TEKNIKUM

UPPSALA UNIVERSITY
SCHOOL OF ENGINEERING
Department of Technology

UPTEC 8792R

Date of issue
SEPTEMBER 1987

Project name

Sponsoring organization

Author(s)

Bengt Carlsson, Torsten Söderström and Anders Ahlén

Title (Swedish)

Title (English)

Digital Differentiating Filters

Abstract

The design and analysis of digital differentiating filters are discussed. Some general characterizations of differentiating filters are given. A number of methods are surveyed. Both frequency and time domain methods are handled. Several approaches are compared and further analysed in a simulation study.

Keywords / indexing

Security

Language
English

Supplementary bibliographical information

Pages
82

Classification

ISSN 0346-8887

ISBN

Postal address
Box 534
S-751 21 Uppsala
Sweden

Visiting address
Villavägen 4
Uppsala

Phone
+ 46-18-18 30 08 (librarian)
18 30 03 (office)

Telex
76143
UPTEC-S

DIGITAL DIFFERENTIATING FILTERS

A UNIFIED DESCRIPTION AND COMPARATIVE EVALUATION

Bengt Carlsson, Torsten Söderström and Anders Ahlén

TABLE OF CONTENTS

1. INTRODUCTION

The estimation of the time derivative of a noisy signal is a problem that occurs in many fields. A typical example is found when a velocity is to be estimated from position measurements. Another common example is the problem to estimate the net flow in a tank from measurements of the level.

In this report we will discuss the design and analysis of digital differentiating filters. Assume that a continuous-time differentiable signal is sampled. The problem under study is to estimate the derivative of the signal. The design of a discrete-time differentiating filter is non-trivial for at least two reasons; the relation between the discrete-time measurements and the continuous-time derivative and the sensitivity for noise in the signal. The problem has been pursued by a number of approaches. Naturally, standard methods for design of digital filters can be applied. The design is then often made in the frequency domain. Other possibilities also exist, for example to estimate a parametric time-domain model from measurement data. The derivative can then be found by differentiating the model. Several methods will be described in this paper.

The paper is organised as follows. The next section contains a statement of the problem and some general observations on differentiating filters. In Section 3, some ways to characterize the properties of a differentiating filter is discussed. A description of a number of different approaches for designing differentiating filters are given in Section 4. In the following section some approaches are numerically illustrated. Comparisons and analysis are also given. In the final section some conclusions are drawn.

2. PROBLEM SET-UP AND PRELIMINARIES

The problem we will deal with is analysis and design of linear digital differentiating filters. The input data to the filter is assumed to be contamined by noise. We describe the problem in the following way, cf also Figure 2.1.

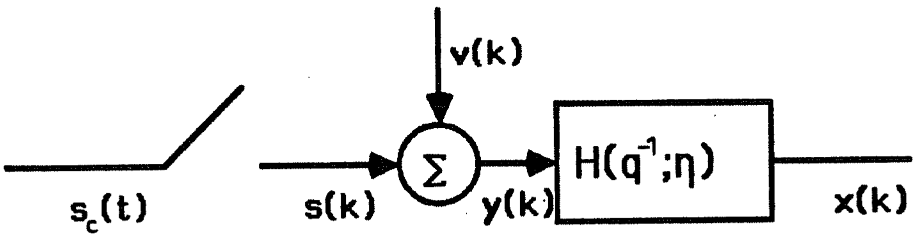


Figure 2.1. The problem set-up.

The continuous-time signal $s_c(t)$ is sampled which gives the discrete-time signal $s(k)$, where k is a running index. The sampling interval is T. The sampled signal is measured with noise so that

$$y(k) = s(k)+v(k) = s_c(kT)+v(k) \tag{2.1}$$

is the signal available for filtering. The disturbance $v(k)$ is assumed to be independent of the signal $s(k)$. We will also generally assume in what follows that the disturbance $v(k)$ is a stationary stochastic process. The filter is given by

$$x(k) = H(q^{-1};\eta)y(k) \tag{2.2}$$

where q^{-1} is the backward shift operator, so that

$$q^{-1}y(k) = y(k-1) \tag{2.3}$$

and η is a parameter vector. Different design methods lead to different parametrizations of the filter, i.e. different ways to let $H(q^{-1};\eta)$ depend on the vector η. We can write the filter as

$$H(q^{-1};\eta) = \sum_{j=-N}^{M} h_j(\eta)q^{-j} \tag{2.4}$$

In (2.4) we have allowed the filter $H(q^{-1};\eta)$ to be noncausal if $N>0$. This can give a beneficial influence on the filter performance. In a strict sense the linear operator $H(q^{-1};\eta)$ should be used under different names depending on the sign of N, namely:

- a smoother if $N>0$
- a filter if $N=0$
- a predictor if $N<0$

It will, however, be convenient, to often call $H(q^{-1};\eta)$ a filter regardless the sign of N. It is very common to classify the filter in (2.4) depending on the length of the impulse response. The filter in (2.4) is called

- an IIR-filter (infinite impulse response filter) if $M=\infty$

- a FIR-filter (finite impulse response filter) if $M<\infty$

The purpose of the filtering is that the filter output should be close to the time derivative of the signal in the sampling points

$$x(k) \approx \left.\frac{ds_c(t)}{dt}\right|_{t=kT} \triangleq \dot{s}_c(kT) \tag{2.5}$$

where T is the sampling interval.

The design of a differentiating filter $H(q^{-1};\eta)$ is non-trivial for at least two reasons, which we will discuss below:

- The relation between the discrete-time values $s(k)$ and the continuous-time derivative at the sampling points is not known in general. There are many ways a continuous-time signal $s_c(t)$ can vary between the discrete-time values $s(k)=s_c(kT)$.

- The noise corruption.

The output of the filter can be written as

$$x(k) = H(q^{-1};\eta)y(k) = H(q^{-1};\eta)s(k)+H(q^{-1};\eta)v(k) \tag{2.6}$$

Hence we can decompose the total error as follows

$$\dot{s}_c(kT)-x(k) = [\dot{s}_c(kT)-H(q^{-1};\eta)s(k)]-[H(q^{-1};\eta)v(k)] \tag{2.7}$$

The first part is sometimes referred to as the <u>systematic error</u>. It is the error that occurs in the noise-free case. The second part of (2.7), the <u>stochastic error</u>, describes the effect of the disturbances. Since the signal and the disturbances are assumed to be independent the two types of errors described above will be independent. It hence follows that the total mean square error variance (assuming zero means) is given by

$$V = E[\dot{s}_c(kT)-x(k)]^2$$
$$= E[\dot{s}_c(kT)-H(q^{-1};\eta)s(k)]^2+E[H(q^{-1};\eta)v(k)]^2 \tag{2.8}$$

The first of the two difficulties mentioned above concerns the relation between the time derivative $\dot{s}_c(kT)$ and the discrete-time values $\{s(k)\}$. There are essentially two ways to cope with this issue, namely

- To use a <u>frequency domain approach</u> to construct an ideal differentiating filter.

- To use a <u>parametric model</u> of the signal, from which the derivative can be found.

<u>A frequency domain approach</u>

Assume that the (continuous-time) signal $s_c(t)$ has a spectral density that vanishes completely above the Nyquist frequency $\omega_n=\pi/T$. According to the sampling theorem we can then reconstruct $s_c(t)$ exactly from the discrete-time values $\{s(k)\}$. This is sometimes called a <u>Shannon reconstruction</u> and can be found in, for example, Oppenheim and Schafer (1975), Åström and Wittenmark (1984). The reconstruction can be described as

$$s_c(t) = \sum_{j=-\infty}^{\infty} \phi_j(t)s(j) \tag{2.9}$$

$$\phi_j(t) = \begin{cases} 1 & t=jT \\ \dfrac{\sin\frac{\pi(t-jT)}{T}}{\frac{\pi(t-jT)}{T}} & t\neq jT \end{cases} \tag{2.10}$$

By straightforward differentiation of (2.9) we find that the derivative will satisfy

$$\dot{s}_c(t) = \sum_{j=-\infty}^{\infty} \dot{\phi}_j(t)s(j) \tag{2.11}$$

In the sampling points we get

$$\dot{s}_c(kT) = \sum_{j=-\infty}^{\infty} \dot{\phi}_j(kT)s(j) \tag{2.12}$$

Differentiation of the expression (2.10) for the weights $\{\phi_j(t)\}$ gives

$$\dot{\phi}_j(kT) = \begin{cases} \dfrac{(-1)^{k-j}}{T(k-j)} & k \neq j \\ 0 & k = j \end{cases} \tag{2.13}$$

Inserting this result into (2.12) gives after some calculations

$$\dot{s}_c(kT) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{s(k+n)-s(k-n)}{nT} \tag{2.14}$$

The filter can hence in this case be written as

$$\boxed{H(q^{-1}) = \sum_{j=1}^{\infty} (-1)^{j+1} \frac{q^{j}-q^{-j}}{jT}} \tag{2.15}$$

which easily gives the filter coefficients (cf (2.4))

$$h_j = \begin{cases} 0 & j = 0 \\ \dfrac{(-1)^j}{jT} & j \neq 0 \end{cases} \tag{2.16}$$

The frequency response of the filter becomes

$$H(e^{i\omega T}) = 2i \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{jT} \sin j\omega T \tag{2.17}$$

In the noise-free case the filter gives the exact derivative, i.e. the systematic error is zero. As a way to illustrate this we consider an "ideal" differentiator described by the frequency response

$$\bar{H}(e^{i\omega T}) = i\omega \quad |\omega|<\pi/T \tag{2.18}$$

Since $\bar{H}$ is a purely imaginary and odd function it can be expanded in a Fourier series as

$$\bar{H}(e^{i\omega T}) = \sum_{k=1}^{\infty} ia_k \sin k\omega T \tag{2.19}$$

with

$$a_k = \frac{2T}{\pi} \int_0^{\pi/T} \omega \sin k\omega T d\omega = \frac{2}{T} \frac{(-1)^{k+1}}{k} \tag{2.20}$$

Comparing with (2.17) we see that $\bar{H}(e^{i\omega T})=H(e^{i\omega T})$.

Let us illustrate the procedure (2.14) for differentiation of a sinusoidal signal.

Example 2.1. Let $s_c(t)$ be the sinusoid

$$s_c(t) = \sin\omega_o t \tag{2.21a}$$

The sampled signal is

$$s(k) = \sin(\omega_o Tk) \tag{2.21b}$$

where we must assume $|\omega_o T|<\pi$ in order to avoid aliasing effects, cf the sampling theorem. The right hand side of (2.14) becomes

$$x(k) = \sum_{n=1}^{\infty} \frac{1}{nT}(-1)^{n+1}[\sin\omega_o(k+n)T-\sin\omega_o(k-n)T]$$

$$= 2\cos\omega_o kT \sum_{n=1}^{\infty} \frac{1}{nT}(-1)^{n+1}\sin\omega_o nT \tag{2.21c}$$

To proceed we need the following result given in Gradshteyn and Ryzhik (1980):

$$\sum_{n=1}^{\infty} \frac{1}{n}\sin nx = \frac{\pi-x}{2} \quad 0<x<2\pi \tag{2.21d}$$

Now set $x=\pi+\omega_o T$. Then we have

$$x(k) = 2\cos\omega_o kT \sum_{n=1}^{\infty} \frac{1}{nT}(-1)^{n+1}\sin(nx-n\pi)$$

$$= 2\cos\omega_o kT \sum_{n=1}^{\infty} \frac{1}{nT}(-1)^{n+1}(-1)^n \sin nx$$

$$= -2\cos\omega_o kT \sum_{n=1}^{\infty} \frac{1}{nT}\sin nx = -2\cos\omega_o kT \, \frac{1}{T}\,\frac{\pi - x}{2}$$

$$= -\frac{1}{T}\cos\omega_o kT[-\omega_o T] = \dot{s}_c(kT)$$

which is the desired result. ∎

Let us stress again that the filter given by (2.14) is ideal in the sense that its frequency function is given by (2.18) and that there will be no systematic error. However, this filter design is not very practical to use for a number of reasons:

- The effect of noise is not accounted for.
- The algorithm is not suitable for efficient computation. For every (discrete) point k all data points have to be explicitly used in order to compute the derivative. This would require a very large number of operations per time step. In contrast a finite order linear filtering would require only a modest number of operations per time step.

Note that if the signal is not band-limited the correct derivative will not be found. There is no guarantee that the deviation between true and estimated derivative is small.

In (2.18) we introduced an "ideal" differentiator. In practice when also noise effects must be taken into account a "good" differentiator should satisfy (2.18) only for certain frequency regions. Thus, if $H(e^{i\omega T};\eta) \approx i\omega$ for

$0<\omega<\pi/T$ $H(q^{-1};\eta)$ is called a full-band differentiator

$0<\omega<0.7\pi/T$, say, $H(q^{-1};\eta)$ is called a wide-band differentiator

small ω $H(q^{-1};\eta)$ is called a narrow-band differentiator

Parametric models

This is the other approach for finding the derivative from discrete-time data. We then describe the signal with the help of a number of parameters, which we may collect in a vector θ. Note that to uniquely determine a continuous-time model from discrete-time data, the sampling theorem must not necessarily be fulfilled. A necesssary condition is, however, that the imaginary parts of the poles of the underlying system are less than π/T, cf Åström and Wittenmark (1984).

As simple examples of parametric models consider

$$\begin{aligned} s_1(t) &= A\sin(\omega_o t+\varphi) \qquad \theta = (A\ \omega_o\ \varphi)^T \\ s_2(t) &= at+b \qquad\qquad\quad \theta = (a\ b)^T \end{aligned} \tag{2.22}$$

From a limited number of the (discrete-time) data, we can in the noise free case determine the parameter vector. Once the vector θ is known, we know the signal completely and it is an easy task to compute the derivative for any desired time argument. When the signal is corrupted by noise, the parameters might be found by estimation. Many of the methods described later in this report can be viewed as based on parametric models. Both deterministic and stochastic signal models will be considered.

Consideration of the measurement noise

So far we have discussed only the relation between the signal and its derivative. Phrased differently, our interest has been focused to the systematic error, or the first term in the right hand side of (2.7). In practice we must of course take the measurement noise into consideration. To achieve a small mean square error variance, cf (2.8), the filter must be a compromise between the two objectives

- $H(q^{-1};\eta)$ should give approximately the true derivative in the noise-free case (the systematic error should be small), i.e.

$$H(q^{-1};\eta)s(k) \approx \dot{s}_c(kT) \tag{2.23}$$

- The noise should not be amplified too much. This means that

$$E[H(q^{-1};\eta)v(k)]^2 \approx 0 \tag{2.24}$$

Let us now describe how the trade-off between the objectives (2.23) and (2.24) can be met in an optimal way. By using the notations

$$\Phi_{vx}(\omega) = \sum_{\tau=-\infty}^{\infty} r_{vx}(\tau) e^{-i\tau\omega T} \tag{2.25}$$

$$r_{vx}(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} v(k+\tau)x(k)$$

we can write the mean square error variance as (for convenience we set $\dot{s}(k) \overset{\Delta}{=} \dot{s}_c(kT)$)

$$V = E[x(k)-\dot{s}(k)]^2$$

$$= \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} [\Phi_z(\omega)-\Phi_{z\dot{s}}(\omega)-\Phi_{\dot{s}z}(\omega)+\Phi_{\dot{s}}(\omega)]$$

$$= \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} [|H(e^{i\omega T};\eta)|^2\Phi_y(\omega)-H(e^{i\omega T};\eta)\Phi_{y\dot{s}}(\omega)$$

$$-\Phi_{\dot{s}y}(\omega)H(e^{i\omega T};\eta)+\Phi_{\dot{s}}(\omega)]d\omega \tag{2.26}$$

We can then minimize the integrand with respect to H for every value of ω in order to make V as small as possible. Noting that $\Phi_{y\dot{s}}(\omega)=\bar{\Phi}_{\dot{s}y}(-\omega)$, we can rewrite V as follows

$$V = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} [\Phi_y(\omega)\{H(e^{i\omega T};\eta)- \frac{\Phi_{\dot{s}y}(\omega)}{\Phi_y(\omega)}\}^2$$

$$+\{\Phi_{\dot{s}}(\omega)- \frac{|\Phi_{\dot{s}y}(\omega)|^2}{\Phi_y(\omega)}\}]d\omega \tag{2.27}$$

The first term in 2.27 is quadratic and the second term is independent of $H(e^{i\omega T};\eta)$. Hence the optimal filter can be written as

$$H(e^{i\omega T};\eta) = \frac{\Phi_{\dot{s}y}(\omega)}{\Phi_y(\omega)} = \frac{\Phi_{\dot{s}s}(\omega)}{\Phi_s(\omega)+\Phi_v(\omega)} \tag{2.28}$$

This is the so-called <u>unrealisable Wiener filter</u>.

Assume, in particular, that the signal is band-limited so that its continuous-time spectral density vanishes above the Nyquist frequency $\omega_N=\pi/T$. Then no aliasing effect occurs at the sampling and the continuous-time and discrete-time spectral densities of the signal coincide.

