimization [9]</td>
<td>Minimize ( R(\alpha) ) w.r.t. ( \alpha = (p, \lambda) ): [ R(\alpha) = \frac{1}{N} | (I - A(\alpha)) y_N | ^2 + \frac{2\sigma^2}{N} \text{Tr}(\alpha) + \sigma^2 ]</td>
<td>( \beta \geq 2 )</td>
</tr>
<tr>
<td>Cross-validation [10]</td>
<td>Minimize (T3) w.r.t. ( \alpha ): [ \frac{1}{N} | (I - A(\alpha)) y_N | ^2 ] (T3)</td>
<td>( \beta \geq 2 )</td>
</tr>
<tr>
<td>Maximum likelihood [12]</td>
<td>Minimize (T4) w.r.t. ( \alpha ): [ \frac{y^T (I - A(\alpha)) y}{[\text{det}^+(I - A(\alpha))]^{1/(N-1)}} ] (T4)</td>
<td>( \beta \geq 3 )</td>
</tr>
</tbody>
</table>

(\( \text{det}^+ \) denotes omission of zero eigenvalues).
it can be shown that
\[
\lambda_0^{(m)} = O\left(\frac{\sigma^2}{N} \frac{2v/(4v+2m+1)}{N} \right) \quad \beta > 2,
\]
and
\[
ED\left(\lambda_0^{(m)} \right) = O\left(\frac{\sigma^2}{N} \frac{2v/(4v+2m+1)}{N} \right), \quad 0 < \beta < 2.
\]
From these rates we may deduce that an S-optimal value of \( \lambda \) is D-suboptimal.

To achieve weak D-optimality is not difficult in principle. For example if \( \sigma^2 \) is known then for a fixed \( p \) satisfying (21) we can choose
\[
\lambda = \left(\frac{\sigma^2}{N} \right)^\mu
\]
and
\[
\mu = \begin{cases} 
\frac{2v}{(4v+2m+1)}, & \beta > 2 \\
\frac{2v}{(2\beta v+2m+1)}, & 0 < \beta < 2.
\end{cases}
\]
This guarantees the correct convergence rates in (22)-(23), although the choice of constant of proportionality in (24) will greatly affect the quality of the solution.

Weak D-optimality can sometimes be achieved by cross-validation also. Let \( p, p' \) and \( \beta > 2 \) be such that
\[
p' = (2m+1)p + 2m^2
\]
and (21) is satisfied with \( v \) replaced by \( v' = m+p' \). Clearly (8) is also satisfied. Using cross-validation with order of regularization \( p \) determines an S-optimal value of \( \lambda_{CV} \) with the associated expectational property
\[
\lambda_{CV} = O\left(\frac{\sigma^2}{N} \frac{2v/(4v+1)}{N} \right) \text{ as } N \to \infty.
\]
The choice of $p'$ in (25), however, is such that
\[
\frac{2v}{4v+1} = \frac{2v'}{4v'+2m+1}
\]
and so $\lambda_{CV}$ is also weakly D-optimal for the order of regularization $p'$ higher than $p$.

To achieve stronger levels of D-optimality for a given $p$ appears to be a greater challenge in practice. This challenge arises from the fundamental nature of ill-posed inverse problems and is illustrated by the following observations.

In the degenerate case $m = 0$, which is the case of data smoothing, there is clearly no difference between S- and D-optimality. Thus cross-validation applied to a smoothing problem achieves exactly the right level of optimality in the present context. (We are not discussing absolute optimality, which is not achieved.) What then if the problem of differentiating inexact data is first converted to a smoothing problem, and then an S-optimal method for choosing a regularization parameter for the smoothing problem is used? Does this do better than S-optimality for the original problem?

Consider first the direct approach of $p$th-order regularization applied to $m$th-order Fourier differentiation, i.e.

\[
\min_{f \in T_N} \left\{ \frac{1}{N} \sum_{n=0}^{\infty} [(Kf) (x_n) - y_n]^2 + \lambda \| f^{(p)} \|_{L^2}^2 \right\}
\]

where $T_N$ denotes the space of trigonometric polynomials on $[0,1]$ of degree at most $\frac{1}{2}N$. The minimizer $f_{N;\lambda}$ of (26) has discrete Fourier coefficients

\[
\frac{(i\omega)^m \gamma_N}{1 + \lambda \omega^{2v/q}} \quad \text{for} \quad \frac{1}{2} N \leq q \leq \frac{1}{2} N, \quad v = m + p,
\]
where \( \hat{y}_{N,q} \) denotes the discrete Fourier transform (DFT) of the data \( \{y_n\} \).

The factor \( (i\omega_q^m) \) in (27) appears as the result of \( m \)th-order differentiation whereas the factor \( (1 + \lambda \omega_q^{2^q-1}) \) is the regularization filter acting on the \( q \)th coefficient.