Furthermore, we have for this case

$$\phi_{\dot{s}s}(\omega) = i\omega\phi_s(\omega) \tag{2.29}$$

and hence

$$\boxed{H(e^{i\omega T};\eta) = i\omega \frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_v(\omega)}} \tag{2.30}$$

This result has a nice interpretation. The "ideal" differentiator is $i\omega$. The other factor $\phi_s/(\phi_s+\phi_v)$ describes the optimal trade-off between ideal differentiation and suppression of the noise.

There are many ways to derive a finite-order filter from (2.30). A number of methods are reviewed in Section 4.2.

3. CHARACTERIZATION OF DIFFERENTIATING FILTERS

In this section we will describe some ways to characterize the properties of a differentiating filter $H(q^{-1};\eta)$. We will organize the characterization around the following issues

- Algorithmic properties
- Deterministic behaviour
- Stochastic behaviour

3.1 <u>Algorithmic properties</u>

<u>Causality</u>

The first property concerns the characterization of the differentiator

$$H(q^{-1};\eta) = \sum_{j=-N}^{\infty} h_j(\eta)q^{-j} \tag{3.1}$$

as a smoother $(N>0)$, a filter $(N=0)$, or a predictor $(N<0)$. The differentiator is causal if and only if $N\leq 0$. When $N>0$, the estimated derivative is based on more data, compared to $N\leq 0$, and hence a better performance is potentially possible. In particular, ideal phase properties are possible to achieve with a smoother, see (3.7) below. The price to pay is that a smoother introduces a time-lag in the computations.

<u>Complexity</u>

This property concerns the form of the filter and the amount of computations needed when applying it. When used as a quantitative measure it would be appropriate to express the complexity in the number of floating point operations needed to implement the filter. This is an important implementation issue. However, we will not deal with this type of problems in the report.

<u>Parametrization</u>

This property concerns how the filter $H(q^{-1};\eta)$ depends on the parameter vector η. This dependence can be very direct, while in other cases extensive computations are needed to form the filter once the parameter vector has been chosen. A particular issue is the <u>possibility for adaptation</u>. If the differentiator is to be used in adaptive filtering, it must be possible to tune the parameter vector on-line when processing the data.

Higher order derivatives

The last property to mention in this section concerns extensions to higher order derivatives. In many cases it is easy to extend the filters so that a higher order derivative is computed. This is mostly a better way than to use a filter for the first order derivative repeatedly. Note that the frequency response for an ideal r:th order differentiator is

$$H(e^{-i\omega T};\eta) = (i\omega)^r \qquad |\omega|<\pi/T$$

3.2 Deterministic behaviour

In this section we focus our interest on the filter behaviour in the noise-free case.

Relation between $s(k)$ and $\dot{s}_c(kT)$

In most approaches there is an explicit or implicit modelling on how the discrete-time signal $s(k)$ and its continuous-time derivative at the sampling points $\dot{s}_c(kT)$ are related. For example, if $s(k)$ is assumed to be a polynomial in k, then the derivative $\dot{s}_c(kT)$ is easily found.

Frequency response

The frequency response of the filter is of much interest. It is given by

$$H(e^{i\omega T};\eta) = \sum_{j=-N}^{\infty} h_j(\eta)e^{i\omega Tj} \tag{3.2}$$

The frequency response describes how the various frequency components of a signal is affected by the filter. For example, if the signal is a sampled sinusoid

$$s(k) = A\sin(\omega_o Tk+\varphi) \qquad |\omega_o T|<\pi \tag{3.3a}$$

the filter will produce

$$x(k) = \sum_{j=-N}^{\infty} h_j s(k-j)$$

$$= A|H(e^{i\omega_o T})|\sin(\omega_o kT+\varphi-\arg H(e^{i\omega_o T})) \qquad (3.3b)$$

The frequency response $H(e^{i\omega_o T})$ gives the amplification and phase shift of a sinusoid with angular frequency ω_o. A good differentiating filter should have a frequency response

$$H(e^{i\omega T};\eta) = i\omega \qquad (3.4)$$

for the frequency band of interest. This will, in general, mean (at least) "small" values of ω. Next we give an illustration that it is natural and important to require (3.4) to hold.

<u>Proposition 3.1</u>. Assume that a filter gives exact differentiation for all signals being polynomials in t of degree n. Then

$$H(e^{i\omega T};\eta) = i\omega+O(|\omega|^{n+1}) \qquad (3.5)$$

<u>Proof</u>. Let the continuous-time signal be

$$s_c(t) = \sum_{m=0}^{n} \alpha_m^c t^m \qquad (3.6a)$$

The derivative in the sampling points is

$$\dot{s}_c(kT) = \sum_{m=0}^{n} \alpha_m^c T^{m-1} mk^{m-1} \qquad (3.6b)$$

Denote the sampled signal (cf Figure 2.1)

$$s(k) = \sum_{m=0}^{n} \alpha_m k^m \qquad (3.6c)$$

where $\alpha_m = \alpha_m^c T^m$, $m=0,1,\ldots n$.

Using the discrete-time signal, the derivative in (3.6b) can be written as

$$\dot{s}_c(kT) = \frac{1}{T}\sum_{m=0}^{n} \alpha_m mk^{m-1} \qquad (3.6d)$$

We thus require that

$$\frac{1}{T}\sum_{m=0}^{n} \alpha_m mk^{m-1} \equiv \sum_{j=-N}^{\infty} h_j \sum_{m=0}^{n} \alpha_m (k-j)^m \qquad (3.6e)$$

By equating the coefficients for α_m $(m=0,\dots,n)$, we get

$$mk^{m-1} \equiv T \sum_{j=-N}^{\infty} h_j (k-j)^m \qquad m=0,\dots,n$$

$$mk^{m-1} \equiv T \sum_{j=-N}^{\infty} h_j \sum_{i=0}^{m} \binom{m}{i} k^i (-j)^{m-i} \qquad m=0,\dots,n$$

$$mk^{m-1} \equiv T \sum_{i=0}^{m} \binom{m}{i} k^i (-1)^{m-i} \sum_{j=-N}^{\infty} j^{m-i} h_j \qquad m=0,\dots,n$$

This can be written as

$$\sum_{j=-N}^{\infty} j^k h_j = \begin{cases} -1/T & k=1 \\ 0 & k=0,2,\dots n \end{cases} \tag{3.6f}$$

We now use this result to evaluate the filter for small ω. We have by a Taylor series expansion

$$H(e^{i\omega T}) = \sum_{k=0}^{n} \frac{1}{k!} D^k H(e^{i\omega T}) \Big|_{\omega=0} \omega^k + O(\omega^{n+1})$$

$$= \sum_{k=0}^{n} \frac{1}{k!} D^k \Big[\sum_{j=-N}^{\infty} h_j e^{-i\omega T j} \Big] \Big|_{\omega=0} \omega^k + O(\omega^{n+1})$$

$$= \sum_{k=0}^{n} \frac{1}{k!} \Big[\sum_{j=-N}^{\infty} h_j (-iTj)^k e^{-i\omega T j} \Big] \Big|_{\omega=0} \omega^k + O(\omega^{n+1})$$

$$= \sum_{k=0}^{n} \frac{1}{k!} \Big[(-iT)^k \sum_{j=-N}^{\infty} j^k h_j \Big] \omega^k + O(\omega^{n+1})$$

$$= (-iT) \sum_{j=-N}^{\infty} j h_j \omega + O(\omega^{n+1}) = i\omega + O(\omega^{n+1})$$

This proves (3.5). ∎

The ideal <u>phase shift</u> of a differentiator is $\pi/2$, cf (3.4). When the phase shift differs form this value, the computed derivative will be delayed, cf (3.3b).

A necessary condition for the phase shift of $H(e^{i\omega T};\eta)$ to be equal to the ideal $\pi/2$ is the following antisymmetry conditions:

$$h_j = -h_{-j} \quad \text{all } j \tag{3.7}$$

This can be shown as follows: We have

$$\pi/2 = \arg H(e^{i\omega T}) = \arg[\sum_j h_j e^{ij\omega T}]$$

$$= \arg[\{\sum_j h_j \cos j\omega T\} + i\{\sum_j h_j \sin j\omega T\}]$$

Since this must hold for all ω we have $\sum_j h_j \cos j\omega T \equiv 0$ from which (3.7) follows.

Step response

The step response of the filter is of interest to describe the filter behaviour when the derivative change from one constant level to another. Then the filter output will move gradually to a new level when the derivative changes. Consider a differentiating filter $H(z)$. In order to obtain an unbiased estimate of a linear trend (with z-transform $z/(z-1)^2$) we have to require

$$\lim_{z\to 1} (z-1)H(z)\frac{z}{(z-1)^2} = \frac{1}{T} \tag{3.8}$$

This follows directly from the properties of the z-transform and the relation (3.6d).

The speed with which the filter output changes is of interest. From a deterministic point of view it is desired that the changes can occur fast, so that the filter can easily adapt to new situations. This implies that the sequences of filter coefficients $\{h_j\}$ should tend to zero quickly. However, such a case has certainly drawbacks as well. Then only a quite limited amount of the data are effectively used, which can make the computations vulnerable for disturbances and noise effects. The step responses for some differentiating filters are illustrated and analysed in Section 5.2.

3.3 Stochastic behaviour

In this section we will discuss how to characterize the effects of the measurement noise $v(k)$ on the filter output. According to (2.8) the noise will give a variance of the filter output (or the stochastic error) as

$$
\begin{aligned}
W &= E[H(q^{-1};\eta)v(k)]^2 \\
&= \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} |H(e^{i\omega T};\eta)|^2 \Phi_v(\omega)d\omega
\end{aligned}
\tag{3.9}
$$

where $\Phi_v(\omega)$ is the spectral density of the disturbance $v(k)$. Again, the <u>frequency response</u> $H(e^{i\omega T};\eta)$ is a useful quantity for describing the effect of the noise. To keep W small it is necessary that the frequency response is small whenever the spectral density $\Phi_v(\omega)$ is large.

In a typical situation the signal is of low-frequency type while the measurement noise is white (meaning that $\Phi_v(\omega)$ is constant). Then the frequency response should be close to $i\omega$ for small frequencies (the frequency band where the signal clearly dominates over the noise). This will make the systematic error small. On the other hand for large frequencies (where the signal does not have a significant content) the frequency function should be small in order to make W small and reduce the noise effects. More generally, whether a full-band, wide-band or narrow-band differentiator should be used, depends very much on the characteristics of the disturbances (of course, the intention of the differentiation also plays an important role).

4. SOME DIFFERENTIATING FILTERS

4.1. Introduction

In this section methods for designing digital differentiating filters will be surveyed. Some comparisons and analysis will also be given.

A classification of different approaches can be done in many ways. Some typical classifications are given below

(i) on-line - off-line methods
(ii) narrow-band - wide-band methods
(iii) frequency domain - time domain methods

The first two classifications reflect the applicability of the filter while the third classification mainly reflects the design philosophy. In this report the third classification will be used. The ambition is not to cover all different methods described in the literature. Instead a number of typical methods are presented. Most methods are discussed quite briefly. The choice of a suitable differentiating filter is obviously application dependent. Different requirements (such as linear phase) and implementation issues will restrict the number of suitable design methods for a specific application. One important aspect is the underlying assumptions on the signal and noise characteristics. For a general description of digital filter design techniques we refer to Oppenheim and Schafer (1975).

4.2. Frequency domain techniques

The design of digital filters in the frequency domain is a popular approach. We will here describe some methods which have been used for designing differentiating filters. It is rather common to design wide-band and full-band differentiators using frequency domain techniques. This corresponds to minimizing the systematic error between the true and the estimated derivative, cf Section 2. Note that information (or assumptions) about the signal and noise spectrum makes it reasonable to use the non-causal differentiating Wiener filter defined in (2.30) as a starting point for the design.

4.2.1 Design from analog filters

The idea of this approach is to make a design in continuous-time and then approximate it to get a discrete-time filter. The continuous-time filter is set to be of the form

$$H(s) = sH_o(s) \tag{4.1}$$

where $H_o(s)$ typically is a low-pass filter ($H_o(0)=1$, $H_o(s)$ small for large frequencies). The frequency properties of the low-pass filter should reflect the properties of the signal. ($H_o(s)$ should be close to 1 in the frequency band of interest for the signal.)

Next the filter H(s) is approximated into discrete-time. This can be done in different ways. One simple possibility is to set

$$s = \frac{1-q^{-1}}{T} \tag{4.2}$$

while the bilinear transform (also called the Tustin's approximation)

$$s = \frac{2}{T}\frac{1-q^{-1}}{1+q^{-1}} \tag{4.3}$$

is another alternative.

A third procedure for transforming the analog filter (4.1) to a discrete-time filter is to use an impulse-invariance technique. That is we set

$$h(k) = h_a(t)\Big|_{t=kT} \tag{4.4}$$

where $h_a(t)$ is the impulse-response of (4.1). Kaiser (1966) gives an example of designing a wide-band differentiator based on this technique.

The method above apparently covers a large number of variants. The low-pass filter $H_o(s)$ can be chosen in a number of ways as can the approximation to discrete-time.

To make the discussion more specific let us give two examples.

Example 4.1. Set $H_o(s)=1/(1+s\tau)^2$ and use the simple approximation (4.2). In this case the parameter vector η consists of the single parameter τ. It should be chosen so that the frequency $1/\tau$ is small, yet that the signal has most of its frequency contents below $1/\tau$. The digital filter becomes

$$H(q^{-1};\eta) = \frac{s}{(1+s\tau)^2}\bigg|_{s=\frac{1-q^{-1}}{T}} = \frac{T(1-q^{-1})}{[T+\tau(1-q^{-1})]^2}$$

$$= \frac{1-q^{-1}}{T} \frac{1}{[1+\frac{\tau}{T}(1-q^{-1})]^2} \tag{4.5}$$

■

Example 4.2. Set $H_o(s)=\omega_o^2/(s^2+2\zeta\omega_o s+\omega_o^2)$ and use the bilinear transformation. In this case

$$\eta = (\omega_o \quad \zeta)$$

The choice of the frequency ω_o and the damping factor ζ should reflect the frequency characteristics of the signal. The digital filter becomes

$$H(q^{-1};\eta) = \frac{s\omega_o^2}{s^2+2\zeta\omega_o s+\omega_o^2}\bigg|_{s=\frac{2}{T}\frac{1-q^{-1}}{1+q^{-1}}}$$

$$= \frac{2T\omega_o^2(1-q^{-1})(1+q^{-1})}{4(1-q^{-1})^2+4\zeta\omega_o T(1-q^{-1})(1+q^{-1})+\omega_o^2T^2(1+q^{-1})^2} \tag{4.6}$$

■

4.2.2 Fourier series approaches

The pulse transfer function of the ideal (full-band) differentiating filter is (cf (2.15))

$$H_i(z) = \frac{1}{T}\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}(z^n-z^{-n}) \tag{4.7}$$

which has the frequency response

$$H_i(e^{i\omega T}) = i\omega \qquad |\omega|<\frac{\pi}{T} \tag{4.8}$$

A straightforward approach to get a practical (non-causal) filter is to truncate the infinite sum in (4.7). We then get the following non-causal differentiating FIR-filter

$$H(z;\eta) = \frac{1}{T}\sum_{n=1}^{N} \frac{(-1)^{n+1}}{n}(z^n-z^{-n}) \tag{4.9}$$

The parameter vector η just consists of the filter length parameter N. From (4.9) it follows directly that H(1)=0, i.e. the output from the filter is zero for a constant signal. The filter in (4.9) minimizes the following criterion

$$E = \int_{-\pi/T}^{\pi/T} |H_i(e^{i\omega T})-H(e^{i\omega T};\eta)|^2 d\omega \tag{4.10}$$

This follows directly from the properties of the Fourier series, cf also Usui and Amidror's approach described below.

The discontinuity at $\pm \frac{\pi}{T}$ of the ideal frequency response will produce an oscillating behaviour in the amplitude function of the truncated filter (4.9). This effect is known as the Gibb's phenomenon. The oscillations can be reduced (at the expense of the bandwidth) by multiplying the pulse response in (4.9) with a suitable weighting function w(t), i.e.

$$H(z;\eta) = \frac{1}{T}\sum_{n=1}^{N} w(n)\frac{(-1)^{n+1}}{n} (z^n-z^{-n}) \tag{4.11}$$

Typically w(0)=1 and w(t) goes to zero as t increases. A review of a number of different weighting functions is given in Harris (1978).

Another approach to reduce the Gibb's oscillations is to make a least squares optimization over a smaller interval than in (4.10), i.e.

$$E = \int_{-\alpha\frac{\pi}{T}}^{\alpha\frac{\pi}{T}} |H_i(e^{i\omega T})-H(e^{i\omega T};\eta)|^2 d\omega \tag{4.12}$$

where $\alpha<1$.

The Fourier series approach for designing wide-band differentiators have been studied in Kaiser (1966). In a design example it is shown that the Gibb's oscillations can be drastically reduced by using the modified filters (4.11) or (4.12) instead of the direct truncation in (4.9).

A related approach to the Fourier series approach is given by Usui and Amidror (1982a). It will be presented below.

Usui and Amidror's approach

Consider a strictly bandlimited signal (i.e. the signal does not contain frequencies above $\alpha\pi/T$, $\alpha \leq 1$) which is to be differentiated without any systematic error. The frequency response of the ideal differentiator which minimizes the variance of the stochastic error (cf (2.8)) under the constraint above is given by

$$H_{LPI}(e^{i\omega T}) = \begin{cases} i\omega & |\omega| \leq \alpha\frac{\pi}{T} \\ 0 & \alpha\frac{\pi}{T} < |\omega| \leq \frac{\pi}{T} \end{cases} \tag{4.13}$$

A FIR-filter with an odd antisymmetric impulse response will be used to approximate the frequency response in (4.13)

$$H(z;\eta) = \frac{1}{2T}\sum_{n=1}^{N} c_n(z^n - z^{-n}) \tag{4.14}$$

The filter parameter vector $C=(c_1\ c_2,\ldots c_N)^T$ in (4.14) will be chosen so that

$$E(\alpha,C) = \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} |H_{LPI}(e^{i\omega T}) - H(e^{i\omega T};\eta)|^2 d\omega \tag{4.15}$$

is minimized subject to the constraint

$$\left.\frac{dH(e^{i\omega T})}{d\omega}\right|_{\omega=0} = i \tag{4.16}$$

The constraint (4.16) means that $H(e^{i\omega T};\eta) \approx i\omega$ for small ω. This will give good low-frequency differentiation even for small values of N.

Using (4.13) and (4.14) the criterion in (4.15) is straightforward to evaluate. We get

$$E(\alpha,C) = \frac{1}{T^3}[\frac{2}{3}(\alpha\pi)^3 + \pi C^T C + C^T L] \tag{4.17}$$

where

$$L = (l_1 \ldots l_N)^T$$

$$l_p = \frac{4}{p^2}[p\alpha\pi\cos(p\alpha\pi) - \sin(p\alpha\pi)] \quad p=1,2,\ldots,N$$

The constraint (4.16) can be written

$$C^T u = 1 \tag{4.18}$$

where $u=(1,2...N)^T$

The optimization problem has a quadratic loss function and a linear constraint. Using standard optimization theory it can then be seen that the optimal filter parameter vector C_o is given by

$$C_o = \frac{1}{2\pi}\left[\frac{u(2\pi+u^T L)}{u^T u} - L\right] \tag{4.19}$$

The following remarks are now in order

- The parameter vector $\eta=(N\ \alpha)$, i.e. the design parameters are the filter length and the upper limit of the differentiation band. Note that generalization to higher order derivatives is straightforward.

- The constraint (4.16) guarantees unbiased differentiation of a linear trend signal. Furthermore, the structure of the filter (4.14) implies zero dc-gain which is a very reasonable property of a differentiating filter.

- From (4.19) it is easy to verify that $C_o \to -\frac{1}{2\pi}L$ as $N \to \infty$. This is the optimal solution to (4.15) <u>without</u> the constraint (4.16), i.e. the constraint is automatically fulfilled for large N. Furthermore, for $\alpha=1$ we get

  $$c_n \to \frac{2}{n}(-1)^{n+1}$$

  which inserted in (4.14) gives the ideal differentiating filter defined in (2.15).

- For $\alpha=0$ we have $C_o = \frac{u}{u^T u}$. This can be interpreted as an estimate of the slope of a linear trend using local least squares polynomial fitting, see Section 4.3.3. The filter minimizes the noise transmission under the constraint of ideal differentiating at $\omega=0$, cf also Usui and Amidror (1982b).

- The discontinuity at $\pm\alpha\frac{\pi}{T}$ will cause Gibb's oscillations. A possible remedy is to use a suitable weighting function as in (4.11).

4.2.3 Optimal IIR-filters

We will here briefly describe two frequency domain methods for optimizing IIR-filters.

Steiglitz's approach

Steiglitz (1970) proposed a method for designing IIR-filters with an arbitrary amplitude function. The desired amplitude function $|H_d(e^{iwT})|$ is given for a discrete set of frequencies $\{\omega_n\}_{n=1}^{M}$. The following least mean squares criterion is minimized

$$Q = \sum_{n=1}^{M} (|H(e^{i\omega_n T})|-|H_d(e^{i\omega_n T})|)^2 \tag{4.20}$$

where H(z) is a 2N:th order IIR-filter of cascade form

$$H(z) = A \prod_{n=1}^{N} \frac{1+b_{n1}z^{-1}+b_{n2}z^{-2}}{1+a_{n1}z^{-1}+a_{n2}z^{-2}} \tag{4.21}$$

The optimization in (4.20) is done with respect to the parameters in (4.21). Nonlinear optimization technique is used to find the minimum. It is advantageous to use the cascade form because of its relatively low coefficient sensitivity and convenience in calculating the derivatives in the optimization procedure. The filter coefficients in (4.21) will depend on the following parameter vector

$$\eta = (N, |H_d(e^{i\omega_1 T})|, |H_d(e^{i\omega_2 T})|, \ldots, |H_d(e^{i\omega_M T})|)$$

The design of a differentiating filter using (4.20) gives the user a number of choices. The desired frequency response $H_d(e^{i\omega_n T})$ can be chosen, for example, as in (4.13). If the signal and noise spectra are known the differentiating Wiener filter in (2.30) might be another reasonable choice.

Design examples of wide-band differentiators using the mean squares criterion (4.20) are given by Steiglitz (1970) and Rabiner and Steiglitz (1970).

Spriet and Bens's approach

The cascade form (4.21) is also used by Spriet and Bens (1979) for designing wide-band differentiators. To guarantee good low-frequency behaviour the following requirements are imposed

(i) The dc-gain should be zero.

(ii) Differentiation of a linear trend signal should be unbiased.

This implies that, cf (3.8)

$$H(z) = (1-z^{-1})H_1(z) \tag{4.22}$$

where $H_1(1)=1/T$. Using the cascade form, the requirements in (4.22) can be fulfilled by the following structure

$$H(z) = \frac{1}{T}(1-z^{-1})\frac{1-b_{12}z^{-1}}{1-b_{12}} \prod_{n=1}^{N} \frac{1+a_{n1}+a_{n2}}{1+a_{n1}z^{-1}+a_{n2}z^{-2}} \prod_{n=2}^{N} \frac{1+b_{n1}z^{-1}+b_{n2}z^{-2}}{1+b_{n1}+b_{n2}} \tag{4.23}$$

The design idea is to make a compromise between the amplitude and the phase errors in the filter. The following criterion to minimize is proposed

$$J = \int_0^{\pi/T} f(\omega)\varepsilon(\omega)d\omega \tag{4.24}$$

where $f(\omega)$ is a weighting function reflecting the a priori knowledge about the frequency content of the signal. The function $\varepsilon(\omega)$ is a measure of the relative errors in the amplitude and phase and is given by

$$\varepsilon(\omega) = \frac{[\delta A(\omega)]^2}{A(\omega)^2} + [\delta\phi(\omega)]^2 \tag{4.25}$$

where $A(\omega)$ is the desired amplitude gain (for an ideal differentiator $A(\omega)=\omega$). The error in the amplitude gain for the filter is denoted $\delta A(\omega)$ and the error in phase is $\delta\phi(\omega)$. The criterion (4.24) penalizes also phase errors in contrast to the least mean squares criterion in (4.20) where only the amplitude error is minimized. Spriet and Bens found that increasing the parameter N in (4.23) above 2 only gave minor reduction of the loss function in (4.24). If the full-band requirement is relaxed and the effective bandwidth is reduced, an increase in the order of the filter results in a more substantial reduction of the loss function.

4.2.4 Minimax optimization

The use of minimax optimization (Chebyshev approximation) technique for designing FIR-filters is a popular approach, see for example Rabiner et al (1975) for a detailed discussion.

Consider a FIR filter with a purely imaginary frequency response

$$H(z) = \frac{1}{T}\sum_{n=1}^{N} b_n(z^n - z^{-n}) \tag{4.26}$$

The filter parameters $b_1, b_2, \ldots b_N$ are optimized using the following criterion

$$\min_{b_1 \ldots b_N} \max_{\omega \in \Omega} W(e^{i\omega T})|H_d(e^{i\omega T}) - H(e^{i\omega T})| \tag{4.27}$$

where $W(e^{i\omega T})$ is a weighting function which can be used to choose the relative size of the error in different frequency bands. The desired (ideal) frequency response is denoted $H_d(e^{i\omega T})$. For designing a differentiating filter, $H_d(e^{i\omega T})$ can be chosen as in (2.30) or (4.13). The weighting function $W(e^{i\omega T})$ can be set to "high" values at frequencies where it is particularly important to have good performance, i.e. close agreement of $H(e^{i\omega T})$ to $H_d(e^{i\omega T})$.

The filter coefficients in (4.27) will then depend on the parameter vector

$$\eta = (N, W(e^{i\omega T}), H_d(e^{i\omega T}))$$

Various design examples of wide-band differentiators using minimax optimizations techniques can be found in, for example, Hofstetter et al (1971), McClellan and Parks (1973), McClellan et al (1973), Antoniou and Charalambous (1981) and Morgera (1982).

In McClellan et al (1973), a Fortran program for minimax optimization is given. Among various different filter types an example of a full-band differentiator is given. A modification to the program is proposed in Rahenkamp and Kumar (1986) to allow the design of higher order differentiating filters. Antoniou and Charalambous (1981) proposed a method to design equiripple differentiators with prescribed specifications in the error function by searching also over the filter order N. An application to pilot tone syncronization in a communication system is reported in Morgera (1982), where a differentiating FIR-filter based on minimax optimization is used.

4.2.5 Frequency sampling

We will end the frequency domain section with a very brief discussion of another popular method for designing FIR-filters, namely the frequency sampling design.

The method is based on the fact (cf for example Oppenheim and Schafer (1975), pp 251) that an N:th order FIR-filter can be represented in terms of N frequency samples $\{\tilde{H}(n)\}_{n=0}^{N-1}$ of the filters frequency response, i.e.

$$H(z) = \frac{1-z^{-N}}{N} \sum_{n=0}^{N-1} \frac{\tilde{H}(n)}{1-z^{-1}e^{i\frac{2\pi n}{N}}} \tag{4.28}$$

A simple approach to construct a discrete-time filter is then to sample the desired frequency response $H_d(e^{i\omega T})$ and set

$$\tilde{H}(n) = H_d(e^{i\omega T})\Big|_{\omega=\frac{2\pi}{NT}n} \qquad n=0,1,\ldots,N-1 \tag{4.29}$$

The filter in (4.28) will then coincide with the desired frequency response at the N sampling points. The filter in (4.28) is determined by the parameter vector

$$\eta = (N, \{H_d(e^{i\frac{2\pi}{N}n})\}_{n=0,1,\ldots,N-1})$$

To improve the intersample behaviour a few samples can be set free and optimized.

The design of differentiating FIR-filters is straightforward using for example (4.13) or (2.30) as the desired frequency response.

Examples of full-band and wide-band differentiators based on frequency sampling can be found in Rabiner and Steiglitz (1970) and Rabiner et al (1970). In the first reference it is empirically found that a doubling of the filter order N tends to half the maximum magnitude error in the frequency response of the filter. It is also found that by reducing the bandwidth (differentiation band), the peak magnitude error drops significantly.

4.3 Polynomial signal modelling

We will here describe a number of approaches based on polynomial signal modelling. The idea is to model the signal $s_c(t)$ as a (piecewise) polynomial. Having an estimate of the parameters in a polynomial model, an estimate of the derivative is easily obtained (cf (3.6d)). Generalization to higher order derivative is straightforward.

4.3.1 Polynomial interpolation

Interpolation is a way of constructing a curve through a given set of points. A common method is to use polynomials for interpolating, see for example Conte and de Boor (1972) for a detailed presentation. There exists exactly one polynomial of degree n, which coincides with the signal at n+1 given points. The application to differentiation is straightforward:

- Determine the interpolating polynomial for a number of data points.
- Estimate the derivative (of the signal) by differentiating the polynomial.

A classical method for constructing the interpolating polynomial $s_c(t)$ from the discrete-time points $y(k)$, $k=0,1,\ldots,2n$ (for convenience we let the number of data points be odd) is the Lagrange interpolation formula

$$s_c(t) = \sum_{k=0}^{2n} K_k(t)y(k) \tag{4.30}$$

where

$$K_k(t) = \frac{\pi_{2n}(t)}{(t-kT)\pi'_{2n}(kT)}$$

$$\pi_{2n}(t) = (t-0)(t-T)\ldots(t-2nT)$$

$$\pi'_{2n}(kT) = (kT-0)\ldots(kT-(k-1)T)(kT-(k+1)T)\ldots(kT-2nT)$$

Straightforward differentiation of (4.30) gives

$$\dot{s}_c(t) = \sum_{k=0}^{2n} L_k(t)y(k) \tag{4.31}$$

where

$$L_k(t) = \ell_k(t) = \sum_{\substack{j=0\\ j\neq k}}^{2n} \frac{\pi_{2n}(t)}{(t-kT)(t-jT)\pi'_{2n}(kT)} \tag{4.32}$$

Note that (4.31) can be used as a predictor, filter or a smoother. For the smoothing case we present the following theorem

Theorem 4.1. Let $t=nT$ in (4.31). Then we obtain the following differentiation filter

$$H(z) = \sum_{k=1}^{n} b_k(z^k - z^{-k}) \tag{4.33a}$$

where

$$b_k = \frac{1}{T}\frac{(-1)^{k+1}}{k}\frac{n!n!}{(n+k)!(n-k)!} \tag{4.33b}$$

Proof. We will first show that $L_n(nT)=0$ in (4.32). We have

$$L_n(nT) = \sum_{\substack{j=0\\ j\neq n}}^{2n} \frac{nT(nT-T)\ldots(nT-(n-1)T)(nT-(n+1)T)\ldots(nT-2nT)}{(nT-jT)[nT(nT-T)\ldots(nT-(n-1)T)(nT-(n+1)T)\ldots(nT-2nT)]}$$

$$= \sum_{\substack{j=0\\ j\neq n}}^{2n} \frac{1}{(nT-jT)} = 0$$

For $k\neq n$ we have

$$L_k(nT) = \sum_{\substack{j=0\\ j\neq k}}^{2n} \frac{t(t-T)\ldots}{(t-kT)(t-jT)[kT(kT-T)\ldots}$$

$$\left.\frac{\ldots(t-2nT)}{\ldots(kT-(k-1)T)(kT-(k+1)T)\ldots(kT-2nT)]}\right|_{t=nT}$$

$$= \frac{nT(nT-T)\ldots(nT-(n-1)T)(nT-(n+1)T)\ldots(nT-2nT)}{(nT-kT)[(kT(kT-T)\ldots(kT-(k-1)T)(k-(k+1)T)\ldots(k-2nT)]} =$$

$$= \frac{1}{T}\frac{n(n-1)\ldots1(-1)\ldots(-n)}{(n-k)\{k(k-1)\ldots1(-1)\ldots(k-2n)\}} = \frac{1}{T}(-1)^{n-k}\frac{n!n!}{(n-k)k!(2n-k)!}$$

It follows directly that

$$L_k(nT) = -L_{2n-k}(nT)$$

Insertion of L_k in (4.31) gives easily

$$b_k = L_{n+k}(nT) = \frac{1}{T}(-1)^{n-n-k} \frac{n!n!}{(n-n-k)(n+k)!(n-k)!} \qquad k=1,\dots,n$$

which proves (4.33b). ∎

The smoother in (4.33) can be interpreted as a noncausal antisymmetric FIR-filter. The filter parameter vector is just $\eta=(n)$.

Next the relation with the ideal differentiation filter will be discussed. Recall from (2.15) that the ideal differentiating filter is given by

$$H(z) = \sum_{k=1}^{n} \beta_k(z^k - z^{-k}) \tag{4.34a}$$

$$\beta_k = \frac{1}{T}\frac{(-1)^{k+1}}{k} \tag{4.34b}$$

The filter weights in (4.33b) can hence be written as

$$b_k = \beta_k f(n,k) \tag{4.35a}$$

where

$$f(n,k) = \frac{n!n!}{(n+k)!(n-k)!} \tag{4.35b}$$

We then have the following proposition.

<u>Proposition 4.1</u>. Let k be fix, then

$$f(n,k) \to 1 \text{ as } n\to\infty$$

<u>Proof</u>.

$$f(n,k) = \frac{n!n!}{(n+k)!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{(n+1)(n+2)\dots(n+k)} \to 1 \text{ as } n\to\infty$$

∎

The proposition shows that for large n and small k, the filter weights in (4.33b) approximate the "ideal" weights defined in (4.34b). We next give an example to illustrate the differentiating filter in (4.33).