Now consider the a-posteriori smoothing approach where we first compute inexact derivative data \( \{d_n\} \) which we subsequently smooth using \( p \)th-order regularization. The DFT of the derivative data will always have the form

\[
\hat{d}_{N,q} = (i\omega_q^m) \tau_{N,q} \hat{y}_{N,q} \tag{28}
\]

where \( \tau_{N,q} \) is an attenuation factor determined by whatever method we use to generate the derivative data \( \{d_n\} \). For example if \( \{d_n\} \) is obtained from \( \{y_n\} \) by \( m \)th-order central differencing, we have

\[
\tau_{N,q} = \left[ \frac{\sin \left( \omega_q^N \right)}{N} \right]^m.
\]

Alternatively if \( \{d_n\} \) is obtained by direct regularization then \( \tau_{N,q} \) is the associated regularization filter (cf. (27)).

The variance-covariance matrix of the Fourier data \( \{\hat{d}_{N,q}\} \) is

\[
\frac{\sigma^2}{N} \hat{\Sigma},
\]

where \( \hat{\Sigma} = \text{diag}\left( \frac{\omega_q^{2m}}{\tau_{N,q}} \right) \). The associated matrix \( V \) for the data \( \{d_n\} \) is given by

\[
V = \psi \hat{V} \psi^* \]

where \( \psi \) is the DFT matrix. Both \( \hat{V} \) and \( V \) have rank \( < N \). In terms of the generalized inverse \( V^+ \) of \( V \), the a posteriori smoothing problem may be stated:

\[
\min_{\phi \in T_N} \left\{ \frac{1}{N} (\hat{\phi} - \hat{d}_N)^T V^+ (\hat{\phi} - \hat{d}_N) + \lambda \| \phi \|_{L^2}^2 \right\}
\]

(29)
where \( \hat{\phi} \) denotes the N-vector sampling \( \phi \) at \( x_n \). It is easily shown that the minimizer \( \phi_{N;\lambda} \) of (29) has discrete Fourier coefficients

\[
\frac{(i\omega)^m}{q^m} \tau_{N,q} \hat{\phi}_{N,q} \quad \text{for } \frac{1}{2} N \leq q \leq \frac{1}{2} N. 
\]  

(30)

For finite \( N \) there will be a difference, therefore, in the derivatives \( f_{N;\lambda} \) obtained through (27) by the direct approach and \( \phi_{N;\lambda} \) obtained through (30) by the a posteriori smoothing approach. As \( N \to \infty \), however, it transpires that \( \phi_{N;\lambda} \to f_{N;\lambda} \) since \( \tau_{N,q} \to 1 \) for all \( q \). What then if we choose an \( S \)-optimal value of \( \lambda \) for the smoothing problem (29)? It is not difficult to prove the following:

**Theorem 2.** Let \( 0 < \tau_{N} \leq |\tau_{N,q}| < 1 < \infty \) for all \( q \). Then the value of \( \lambda \) which minimizes

\[
E \left\{ \frac{1}{N} \sum_{n=0}^{N-1} \left[ \hat{\phi}_{N;\lambda}(x_n) - f(x_n) \right]^2 \right\}
\]

under the regularity condition (8), may be written

\[
\gamma_0(\text{aps}) = \theta_0 \gamma_0 \quad \text{as } N \to \infty,
\]

where \( \theta_0 \) satisfies

\[
\tau_N^{4\nu/(4\nu+1)} \leq \theta_0 \leq \tau_N^{-2(6\nu+1)/(4\nu+1)} \quad \text{when } \beta \geq 2,
\]

\[
\tau_N^{8\nu/2\beta
\nu+1} \leq \theta_0 \leq \tau_N^{-2(6\nu+1)}, \quad \text{when } 0 < \beta < 2.
\]

**Corollary.** \( \gamma_0(\text{aps}) \to \gamma_0 \) as \( N \to \infty \), whichever way the derivative data \( \{d_n\} \) are derived. This follows immediately from the fact that \( \tau_N \to 1 \) as \( N \to \infty \), and so \( \theta_0 \to 1 \).

Thus, in the limit \( N \to \infty \), \( S \)-optimal a posteriori smoothing (whatever method
is used to provide derivative data) is equivalent to S-optimal regularization of the original differentiation problem. The reason for the lack of improvement is clear. The variance-covariance matrix of the differentiated data reflects the instability of differentiation, and the weighting thus introduced into the a posteriori smoothing problem makes it equivalent to the differentiation problem as $N \to \infty$. Of course there is a difference when $N$ is finite, and improvements in the quality of the derivative may be possible in practice (cf., for example, [13]) using a posteriori smoothing.

REFERENCES