Example 4.3. By using Theorem 4.1, it is straightforward to calculate the filter weights. For n=2 we get

$$H(z) = \frac{1}{12T}(-z^2+8z-8z^{-1}+z^{-2}) \tag{4.36}$$

Since the filter is antisymmetric, the frequency response is purely imaginary. Figure 4.1 shows the amplitude response of the filter in (4.36).

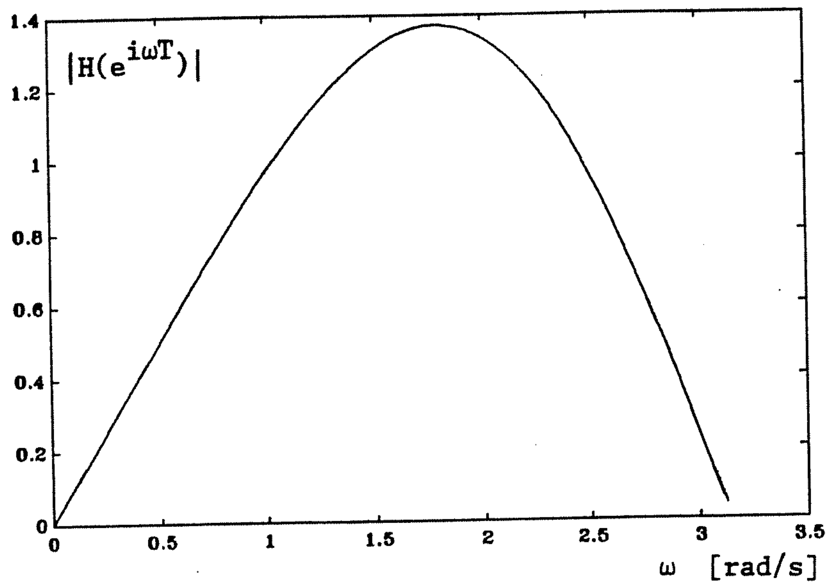


Figure 4.1. Amplitude response of the filter in (4.36), (T=1).

■

The use of interpolation methods for differentiation can be classified as an wide-band method. Interpolation methods are hence very sensitive for noise in the signal, cf for example Koenig (1966), Allum (1975) and Lavery (1979). This is very natural since interpolation is exact in the sense that it passes through each data points.

4.3.2 Splines

A further drawback with polynomial interpolation, apart from the noise sensitivity, is that the interpolation polynomial becomes excessively oscillatory as the number of interpolating points increase. The last drawback can be eliminated by using piecewise polynomials, i.e. different polynomials in different intervals. Of particular interest are splines which are piecewise polynomials of degree n adjusted so that the (n-1):th derivative is continuous. For a detailed description of spline functions we refer to de Boor (1978).

A straightforward approach to estimate the derivative, using splines, is to

(i) approximate the data by a spline function

(ii) differentiate the spline

For noisy data other approximation methods than interpolation should be used in step (i). The problem is to find a balance between the goodness of fit to the data and the sensitivity to noise, see Morozov (1984) for a general treatement.

One attempt to make a trade-off between the two conflicting goals above is to use the following quadratic criterion

$$\min_{s_s(t)} \Big[\sum_{k=1}^{N} |s_s(kT)-y(k)|^2 + p \int_T^{NT} [s_s^m(t)]^2 dt \Big] \qquad (4.37)$$

where $p>0$, and $y(k)$ is discrete-time noisy measurements.

The minimization is done over the class of functions $\{s_s(t)\}$ which are continuous up to the m-1 order of the derivative and squared integrable in the m:th order derivative on the interval (T,NT). The minimization in (4.37) results in a spline of degree 2m-1 (order 2m). The spline function is determined by the choice of p and m. For $p=0$ the problem reduces to interpolation. For the asymptotic case, $p\to\infty$, the minimum in (4.37) is obtained for a single polynomial of degree m-1, which fits the measurement data in a least square sense. This is intuitively seen in (4.37). For a polynomial of degree m-1 the integrand is zero, and the first part in (4.37) is a least squares criterion, cf Section 4.3.3 below.

A nontrivial problem is to choose the regularizing parameter p. A too large value of p produces an oversmoothing of the data (large systematic error) whereas a too small value may give an undesirably high noise sensitivity (large stochastic error). Different types of criteria for selecting p exist. A popular method for selecting p is to use cross-validation, cf Wahba and Wold (1975) and Craven and Wahba (1979). The basic idea of cross-validation is to leave the data points out one at a time and then choose the value of p which best predicts the missing data points. An earlier approach is reported in Anderssen and Bloomfield (1974), who assume that y(k) is a stationary stochastic process and estimate the regularizing parameter from a Fast Fourier Transform of the data.

Differentiation with optimally regularized splines applied to biomechanics is given in, for example, Woltring (1985). A program package based on optimal regularisation of (4.37) is presented in Woltring (1986). Some illustrations of regularized splines (including the choices of p and m) applied to differentiation are presented in Section 5.2.

4.3.3 Local least squares polynomial fitting

Differentiating filters based on local least squares polynomial fitting will be described. A polynomial of degree n will be fitted to 2N+1 measurement points. The estimate is made in the middle of the interval.This implies that a noncausal filter (smoother) is used.

For convenience we consider the r:th (r<n) order estimate of the derivative. The discrete-time measurement data is denoted

$$y(k) \quad k=0,1,2...$$

The problem under study is to fit a polynomial so that the squared sum of the curve fitting error is minimized.

The polynomial is represented in shifted power form

$$s(k_o) = \sum_{i=0}^{n} a_i (k_o - k)^i \tag{4.38}$$

Let $a=(a_o \; a_1, ... a_n)^T$. The local least squares estimate of the vector a is defined as the vector a that minimizes the loss function

$$V(a) = \sum_{j=-N}^{N} [y(k+j)-s(k+j)]^2 \tag{4.39}$$

which has the solution

$$U^T U \hat{a}(k) = U^T Y(k) \tag{4.40}$$

where

$$u_i = ((-N)^i, \dots\ 0^i, \dots\ N^i)^T$$

$$U = (u_o, u_1, \dots, u_n)$$

$$Y(k) = (y(k-N), \dots\ y(k), \dots\ y(k+N))^T$$

The r:th order derivative of the (continuous-time) polynomial is estimated by (cf (3.6d) and (4.38))

$$\left.\frac{d^r \hat{s}_c(t)}{dt^r}\right|_{t=kT} = \hat{s}_c^r(kT) = \frac{r!}{T^r}\hat{a}_r(k)$$

which using (4.40) can be written

$$\hat{s}_c^r(kT) = \frac{r!}{T^r}\, e_{r,n}^T (U^T U)^{-1} U^T Y(k) \tag{4.41}$$

where

$$\begin{array}{rcl} e_{r,n} = & & [0,0,\dots 0\ \ 1\ \ 0,\dots 0]^T \\ & & \ \vdots \qquad\quad \vdots \qquad \vdots \\ & \text{pos} & \ 0,\dots \quad\ r,\dots \quad n \end{array}$$

This gives the following differentiating filter

$$x(k) = \sum_{j=-N}^{N} h_j y(k-j) = h^T Y(k) \tag{4.42a}$$

where

$$h = (h_N \dots h_0 \dots h_{-N})^T = \frac{r!}{T^r} U (U^T U)^{-1} e_{r,n} \tag{4.42b}$$

The filter in (4.42) is an antisymmetric noncausal FIR-filter. The filter is determined by the design parameter vector (T and r are assumed to be given)

$$\eta = (N, n)$$

where 2N+1 is the number of data points and n is the degree of the polynomial.

Example 4.4. Consider the case r=1, N=2 and n=2. From (4.42) we get

$$h = \frac{1}{T}[u_o \; u_1 \; u_2][\begin{bmatrix} u_o^T \\ u_1^T \\ u_2^T \end{bmatrix}(u_o \; u_1 \; u_2)]^{-1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{10T}u_1 = \frac{1}{10T}\begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$$

which gives the following differentiating filter

$$H(z) = \frac{1}{10T}(2z^2+z-z^{-1}-2z^{-2}) \tag{4.43}$$

Figure (4.2) shows the amplitude response of the filter in (4.43).

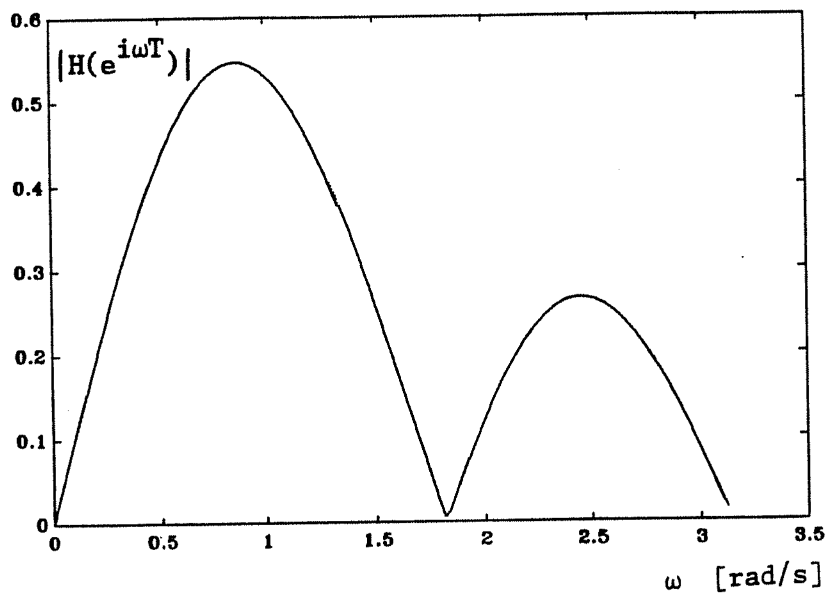


Figure 4.2. Amplitude response of a second order least squares polynomial differentiator based on 5 data points, (T=1).

Note the clear difference with the interpolation polynomial approach in Example 4.3 (the figures have different scales). ∎

Bandwidth and noise amplification for the local least squares differentiating filter have been studied by Lanshammar (1982).

Consider the case r=n=1 in the differentiating filter above. This corresponds to the classical estimate of the first order derivative by local least squares fitting of a linear trend. From (4.42b) we obtain the following impulse response

$$h_j = \frac{j}{2T\sum_{i=1}^{N} i^2} \tag{4.44}$$

This case is widely known as Lanczos method, see Lanczos (1956). Note that the same filter weights are obtained for $\alpha=0$ in Usui and Amidror's approach in Section 4.2.2.

Lanshammar's method

The method described here is given by Gustafsson and Lanshammar (1977). A slight extension is reported by Söderström (1980). The basic assumptions in Lanshammar's approach are the following:

(i) The measurements are given by

$$y(k) = s_n(k)+e(k) \tag{4.45}$$

where $e(k)$ is white noise with zero mean and variance λ^2. The signal $s_n(k)$ is approximated by a polynomial of degree n $(n>r)$ where r is the order of the derivative to be estimated.

(ii) A noncausal FIR-filter is used for estimating the r:th order derivative

$$\hat{s}_c(kT) = \sum_{j=-N}^{N} h_j y(k-j) = h^T Y(k) \tag{4.46}$$

$$h = [h_N, \ldots h_0, \ldots h_{-N}]^T$$

$$Y(k) = [y(k-N), \ldots y(k), \ldots y(k+N)]^T$$

(iii) The systematic (deterministic) error for polynomials of degree $n-1$ is constrained to be zero.

(iv) The filter weights $\{h_j\}$ are determined such that the variance of the total error (see (2.8)) is minimized.

Assuming $s_n(k)$ to be a polynomial of degree n gives

$$s_n(k+j) = \sum_{i=0}^{n} \frac{(jT)^i}{i!} s_c^i(kT) \tag{4.47}$$

where $s_c^i(t)$ denotes the i:th order derivative of the underlying continuous-time polynomial.

Using (4.45) and (4.46), the variance of the total error can be written

$$V(h) = [h^T S_n^T(k) - s_c^r(kT)]^2 + \lambda^2 h^T h \tag{4.48}$$

where $S_n(k) = [s_n(k-N), \dots s_n(k), \dots s_n(k+N)]^T$

The constraint in (ii) gives

$$h^T S_{n-1}(k) = s_c^r(kT) \tag{4.49}$$

We have a quadratic criterion with a linear constraint. A solution in closed form then exists. The solution is presented in the following lemma.

<u>Lemma 4.2</u>. Consider the loss function V (h) in (4.48) and the constraint in (4.49). The optimal vector h^o is given by

$$h^o = [\varrho u_n u_n^T + \lambda^2 I]^{-1} U_{n-1} [U_{n-1}^T (\varrho u_n u_n^T + \lambda^2 I)^{-1} U_{n-1}]^{-1} \frac{r!}{T^r} e_{r,n-1} \tag{4.50}$$

where

$$u_n = [(-N)^n, \dots 0^n, \dots N^n]^T$$

$$U_{n-1} = [u_o, u_1 \dots u_{n-1}]$$

$$\begin{array}{llll} e_{r,n-1} = [0 \dots & 010 & \dots & 0]^T \\ \text{pos} \quad 0 \dots & r \dots & & n-1 \end{array}$$

$$\varrho = [\frac{T^n}{n!} s_c^n(kT)]^2$$

<u>Proof</u>. From (4.47) it is clear that

$$S_n(k) = \sum_{i=0}^{n} \frac{T^i}{i!} u_i s_c^i(kT) \tag{4.51}$$

where $u_i = [(-N)^i \dots 0^i \dots N^i]^T$.

Using the constraint (4.49) and (4.51), the loss function (4.48) can be written

$$V(h) = [\frac{T^n}{n!} s_c^n(kT)u_n^T h]^2+\lambda^2 h^T h = \varrho h^T u_n u_n^T h+\lambda^2 h^T h \tag{4.52}$$

where

$$\varrho=[\frac{T^n}{n!} s_c^n(kT)]^2.$$

Using (4.51), we then get from (4.49)

$$u_i^T h = \begin{cases} 0 & i\neq r \\ \frac{r!}{T^r} & i=r \end{cases} \qquad i=0,1,\ldots n-1$$

which, using the notations above can be written

$$U_{n-1}^T h = \frac{r!}{T^r} e_{r,n-1} \tag{4.53}$$

The static optimization problem with the quadratic loss (4.52) is straightforward to solve and directly gives (4.50). ∎

By using the matrix inversion lemma (cf for example Kailath (1980)) the optimal filter weights vector can be written

$$\begin{aligned} h^o &= [(1+\alpha u_n^T u_n)I-\alpha u_n u_n^T]U_{n-1} \\ &\quad [(1+\alpha u_n^T u_n)U_{n-1}^T U_{n-1}-\alpha U_{n-1}^T u_n u_n^T U_{n-1}]^{-1} \frac{r!}{T^r} e_{r,n-1} \end{aligned} \tag{4.54}$$

where $\alpha=\varrho/\lambda^2$.

From (4.52) it is clear that the systematic error is proportional to ϱ and the stochastic error is proportional to λ. Therefore α represent a balance between these two errors. The factor α can be regarded as a design variable. The filter parameters are determined by the design parameter vector

$$\eta = (N \; n \; \alpha)$$

From (4.50) it follows directly that

$$\lim_{\alpha\to 0} h^{o} = U_{n-1}[U_{n-1}^{T}U_{n-1}]^{-1} \frac{r!}{T^{r}} e_{r,n-1}$$

which can be interpreted as a local least squares fitting of a polynomial of degree n-1, cf (4.41).

From (4.54) it can be verified that

$$\lim_{\alpha\to\infty} h^{o} = U_{n}(U_{n}^{T}U_{n})^{-1} \frac{r!}{T^{r}} e_{r,n}$$

i.e. a least squares fitting of a polynomial of degree n.

The factor α can be used to balance the systematic and stochastic error terms in the criterion function. It turns out in practice that the use of a nonzero α can have a substantial and beneficial effect on the filter, especially on the damping for high frequencies, see Gustafsson and Lanshammar (1977) for a discussion.

4.3.4 Recursive estimation of polynomial models

We have above described some non-causal polynomial approaches for differentiation. The natural solution for an on-line application is to use a recursive algorithm. Some common recursive algorithms will be briefly reviewed below.

A polynomial trend

$$s(k) = a_{o}+a_{1}k+,\ldots a_{n}k^{n} \tag{4.55}$$

can be written in the linear regression form

$$s(k) = \varphi^{T}(k)\theta \tag{4.56}$$

where $\varphi^{T}(k)=(1\ k,\ldots k^{n})$ and $\theta^{T}=(a_{o}\ a_{1}\ldots a_{n})$.

Furthermore, let the measurement be given by

$$y(k) = s(k)+e(k) \qquad k=0,1,2\ldots \tag{4.57}$$

where e (k) is a white noise sequence with zero mean. The least squares estimate of θ minimizes the loss function

$$V_{N}(\theta) = \sum_{k=1}^{N} [y(k)-\varphi^{T}(k)\theta]^{2} \tag{4.58}$$

where N is the number of data points. The classical recursive least squares algorithm for computing this estimate is given by

$$\varepsilon(k) = y(k)-\varphi^T(k)\hat{\theta}(k-1) \tag{4.59a}$$

$$\hat{\theta}(k) = \hat{\theta}(k-1)+P(k)\varphi(k)\varepsilon(k) \tag{4.59b}$$

$$P(k) = P(k-1)-\frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{1+\varphi(k)^T P(k-1)\varphi(k)} \tag{4.59c}$$

see, for example, Ljung and Söderström (1983). An on-line estimate of the derivative is obtained via

$$\hat{\dot{s}}_c(kT) = \frac{1}{T}\varphi_D^T(k)\hat{\theta}(k) \tag{4.60}$$

where $\varphi_D^T(k)=(0,1,2k,\ldots nk^{n-1})^T$. The algorithm (4.59) weights all data equally. This will give poor trackability to parameter changes.

Unknown time variations in the parameters may be tracked by introducing a forgetting factor in the criterion

$$V_N(\theta) = \sum_{k=1}^{N} \lambda^{N-k}[y(k)-\varphi^T(k)\theta]^2 \quad 0<\lambda\leq 1 \tag{4.61}$$

which changes (4.59c) to

$$P(k) = [P(k-1)-\frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{\lambda+\varphi(k)^T P(k-1)\varphi(k)}]\frac{1}{\lambda} \tag{4.59c'}$$

Another method for tracking time variations is to use an algorithm with rectangular data weighting which corresponds to the criterion

$$V_N(\theta) = \sum_{k=N-n}^{N} [y(k)-\varphi^T(k)\theta]^2 \tag{4.62}$$

The estimate of θ is then based on the last n samples. For a general treatment of the recursive least squares algorithm with rectangular data weighting see Young (1984).

Allum (1975) have used a cubic polynomial model. The parameters are estimated recursivly with a rectangular data weighting. The derivative is estimated from (4.60).

General exponential smoothing

We will describe general exponential smoothing which is an algorithm for recursive estimation. This method is extensively described in many textbooks in statistical forecasting, see for example Brown (1962) and Abraham and Ledolter (1983).

We will here focus on polynomial estimation applied to differentiation. The presentation here is based on Carlsson (1987). First the general ideas will be given and thereafter the application to differentiation will be presented.

Consider the following (discrete-time) model

$$y(k) = \varphi^T\theta(k)+v(k) \tag{4.63a}$$

$$\theta(k+1) = F\theta(k) \tag{4.63b}$$

where F is a fixed transition matrix, $\theta(k)$ is a parameter vector, φ is a constant regression vector and $v(k)$ is a zero mean white noise sequence.

The discounted least squares estimate of $\theta(k)$ is defined as the vector $\hat{\theta}(k)$ that minimizes the following loss function

$$V_N(\theta) = \sum_{k=1}^{N} \lambda^{N-k}[y(k)-\varphi^T(k)F^k\theta(0)]^2 \tag{4.64}$$

where $0<\lambda\leq 1$.

The recursive solution is given by

$$\hat{\theta}(k) = F\hat{\theta}(k-1)+P(k)\varphi[y(k)-\varphi^T F\hat{\theta}(k-1)] \tag{4.65a}$$

$$P(k) = [FP(k-1)F^T - \frac{FP(k-1)F^T\varphi\varphi^T FP(k-1)F^T}{\lambda+\varphi^T FP(k-1)F^T\varphi(k)}]\frac{1}{\lambda} \tag{4.65b}$$

Denote $P=\lim_{k\to\infty} P(k)$. Replacing $P(k)$ in (4.65a) with its asymptotic value P gives

$$\hat{\theta}(k) = F\hat{\theta}(k-1)+P\varphi[y(k)-\varphi^T F\hat{\theta}(k-1)] \tag{4.66a}$$

$$P = [\sum_{s=0}^{\infty} \lambda^s (F^{-s})^T\varphi\varphi^T F^{-s}]^{-1} \tag{4.66b}$$

The derivation of (4.65) and (4.66) can be found in Carlsson (1987). The algorithm (4.66) is known as <u>general exponential smoothing</u>. The general exponential smoothing algorithm (4.66) needs less computations than (4.65) since the gain in (4.66) can be computed in advance. A straightforward use of (4.65) may also be numerically unstable.

Next the applicability of general exponential smoothing for estimating the parameters in a polynomial model will be described.

A polynomial

$$s(k) = a_o + a_1 k + a_2 \frac{k^2}{2!} + \ldots a_n \frac{k^n}{k!} \tag{4.67}$$

can be represented by the structure (4.63) by choosing

$$\varphi^T = (1\ 0 \ldots 0) \tag{4.68a}$$

$$\theta^T(0) = (a_o\ a_1 \ldots a_n) \tag{4.68b}$$

$$F = \begin{bmatrix} 1 & 1 & \frac{1}{2!} & \cdots & \frac{1}{n!} \\ 0 & 1 & \cdots & & \frac{1}{(n-1)!} \\ & \mathbf{0} & & & \\ & & & & 1 \end{bmatrix} \tag{4.68c}$$

In Carlsson (1987) it is shown that

$$F^t = \begin{bmatrix} 1 & k & \frac{k^2}{2!} & \cdots & \frac{k^n}{n!} \\ 0 & 1 & k & \cdots & \\ & \mathbf{0} & & & \\ & & & & 1 \end{bmatrix} \tag{4.69}$$

The structure (4.67) is then directly obtained by observing that (4.63) with v(k)=0 can be written

$$y(k) = \varphi^T F^k \theta(0) \tag{4.70}$$

Next we will give an example for estimating the parameters in a linear trend model.

Example 4.5. Consider a linear trend corrupted by noise $v(k)$

$$y(k) = a_o + a_1 k + v(k) \tag{4.71}$$

Using (4.68) we can represent the signal by the model structure (4.63) with $\varphi = (1\ 0)^T$, $\theta(k) = [a_o(k)\ a_1(k)]^T$, $\theta(0) = (a_o\ a_1)^T$ and $F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, i.e.

$$\begin{aligned} y(k) &= a_o(k) + v(k) \\ a_o(k+1) &= a_o(k) + a_1(k) \\ a_1(k+1) &= a_1(k) \end{aligned} \tag{4.72}$$

The parameters can be estimated with the algorithm (4.66). Straightforward calculations give, using (4.66b)

$$P\varphi = \begin{bmatrix} 1-\lambda^2 \\ (1-\lambda)^2 \end{bmatrix}$$

which inserted in (4.66) gives the following algorithm for estimating the parameters in the linear trend

$$\hat{a}_o(k) = \hat{a}_o(k-1) + \hat{a}_1(k-1) + (1-\lambda^2)(y(k) - \hat{a}_o(k-1) - \hat{a}_1(k-1)) \tag{4.73a}$$

$$\hat{a}_1(k) = \hat{a}_1(k-1) + (1-\lambda)^2 (y(k) - \hat{a}_o(k-1) - \hat{a}_1(k-1)) \tag{4.73b}$$

The algorithm (4.73) is widely known as double exponential smoothing, cf Brown (1962). ∎

The following example illustrate double exponential smoothing applied to differentiation.

Example 4.6. (Differentiation with double exponential smoothing)

Using the parametrization (4.72) the derivative of a linear trend can be written

$$\dot{s}_c(kT) = \frac{1}{T} a_1(k) \tag{4.74}$$

Double exponential smoothing gives the following estimate

$$\hat{\dot{s}}_c(kT) = \frac{1}{T} \hat{a}_1(k) \tag{4.75}$$

Double exponential smoothing is a linear operation. Hence the estimate in (4.75) can be described by a pulse-transfer operator. From (4.73) and (4.75) the following differentiating filter is obtained

$$H(q^{-1}) = \frac{1}{T}(1-\lambda)^2 \frac{1-q^{-1}}{(1-\lambda q^{-1})^2} \tag{4.76}$$

The design variable is the forgetting factor λ. The filter has a zero for $z=1$ and a double pole for $z=\lambda$.

Note that the same filter structure is obtained in Example 4.1. It is easy to verify that the filter in (4.76) becomes equal to (4.5) if we set

$$\lambda = \frac{\tau}{T+\tau} \tag{4.77}$$

We have hence a relation between the cut-off frequency $1/\tau$ and the forgetting factor λ. A priori knowledge of the signals frequency characteristics can then be used to choose the forgetting factor λ. Relations between the double exponential smoothing and the Kalman filter are given in Carlsson (1987). ∎

General exponential smoothing can be used for filtering (smoothing) a noisy signal. From (4.66) we obtain

$$\hat{y}(k/k) = \varphi\hat{\theta}(k) = \varphi[I-(I-P\varphi\varphi^T)Fq^{-1}]^{-1}P\varphi y(k) \tag{4.78}$$

An estimate of the derivative can then be obtained by differentiating the output from the filtered signal in (4.78).

4.4 Stochastic signal modelling

A stochastic signal model is a more general signal model than polynomials. Two approaches applicable to differentiation based on stochastic signal modelling will be described.

4.4.1 A state space continuous-time formulation

This approach is described in more detail by Söderström (1980, 1982) and Ahlén (1984). A similar approach adapted to velocity estimation based on position measurements can be found in Ljung and Glad (1984). A related contribution is also given by Anderssen and Bloomfield

(1974) where time-series analysis is used to describe the signal but the filter is obtained via spectral properties.

The underlying continuous-time signal $s_c(t)$ is assumed to be described as a process with the spectral density

$$\phi_s(\omega) = \frac{B_p(i\omega)B_p(-i\omega)}{A_p(i\omega)A_p(-i\omega)} \tag{4.79}$$

where

$$A_p(s) = s^n + a_1 s^{n-1} + \dots a_n$$

$$B_p(s) = b_{r+1} s^{n-r-1} + \dots b_n$$

The problem is to estimate the r:th order derivative of the underlying continuous-time signal at the sampling points. Note that the coefficients $b_1, b_2 \dots b_r$ must be assumed to be zero to guarantee the existence of the derivative $s_c^r(t)$.

The signal can be represented in state space form as

$$\begin{aligned} dx(t) &= Ax(t)dt + Bdv \\ s_c(t) &= Cx(t) \end{aligned} \tag{4.80}$$

with

$$A = \begin{bmatrix} -a_1 & -a_2 & \cdot & \cdot & \cdot & -a_n \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \\ & \cdot & & & & \cdot \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & & & & 1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

$$C = [0 \; \dots \; b_{r+1} \; \dots \; b_n]$$

and $v(t)$ being a Wiener process with unit incremental variance.

The r:th order derivative $s_c^r(t)$ is given by

$$\begin{aligned} s_c^r(t) &= Dx(t) \\ D &= (b_{r+1}, b_r, \dots b_n \; 0, \dots 0) \end{aligned} \tag{4.81}$$

Sampling of the process gives, see for example Åström (1970),

$$\begin{aligned} x(k+1) &= Fx(k)+w(k) \\ s(k) &= Cx(k) \end{aligned} \tag{4.82}$$

where $F=e^{AT}$, and $w(k)$ is a vector-valued white noise sequences with zero mean and covariance matrix

$$Ew(k)w^T(k) \overset{\Delta}{=} R = \int_0^T e^{As}BB^Te^{A^Ts}ds \tag{4.83}$$

The measurements are assumed to be corrupted by noise which give

$$\begin{aligned} x(k+1) &= Fx(k)+w(k) \\ y(k) &= Cx(k)+e(k) \end{aligned} \tag{4.84}$$

where $e(k)$ is discrete-time white noise with zero mean and variance λ^2.

Optimal state estimation can now be used for estimating the derivative. One reasonable approach is to apply a time-invariant fixed lag smoother, cf Anderson and Moore (1979)

$$\hat{x}(k+1|k) = F\hat{x}(k|k-1)+K\tilde{y}(k) \tag{4.85a}$$

$$\tilde{y}(k) = y(k)-C\hat{x}(k|k-1) \tag{4.85b}$$

$$\hat{x}(k|k+m) = \hat{x}(k|k+m-1)+K_m^a\tilde{y}(k+m) \tag{4.85c}$$

$$K_m^a = P((F-KC)^T)^mC^T(CPC^T+\lambda^2)^{-1} \tag{4.85d}$$

$$P = FPF^T+R-FPC^T(CPC^T+\lambda^2]^{-1}CPF^T \tag{4.85e}$$

$$K = FPC^T[CPC^T+\lambda^2]^{-1} \tag{4.85f}$$

Let $R=QQ^T$. Sufficient conditions for stability of the filter above are then that (F,Q) is stabilizable (all uncontrollable modes have eigenvalues inside the unit circle) and that (C,F) is detectable (all unobservable modes have eigenvalues inside the unit circle).See eg Anderson and Moore (1979) for a proof. Note that for unstable system the spectrum in (4.79) is not defined.

The derivative in the sampling points can then be estimated by

$$\hat{s}_c^r(kT) = D\hat{x}(k|k+m) \tag{4.86}$$

The design variables are the number of lags m, the order of the model n, the coefficients in $A_p(s)$ and $B_p(s)$ and the variance of the measurement noise, i.e.

$$\eta = (m,\ n,\ a_1,\dots a_n,\ b_{r+1},\dots b_n,\ \lambda)$$

It is straightforward to handle correlated measurement noise by adding some extra states in (4.84).

In the limiting case when m tends to infinity the optimal accuracy becomes

$$\min E[s_c^r(kT)-\hat{s}_c^r(kT)]^2 = D(P-PQP)D^T \tag{4.87}$$

where

$$Q = (F-KC)^T Q(F-KC)+C^T(CPC^T+\lambda^2)^{-1}C$$

cf Anderson and Moore (1979).

In practice a finite number of lags are sufficient. When the measurement noise is small it might be only marginally better to use a smoothing estimate than a filtering estimate. For a high noise level it pays off to apply smoothing for several lags (for example m=5-20). Some numerical examples for the optimal accuracy as a function of the lags m (for standard state-estimation) can be found in Söderström (1985). An optimal filter would have to be time-varying to take care of the transient effects for small k optimally.

4.4.2 Differentiation viewed as an input estimation problem

The idea here is to consider the differentiation problem as a special case of input estimation (deconvolution). The signal to be differentiated is then regarded as the output from an integrator. The idea was developed in a state space formulation in Ahlén (1984). We will here apply the more recent ideas presented in Ahlén and Sternad (1985,1987) to differentiation. A polynomial approach is used to develop the optimal filter.

Consider a discrete-time linear stochastic system given by

$$y(k) = \frac{B(q^{-1})}{A(q^{-1})}\, u(k-d) + \frac{M(q^{-1})}{N(q^{-1})}\, v(k) \tag{4.88a}$$

where the unknown input sequence $u(k)$ is modelled as

$$u(k) = \frac{C(q^{-1})}{D(q^{-1})} e(k) \qquad Ev(k)^2/Ee(k)^2 = \rho \tag{4.88b}$$

The input signal and the measurement noise are described by independent ARMA processes. All polynomials except B are assumed to be monic. We further assume that CBN and MAD have no common factors with zeros on the unit circle, and that the system BC/AD is detectable. The degrees of the polynomials are na, nd etc. The white noise sequences $v(k)$ and $e(k)$ are stationary, zero mean and mutually uncorrelated.

The problem is to find a stable linear estimator of the input $u(k)$ based on noisy measurements $y(k)$

$$\hat{u}(k|k-m) = H(q^{-1})y(k-m) \tag{4.89}$$

where

$$H(q^{-1}) = \frac{Q(q^{-1})}{R(q^{-1})}$$

which minimizes the mean square estimation error

$$E(u(k)-\hat{u}(k|k-m))^2 \tag{4.90}$$

The point is that the system in our application should approximate a continuous-time integrator. An estimate of $u(k)$ can then be regarded as an estimate of the derivative. A simple approach is to use the transform (4.2) which gives

$$q^{-d} \frac{B}{A} = \frac{T}{1-q^{-1}} \tag{4.91}$$

The approximation (4.91) is exact for signals whose derivative is piecewise constant between the sampling instants.

Another idea is to use the bilinear approximation which gives

$$q^{-d} \frac{B}{A} = \frac{T}{2} \frac{1+q^{-1}}{1-q^{-1}} \tag{4.92}$$

These approximations lead to an insignificant systematic error if most of the signal energy is concentrated well below π/T. They are adequate for narrow and moderately wide band differentiation. For wide band differentiation other approximations should be used, cf section 4.2 for some alternatives.

The general problem is illustrated in Figure 4.3.

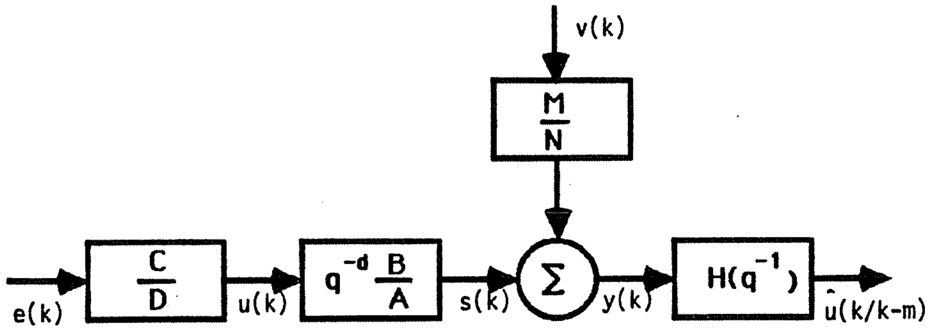


Figure 4.3. The input estimation or deconvolution problem.

Depending on the sign of m, we get an input prediction (m>0), filtering (m=0) or a fixed lag smoothing problem (m<0).

For the solution we need the spectral factorization (* indicates conjugate polynomials)

$$r\beta\beta_* = CBNC_*B_*N_*+\varrho MADM_*A_*D_* \tag{4.93}$$

where r is a positive scalar factor and $\beta(z)$ is a stable and monic polynomial in z with degree

$$n\beta = \begin{cases} nc+nb+nn & \text{if } \varrho=0 \\ \max\{nc+nb+nn, nm+na+nd\} & \text{if } \varrho>0 \end{cases}$$

This is the same spectral factorization that shows up when the output y(t) is described by an equivalent ARMA process. The scalar r can be interpreted as the variance ratio $\lambda_\varepsilon^2/\lambda_e^2$, where λ_ε^2 is the variance of the innovation sequence driving the equivalent ARMA model.

Theorem 4.3. Assume the system (4.88) to be detectable and that CBN and MAD have no common factors with zeros on the unit circle. The optimal input estimation filter (4.89) then attains the global minimum value of the estimation error (4.90) for the system (4.88) if

$$H(q^{-1}) = \frac{Q}{R} = \frac{Q_1NA}{\beta} \tag{4.94}$$

where β is the stable spectral factor from (4.93) and $Q_1(z^{-1})$ and $L(z)$ are the minimum degree solution of

$$z^{m+d}CBNC_* = r\beta Q_{1*}+D_*zL \tag{4.95}$$

with degrees

$$nQ_1 = \max\{nc-m-d, nd-1\}$$
$$nL = \max\{nc+nb+nn+m+d, n\beta\}-1 \tag{4.96}$$

Proof. See Ahlén and Sternad (1987).

After deciding a suitable integrator model (in this case the derivative is to be estimated) the design variables are the number of lags in the filter m, the ARMA model parameters for the input signal and the measurement noise and the quotient between the variances of the driving noises e(k) and v(k)

$$\eta = (m, c_1 \ldots c_{nc}, d_1 \ldots d_{dn}, m_1 \ldots m_{mn}, n_1 \ldots n_{nn}, \rho)$$

The input model need not to be stable. Inputs such as polynomial trends and sinusoids can be modelled by D-polynomials with zeros on the unit circle. If no specific knowledge of the input properties are available, a white noise model (C/D=1) is a natural choice.

To obtain an optimal estimate of the input signal, in this case the derivative, the parameters have to be known a priori or correctly estimated. When the parameters are unknown they have to be estimated in some way. An attempt to estimate the parameters of the input model is discussed in Ahlén (1984). An important question is under what conditions it is possible to uniquely determine the parameters of the input model and measurement noise model. The identifiability properties are analysed in Ahlén (1986). Note that in some cases it might be natural to estimate the ARMA models from a spectral analysis of the data.

The method described here can be seen as a simple time-domain algorithm for calculating discrete-time realizable Wiener filters for deconvolution or, in our case, differentiation. If the measurement noise is coloured and/or the lag m is large, the polynomial approach requires less calculations compared to a state space approach. Another nice property when using a polynomial approach is that the frequency characteristics can be studied directly. A drawback with the polynomial approach is that transient effects cannot be handled optimally.

In Section 5.2 a differentiating filter, based on the results presented here, will be derived.

5. NUMERICAL EXAMPLES

In this section numerical examples and comparisons are given for some of the methods described in the previous section.

In Section 5.1 differentiation of some stationary stochastic signals is studied. The methods used for differentiation are the regularized splines approach and the frequency domain method proposed by Usui and Amidror. A comparison with the optimal accuracy is also given.

In Section 5.2 the problem of tracking the derivative of a noisy trend signal with abrupt changes in the derivative is studied. Three on-line methods are analyzed and compared.

5.1 Differentiation of stationary signals

We will here illustrate two non-causal methods for differentiation of noisy data. The methods under study are the regularized spline approach described in Section 4.3.2 and Usui and Amidror's approach in Section 4.2.2. A comparison with the accuracy of the optimal Kalman filter (cf Section 4.4.1) will also be made. Some concluding remarks are given last in this subsection.

Description of the signals used to generate the data sets

A stochastic process is used for generating the signal $s_c(t)$. The chosen signals are all described by the spectral density

$$\phi_s(\omega) = |G(i\omega)|^2 \tag{5.1a}$$

where

$$G(s) = \frac{\omega_o^2}{s^2+2s\zeta\omega_o+\omega_o^2} \tag{5.1b}$$

The signal is sampled, cf (4.82), and white measurement noise with variance λ^2 is added to the signal.

The sampling interval T is 1s. Simulation of N=1000 data points for each example were carried out. The different examples discussed below are described in Table 5.1.

Example	1a	1b	2a	2b	2c	3a	3b
ω_o	0.2	0.2	0.8	0.8	0.8	0.8	0.8
ζ	0.1	0.1	0.1	0.1	0.1	1	1
λ^2	0.3^2	0.8^2	0.3^2	1^2	4^2	0.3^2	0.8^2
$Es(t)^2/\lambda^2$	5.6	0.8	22	2.0	0.12	2.2	0.3

Table 5.1. Description of the signal parameters.

The spectral densities are shown in Figure 5.1. In Example 1 the spectral density has a sharp resonance peak at $\omega=0.2$, in Example 2 the resonance peak is at $\omega=0.8$. Example 3 has a rather flat spectrum from $\omega=0$ to $\omega=0.8$.

In Figure 5.2 realizations of the signals in Table 5.1 are shown. The true derivatives in the figures are calculated from (4.81).

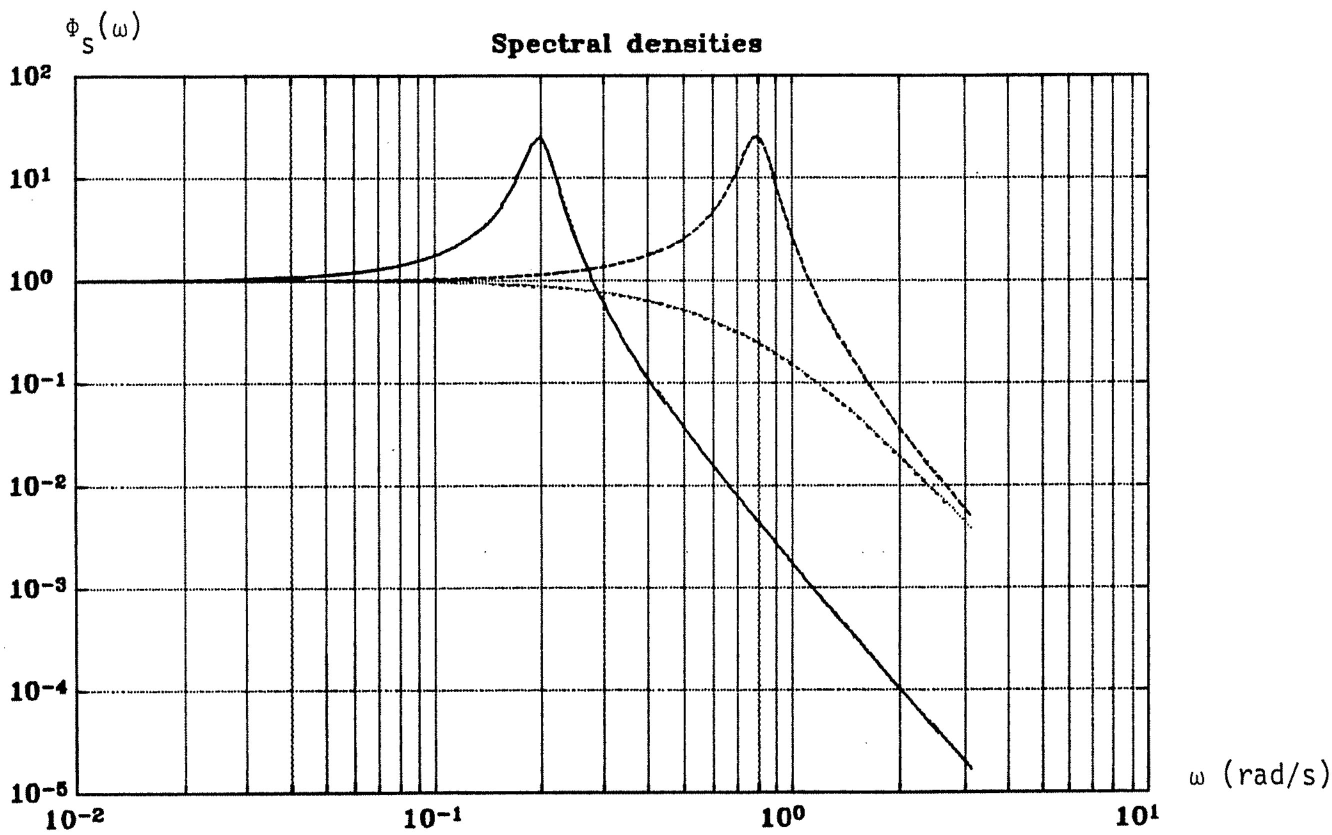


Figure 5.1. Spectral densities for the signals, $\omega_o=0.2$ $\zeta=0.1$ (——), $\omega_o=0.8$, $\zeta=0.1$ (---), $\omega_o=0.8$, $\zeta=1$ (···).

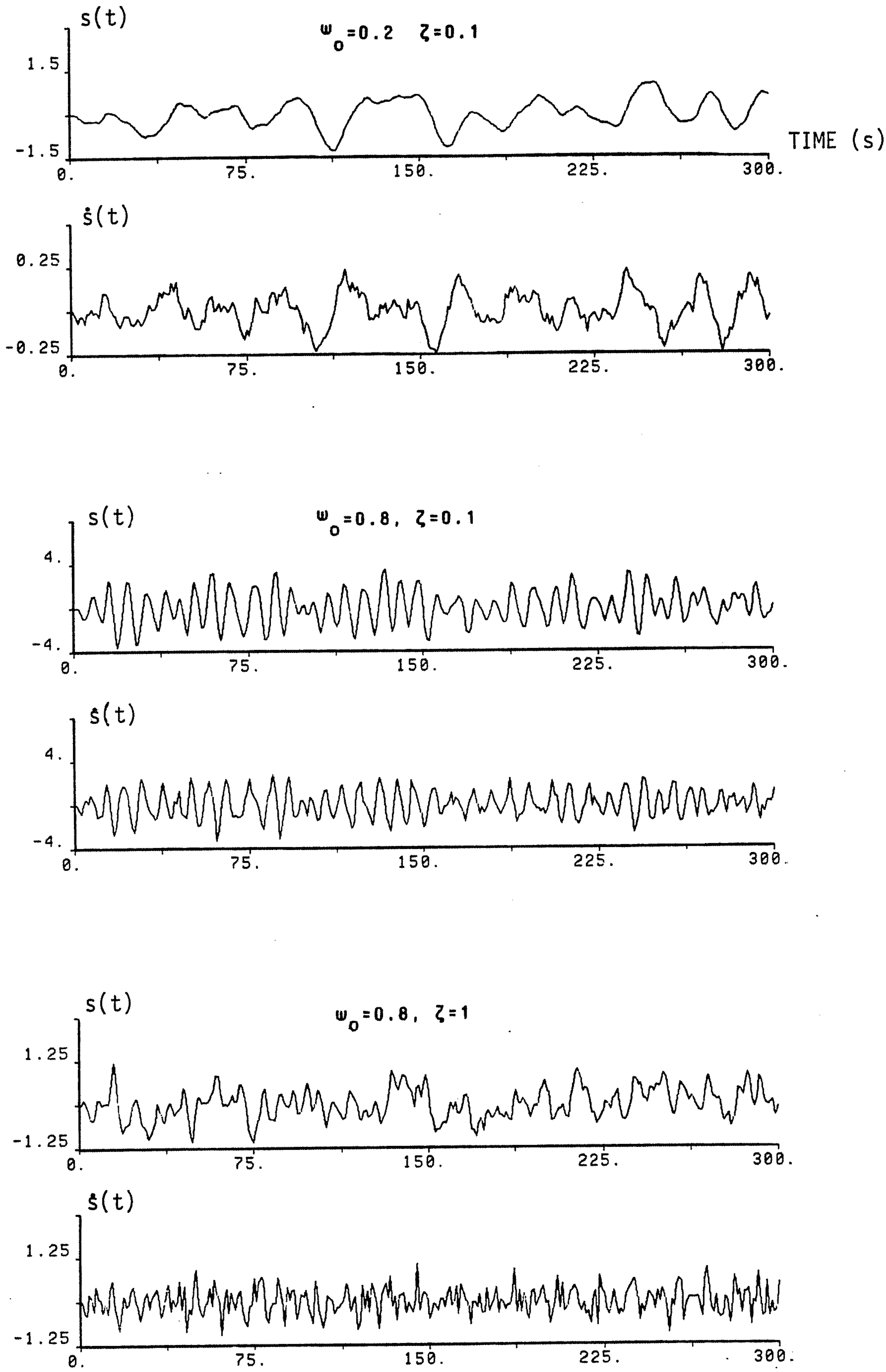


Figure 5.2. The signal s(t) and its derivative ṡ(t).

To evaluate the goodness of the estimated derivatives the following loss function is used

$$V = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} [\dot{s}_c(kT)-\hat{\dot{s}}(kT)]^2} \tag{5.2}$$

The optimal averaged accuracy $\min E[\dot{s}_c(kT)-\hat{\dot{s}}(kT)]^2$ is straightforward to evaluate using (4.87). We will denote

$$V_{opt} = \sqrt{\min E[\dot{s}_c(kT)-\hat{\dot{s}}(kT)]^2} \tag{5.3}$$

<u>Regularized splines applied to differentiation</u>

For convenience we very briefly summarize the method. The following criterion is used

$$\min_{s_s(t)} \left[\sum_{k=1}^{N} |s_s(kT)-y(k)|^2 + p \int_T^{NT} [s_s^m(t)]^2 dt \right] \tag{5.4}$$

The minimization results in a spline function of degree 2m-1. The user choices are m and the regularizing parameter p. Note that p=0 gives interpolating splines, which corresponds to wide-band differentiation, cf Example 4.3. On the other hand, when $p \to \infty$ a polynomial of degree m-1 is fitted to the data in a least squares sense which gives a narrow-band differentiation, cf Example 4.4.

The subroutines presented in Woltring (1986) have been used as a basis for an interactive program package to study regularized splines applied to differentiation.

The influence of the parameters m and p to the criterion V in (5.2) will be studied. A method for selecting p will also be illustrated. Craven and Wahba (1979) introduced the Generalized Cross-Validation Criterion (GCV) for selecting p. The idea of cross-validation is to fit the spline $s_s(kT)$ to all the data points, except the j:th, i.e.

$$\min_{s_s(t)} \sum_{\substack{k=1 \\ k \neq j}}^{N} |s_s(kT)-y(k)|^2 + p \int_T^{NT} [s_s^m(t)]^2 dt \tag{5.5}$$

Let $s_s^j(t)$ be the function which minimizes (5.5) and set

$$L(p) = \frac{1}{N} \sum_{k=1}^{N} w_p(k)(s_s^k(kT)-y(k))^2 \tag{5.6}$$

where $w_p(k)$ is a weighting function to compensate for nonperiodic sampling and nonperiodicity in the data. The GCV estimate of p is defined as the minimizer of L(p) and is denoted p_{GCV}.

Let the noisy measurements be given by

$$y(k) = s(k)+e(k)$$

where s(k) is the signal and e(k) is white noise. Craven and Wahba (1979) show that under mild conditions the cross-validated splines minimize the following true mean square error

$$R(p) = \frac{1}{N} \sum_{k=1}^{N} (s_s(kT)-s(k))^2 \tag{5.7}$$

as $N \to \infty$.

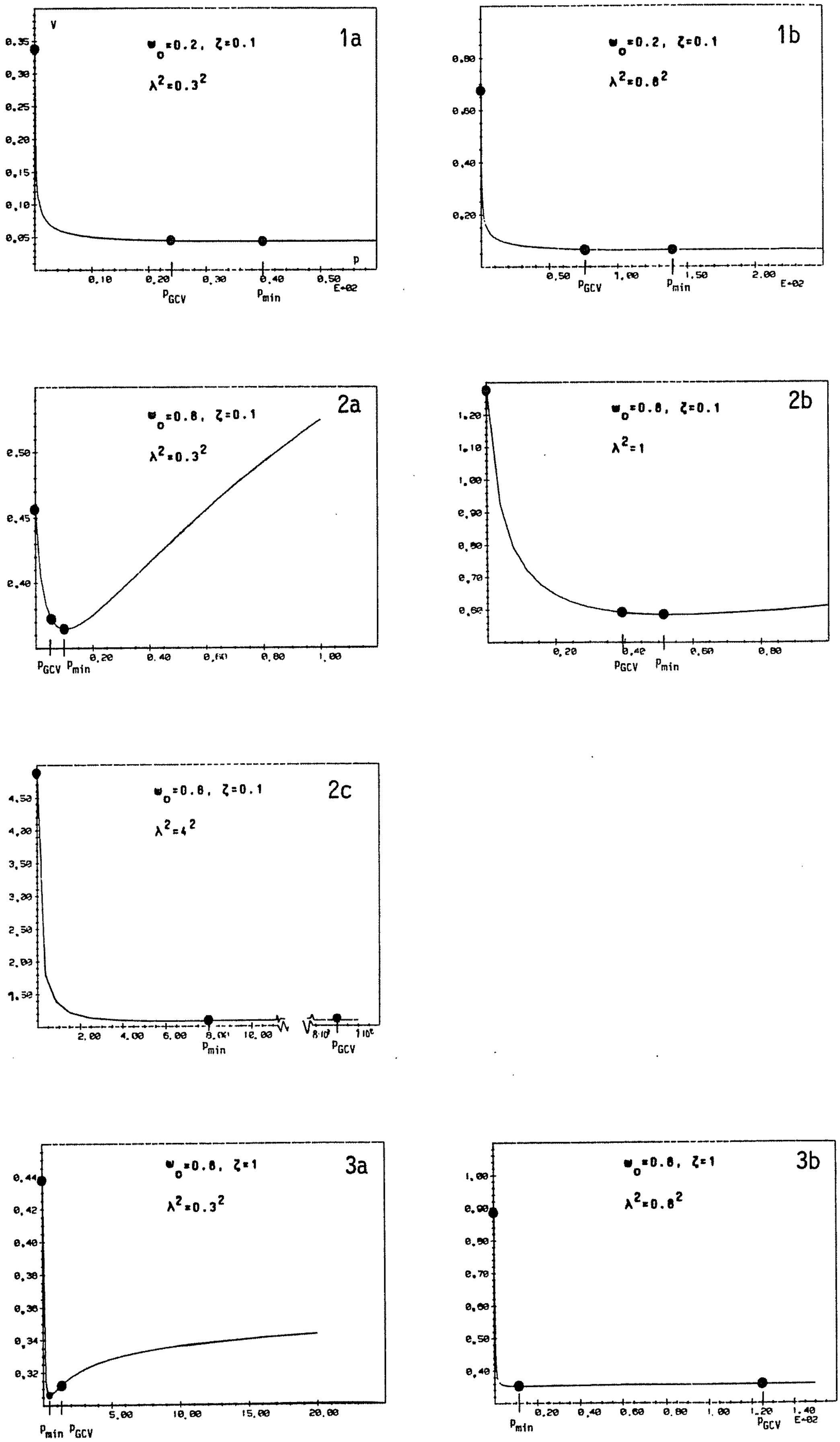
For a more detailed description of the GCV method applied to regularized splines, we refer to Craven and Wahba (1979), cf also Woltring (1986). The Fortran subroutines in Woltring (1986) also include the GCV criterion for selecting p.

The regularized splines were fitted to the data from the signals described in Table 5.1. The derivative was estimated by differentiating the splines. The numerical results are summarized in Table 5.2. Using the true derivative, the loss function in (5.2) was evaluated. In the table below, V_{GCV} is the value of the loss function in (5.2) when using cross-validated splines for estimating the derivative. The minimal value (in the sense of (5.2) using regularized splines) is denoted $V_{min}(RS)$ and is obtained by a numerical search. The corresponding value of p is denoted p_{min}. Note that the true derivative is used to evaluate V in (5.2). As a comparison, the calculated value of V_{opt} is also presented.

As mentioned earlier, the minimizing p was found by a numerical search. In Figure 5.3 the loss function V is plotted as a function of p for m=2. As seen in the figures the function V(p) is quite smooth around the minimum, particularly for the low frequency signals. A too low value on p (more interpolation) increases the stochastic error whereas a too high value will increase the systematic error.

Signal number	m	p_{GCV}	V_{GCV}	p_{min}	$V_{min}(RS)$	V_{opt}
1a $\omega_o=0.2$, $\zeta=0.1$ $\lambda^2=0.3^2$	1	20	0.069	8	0.057	
	2	24	0.044	40	0.043	0.042
	3	186	0.044	300	0.043	
1b $\omega_o=0.2$, $\zeta=0.1$ $\lambda^2=0.8^2$	1	60	0.114	19	0.079	
	2	78	0.065	140	0.062	0.060
	3	993	0.064	1700	0.063	
2a $\omega_o=0.8$, $\zeta=0.1$ $\lambda^2=0.3^2$	1	0	0.70	0.35	0.60	
	2	0.05	0.37	0.10	0.36	0.34
	3	0.05	0.36	0.05	0.36	
2b $\omega_o=0.8$, $\zeta=0.1$ $\lambda^2=1^2$	1	0.08	1.45	1.2	0.80	
	2	0.39	0.60	0.52	0.59	0.52
	3	0.42	0.56	0.41	0.56	
2c $\omega_o=0.8$, $\zeta=0.1$ $\lambda^2=4^2$	1	$5\cdot10^4$	1.11	11	1.06	
	2	$9\cdot10^7$	1.11	8	1.08	0.90
	3	$9\cdot10^{11}$	1.11	12	1.10	
3a $\omega_o=0.8$, $\zeta=1$ $\lambda^2=0.3^2$	1	0.52	0.35	1.7	0.32	
	2	12	0.34	0.50	0.31	0.30
	3	17	0.34	0.25	0.31	
3b $\omega_o=0.8$, $\zeta=1$ $\lambda^2=0.8^2$	1	12.4	0.353	9	0.352	
	2	125	0.356	12	0.353	0.34
	3	1200	0.358	50	0.358	

Table 5.2. Numerical evaluation of regularized splines applied to differentiation.



<u>Figure 5.3</u>. The loss function V as a function of the regularizing parameter p for m=2 (i.e. cubic regularized splines).

One important practical issue is how good the estimated derivative obtained via cross-validated splines is compared to the optimal accuracy. Comparing the loss V_{GCV} with V_{opt}, we see that for m>1 the cross-validated splines gives a value close to the optimal value of V. Note, however that when the noise variance increases so does the differances between the losses (cf Example 2a-2c).

For the signals under study we see that using GCV with m>1 produces a result where V_{GCV} is very close to V_{min}(RS). Note that in practice the true derivative is unknown and the minimizing p cannot be found.

As expected, a higher noise variance increases the optimal p-value, i.e. a "higher degree of regularization" is then needed. Increasing the order of m from 1 to 2 reduces the loss function in all examples, except for Example 2c where the noise level is very high. Note that the reduction is more significant in Example 2 than in 1. It pays off more to increase the order of m when the signal has the dominating spectrum at a higher frequency. Increasing m from 2 to 3 does not result in any significant improvement.

For the low frequency signal in Example 1 we see that the relative increase of p_{min} is larger when the variance is increased. We also notice that when the signal is dominated by low frequencies as in Example 1a and 1b, the optimal value of p is increasing as m increases. For the signal in Example 2a and 2b the opposite occurs due to the higher signal frequency.

Usui and Amidror's approach

The same data sets as previously described have also been applied to Usui and Amidror's (UA) approach. The approach is briefly summarized below. An antisymmetric FIR-filter

$$H(z) = \frac{1}{2T} \sum_{j=1}^{n} c_j (z^j - z^{-j}) \tag{5.8}$$

is used for approximating the following frequency response in a least mean-square sense

$$H_{LPI}(e^{i\omega T}) = \begin{cases} i\omega & |\omega| \leq \alpha\pi/T \\ 0 & \alpha\pi/T < |\omega| < \pi/T \end{cases} \qquad (5.9)$$

under the constraint of ideal differentiation at $\omega=0$.

The user choices are the filter order $2n+1$ and the cut-off frequency $\alpha\pi/T$ of the differentiation band. In this study we will use a fixed filter order of 17. The filter was running with data from the signals described in Table 5.1.

A minimizing α (in the sense of (5.2)) was found by a numerical search, cf Figure 5.4. In practice, the parameter α can be estimated from frequency analysis of the data. The numerical results are summarized in Table 5.3 (for a quick comparison $V_{min}(RS)$ and V_{opt} are also shown).

In all examples, the optimal α obviously decreases when the noise variance is increased. Comparing the results with the regularized splines we see that the latter method gives a slightly better result. The difference is more significant for low noise levels and/or low-frequency signal spectrum. It can, however, be conjectured that increasing the filter order will improve the results.

In Figure 5.4 the loss function V is plotted as a function of the differentiation band α. Note that due to the constraint at $\omega=0$, quite good results are obtained for $\alpha=0$ when the signal has a low frequency character or when the noise variance is high. For low values on α the systematic error is high whereas the stochastic error is small. The optimal α is found when the sum of these errors is minimal.

Figure 5.5 shows the amplitude response of the filter for some values on α. Note the oscillations due to the Gibbs phenomenon.

Signal number	α_{min}	$V_{min}(UA)$	$V_{min}(RS)$	V_{opt}
1a $\omega_o=0.2$, $\zeta=0.1$ $\lambda^2=0.3^2$	0.15	0.046	0.043	0.042
1b $\omega_o=0.2$, $\zeta=0.1$ $\lambda^2=0.8^2$	0.10	0.066	0.063	0.060
2a $\omega_o=0.8$, $\zeta=0.1$ $\lambda^2=0.3^2$	0.45	0.38	0.36	0.34
2b $\omega_o=0.8$, $\zeta=0.1$ $\lambda^2=1$	0.35	0.57	0.56	0.52
2c $\omega_o=0.8$, $\zeta=0.1$ $\lambda^2=4^2$	0.10	1.16	1.08	0.90
3a $\omega_o=0.8$, $\zeta=1$ $\lambda^2=0.3^2$	0.40	0.32	0.31	0.30
3b $\omega_o=0.8$, $\zeta=1$ $\lambda^2=0.8^2$	0.15	0.36	0.36	0.34

Table 5.3. Numerical evaluation of Usui and Amidror's approach. The filter order N=17.

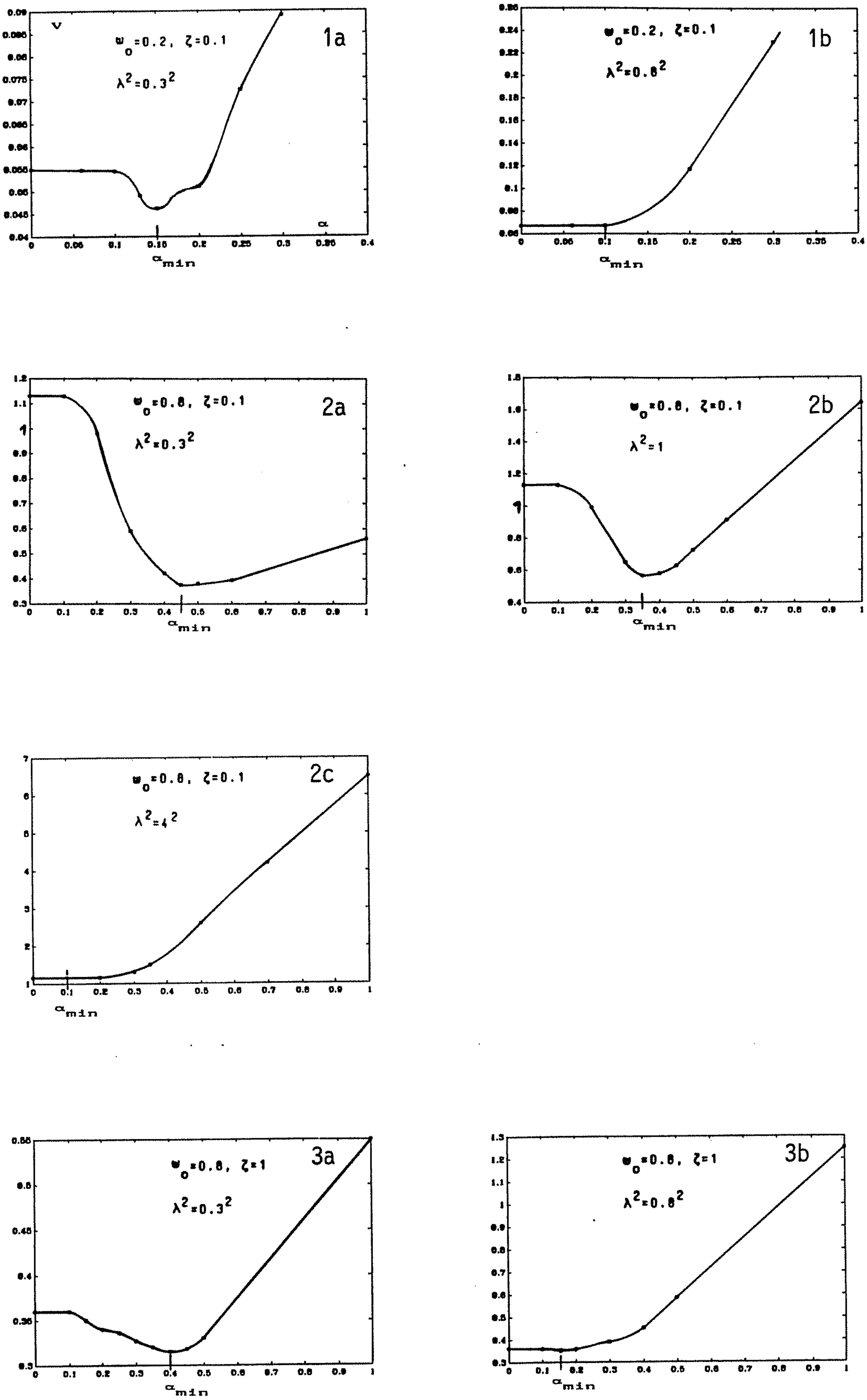


Figure 5.4. The loss function V as a function of the parameter α.

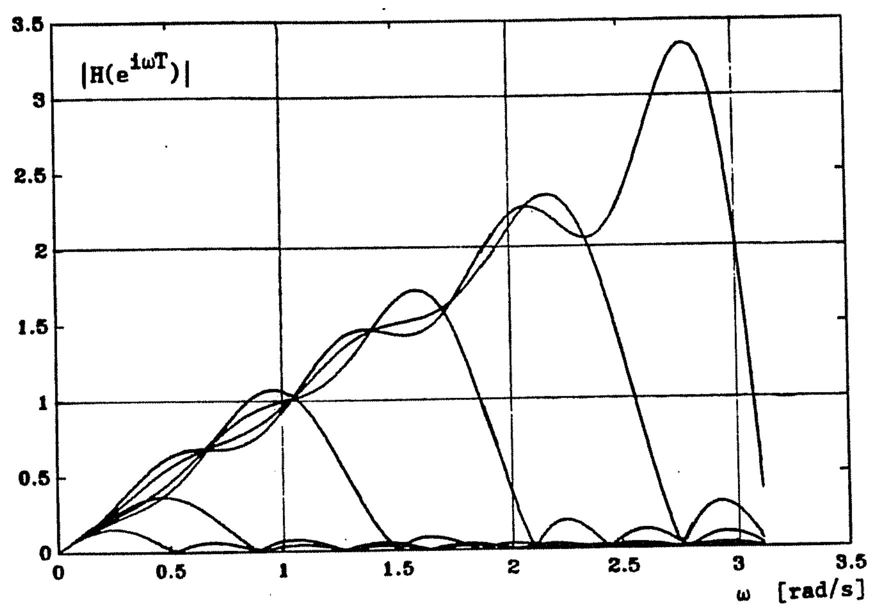


Figure 5.5. Amplitude responses for the UA-method, $\alpha=0$, 0.2, 0.4, 0.6, 0.8, 1.

Non-white measurements noise

We will shortly illustrate the effect of non-white measurement noise. Let the measurement noise be given by a second-order AR-process

$$v(k) = c_1 v(k-1)+c_2 v(k-2)+e(k) \tag{5.10}$$

where $e(k)$ is white noise with variance λ^2.

In the following example we have chosen $c_1=0.8$ and $c_2=-0.8$, which give poles at $0.4\pm i0.8$. The noise spectrum has a resonance at $\omega=1.1$. The variance of the driving noise is $\lambda^2=0.3^2$. The noise was added to the signal 1 in Table 5.1. The results are presented in Table 5.4.

signal characteristics	α_{min}	$V_{min}(UA)$	m	p_{GCV}	V_{GCV}	P_{min}	$V_{min}(RS)$	V_{opt}
$\omega_0=0.2$ $\zeta=0.1$ $c_1=0.8$ $c_2=-0.8$ $\lambda^2=0.3^2$	0.15	0.048	2	$1\cdot10^{-9}$	0.62	60	0.046	0.044

Table 5.4. The effect of non-white measurement noise.

Here the GCV criterion does not work very well. The explanation is that the method cannot distinguish between the true signal (with spectrum around $\omega=0.2$) and the noise (with spectrum around $\omega=1.1$).The cross validated splines are hence fitted to the signal as well as the noise. This gives a too low value on p (more interpolation) which increases the loss function. Note that the minimal loss, using regularized splines comes quite close to the optimal value.

The minimal loss using the UA method is slightly higher than $V_{min}(RS)$. A good value on α can however easily be estimated from a spectrum analysis of the data (if it is a priori known that the dominating noise spectrum is located at higher frequencies than the signal spectrum). Too obtain a good value on p in RS is far from trivial in this case (note that the GCV-method does not work properly for this case).

Some concluding remarks

Two methods for differentiating of noisy data have been illustrated and compared with the optimal accuracy.

The method based on regularized splines (RS) is a true off-line method since all data are used. Usui and Amidror's (UA) approach gives a non-causal filter which can be used on-line if a delay of half the filter order can be accepted.

Both methods have been used in biomechanical applications, cf Woltring (1986) and Usui and Amidror (1983).

The design parameters are very different. The parameter α in the UA approach has a very direct and intuitive meaning - the limit of the differentiation band. The choice of α should be based on frequency domain knowledge (or assumptions). This can be obtained by a frequency analysis of the signal (at least if the signal spectrum is of low-frequency character and the measurement noise is white).

The parameter p in RS is less intuitive, cf (5.4). If the signal and the noise spectrum is a priori known (or estimated), the Butterworth filter interpretation given in Woltring et al (1986) might be useful for determine a good value on p. The GCV-criterion has shown good results when the noise is white and the variance is not too high.

The computation time for determing the value of p_{GCV} in RS on a microVax II is about 33, 79 and 135 seconds for m=1,2 and 3 (N=1000).

The advice in Woltring (1986) to choose m higher than the highest derivative being sought is also found to be valid in this study. We have used stationary stochastic processes for generating the data. It is then natural to use the optimal Kalman filter for estimating the derivative. In practice it is often necessary to estimate the signal parameters, cf Ahlén (1986) for some examples. However, if the data cannot be transformed to a stationary process and a pertinent model structure cannot be found, the RS approach may be advantageous. If the spectrum of the signal is available, a frequency domain method such as the UA-method might be natural to use.

Needless to say, what method to select depends very much on the application and the a priori knowledge of the signal and noise. Note that by fixing the value on m (e.g. m=2) and using the GCV method in RS we have an automatic routine for estimating the derivative. That might be a great advantage for a non-experienced user.

5.2 Tracking of a noisy trend signal

A very simple assumption of the signal to be differentiated is that the true derivative is constant or changes slowly. A more realistic assumption in many cases is to assume that the true derivative is constant (or changes slowly) except for infrequent large changes. This leads to a compromise between the noise sensitivity in the filter and the trackability to step changes in the true derivative. Note that the same problem arises in the transient period even if the true derivative is constant. It may then be important to have a fast transient response from a poor initial guess of the true derivative.

Three discrete-time differentiating filters will be compared

(i) A second order Butterworth filter approach

(ii) Double exponential smoothing applied to differentiation

(iii) An input estimation approach

All three methods can be described by second order IIR-filters but the underlying design philosophies are quite different. Approach (i) is a frequency domain method based on analog filter design, the second approach is based on a polynomial signal model, whereas the third method is based on a stochastic discrete-time model.

A Butterworth filter approach

The basic idea here is to use a continuous-time differentiating filter which is transformed to a discrete-time filter, cf Section 4.2.1. Let the continuous-time filter be

$$H(s) = sH_o(s) \tag{5.11}$$

where $H_o(s)$ is a low-pass filter. In this example a second order Butterworth filter will be used

$$H_o(s) = \frac{\omega_o^2}{s^2+\sqrt{2}s\omega_o+\omega_o^2}$$

Using the bilinear transform gives

$$H(q^{-1}) = sH_o(s)\Big|_{s=\frac{2}{T}\frac{1-q^{-1}}{1+q^{-1}}} =$$

$$= \frac{2}{T}\,\frac{\Omega_o^2(1+q^{-1})(1-q^{-1})}{4+\sqrt{8}\Omega_o+\Omega_o^2+q^{-1}(2\Omega_o^2-8)+q^{-2}(4-\sqrt{8}\Omega_o+\Omega_o^2)} \tag{5.12}$$

where $\Omega_o=\omega_o T$.

The bilinear transform distors the frequency scale, so-called frequency warping. To achieve the desired discrete-time cut-off frequency, the frequency must be prewarped. Set

$$\Omega_o' = 2\tan\frac{\omega_o T}{2} \tag{5.13}$$

The differentiating filter then becomes

$$z(k) = H_{BW}(q^{-1};\omega_o)y(k) \tag{5.14}$$

where

$$H_{BW}(q^{-1};\omega_o) = \frac{2}{T}\,\frac{\Omega_o'^2(1+q^{-1})(1-q^{-1})}{4+\sqrt{8}\Omega_o'+\Omega_o'^2+q^{-1}(2\Omega_o'^2-8)+q^{-2}(4-\sqrt{8}\Omega_o'+\Omega_o'^2)}$$

Note that $H_{BW}(0)=0$, and $\lim_{z\to 1} H_{BW}(z)\frac{1}{z-1}=1/T$, i.e. the dc-gain is zero and a linear trend is differentiated without bias. The design parameter is the cut-off frequency ω_o, which via (5.13) determines the parameters in the filter.

Differentiation with double exponential smoothing

The parameters in a linear trend model can be estimated with double exponential smoothing, cf Example 4.5,

$$\varepsilon(k) = y(k)-\hat{a}_o(k-1)-\hat{a}_1(k-1) \tag{5.15a}$$

$$\hat{a}_o(k) = \hat{a}_o(k-1)+\hat{a}_1(k-1)+(1-\lambda^2)\varepsilon(k) \tag{5.15b}$$

$$\hat{a}_1(k) = \hat{a}_1(k-1)+(1-\lambda)^2\varepsilon(k) \tag{5.15c}$$

The derivative is estimated using

$$\hat{\dot{s}}_c(kT) = \frac{1}{T}\hat{a}_1(k) \tag{5.16}$$

which corresponds to the following differentiating filter, cf Example 4.6,

$$H_{DES}(q^{-1};\lambda) = \frac{(1-\lambda)^2}{T} \frac{1-q^{-1}}{(1-\lambda q^{-1})^2} \tag{5.17}$$

An input estimation approach

Adapting the ideas presented in Section 4.4.2 to the noisy trend signal lead to the model described in Figure 5.6

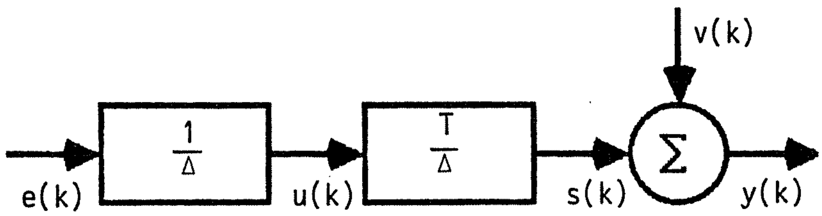


Figure 5.6. A discrete-time input estimation approach for tracking the derivative of a noisy trend signal with abrupt changes in the derivative.

In the figure $\Delta=1-q^{-1}$ and $v(k)$ is white noise with zero mean. Furthermore, let $e(k)$ be a stochastic sequence with the following property

$$e(k) = \begin{cases} +a & \text{with probability } p/2 \\ -a & \text{with probability } p/2 \\ 0 & \text{with probability } 1-p \end{cases} \tag{5.18a}$$

This gives

$$Ee(k_1)e(k_2) = a^2 p \delta_{k_1,k_2} \tag{5.18b}$$

An estimate of $u(k)$ can be regarded as an estimate of the derivative.

Comparing Figure 5.6 with the general model described in Figure 4.3 gives $A=1-q^{-1}$, $B=T$, $C=1$, $D=1-q^{-1}$, $M=1$ and $N=1$, which inserted in (4.93) gives the following spectral factorization

$$r\beta\beta_* = T^2+\varrho(1-2z+z^2)(1-2z^{-1}+z^{-2}) \tag{5.19}$$

where $\varrho=\dfrac{Ev^2(k)}{a^2p}$

Insertion in (4.95) gives for the filter case (m=0)

$$T = r(1+\beta_1 z+\beta_2 z^2)Q_1+(1-z^{-1})z(l_o+l_1 z) \qquad (5.20)$$

which gives

$$Q_1 = \beta(1) = (1+\beta_1+\beta_2)/T \qquad (5.21)$$

From (4.94) we now get the following differentiating filter

$$H_{IEA}(q^{-1};\varrho) = \frac{Q_1 NA}{\beta} = \frac{1+\beta_1+\beta_2}{T} \frac{1-q^{-1}}{1+\beta_1 q^{-1}+\beta_2 q^{-2}} \qquad (5.22)$$

The parameters in the filter (5.22) are determined by the design parameter ϱ via the spectral factorization (5.19). An optimal ϱ can be obtained if the amplitude (a), the frequency of the derivative changes (expressed through p) and the noise variance $E(v^2(k))$ can be correctly estimated.

Comparison between the settling time and the noise transmission

We will here study the settling time and the noise transmission of the filters.

We define the settling time k_s as the time required for the response curve to reach and stay within a range M about the final value when the true derivative makes a unit-step. The value M=10% is chosen.

The noise transmission W is defined as the variance of the filter output when the input is white noise with zero mean and unit variance

$$W = E(H(q^{-1})v(k))^2 = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |H(e^{i\omega T})|^2 d\omega \qquad (5.23)$$

The noise transmission W is straightforward to evaluate analytically, cf for example Åström and Wittenmark (1984). However, the settling time k_s must be determined by simulation (or solved numerically).

A crude relation between W and k_s is as follows

$$W \sim \left(\frac{1}{k_s}\right)^3 \qquad (5.24)$$

i.e. the noise transmission is approximately proportional to the cubed inverse of the settling time. To heuristically derive (5.24) we assume that the bandwidth ω_s of the differentiation band is proportional to the inverse settling time (cf for example the classical rules of thumb in control theory)

$$k_s \sim \frac{1}{\omega_s} \tag{5.25}$$

We further assume that the filter has an ideal frequency response up to ω_s. The noise transmission (5.23) can then be written

$$W = \frac{I}{\pi} \int_0^{\omega_s} \omega^2 d\omega + \int_{\omega_s}^{\pi/T} |H(e^{i\omega T})|^2 d\omega \tag{5.26}$$

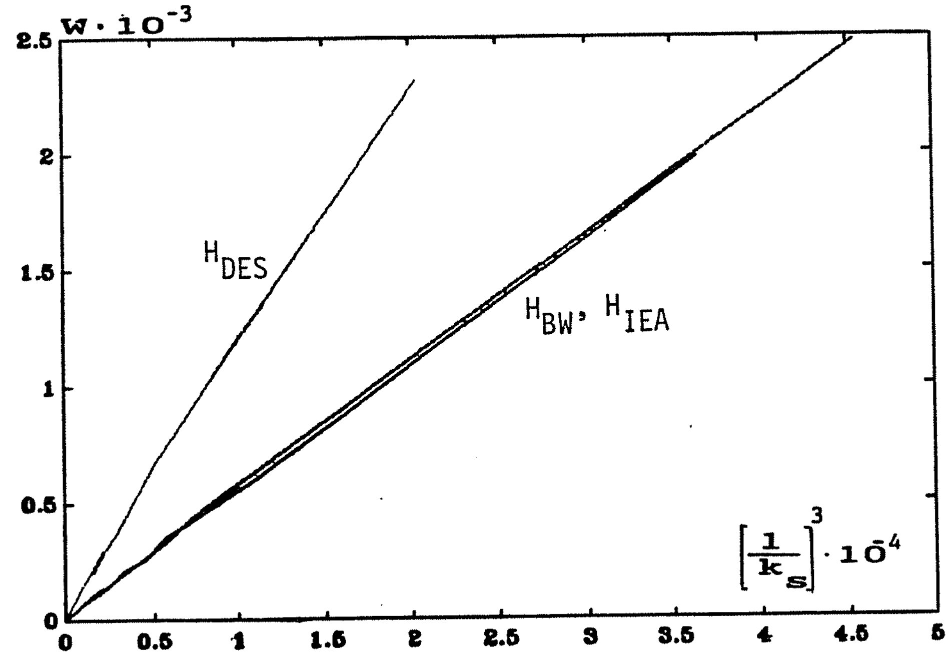
Using (5.25) we see that (5.24) approximately holds if the second part in (5.26) is small.

If we assume that the signal is strictly band-limited with bandwidth ω_s and that the filter is constrained to give no systematic error, the expression (5.26) can also be used to obtain a bound for the achievable accuracy of the estimated derivative, cf Lanshammar (1982).

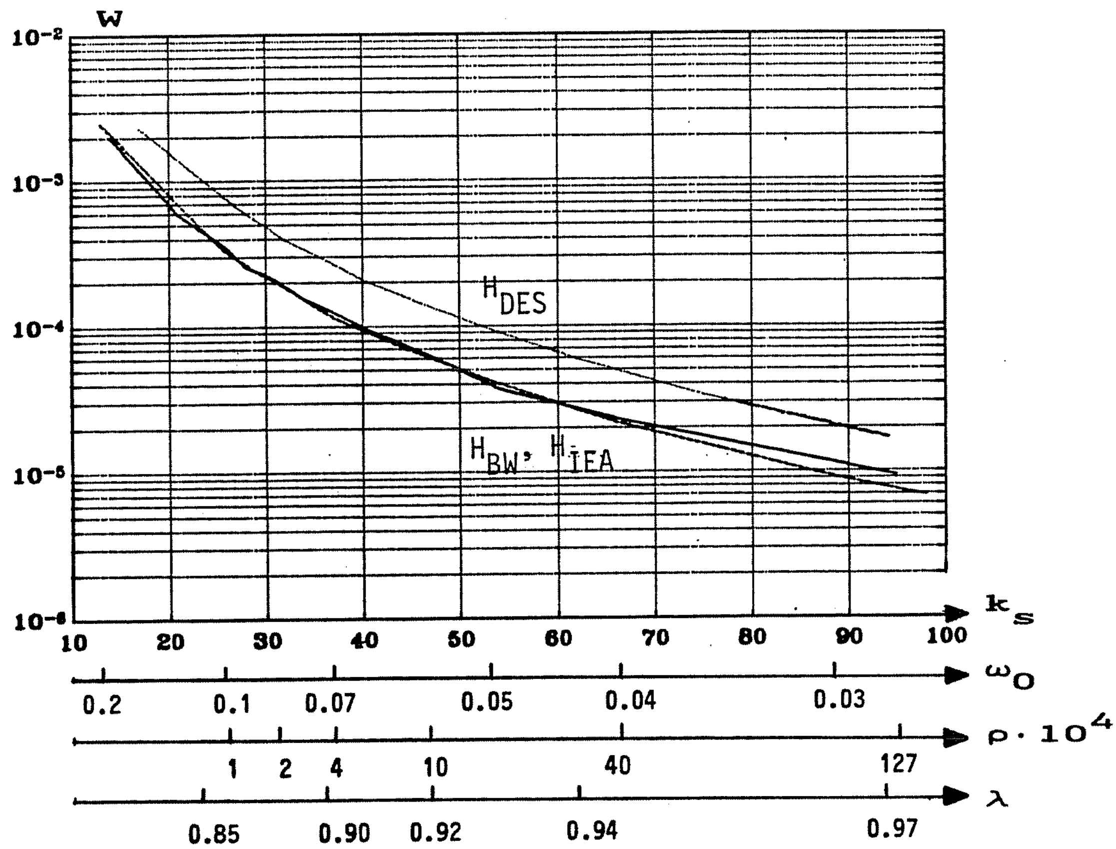
In Figure 5.7a the noise transmission is plotted as a function of the cubed inversed settling time. We see that for the three filters under study, the agreement with (5.24) is quite good.

Figure 5.7b shows the noise transmission W as a function of the settling time k_s, and the filter parameters ω_0, ϱ and λ. For a fixed value of W or k_s, the filter parameters are directly given. The figure might be useful for a quick determination of how requirements in the noise transmission affect the trackability to step changes in the true derivative (and vice versa).

From Figure 5.7b it is seen that the two filters H_{BW} and H_{IEA} produce practically the same result. Note, however, that the numerator in H_{BW} is one degree higher than H_{IEA} and H_{DES}. The filter H_{DES} produces a poorer behaviour than the two others for this value of the settling time parameter M (10%). A slight advantage with H_{DES} is that the filter output gives no overshoot in contrast to the two others which both give an overshoot of about 5% (cf Figure 5.8 below).



<u>Figure 5.7a</u>. Noise transmission as a function of the cubed inversed settling time.



<u>Figure 5.7b</u>. Noise transmission as a function of the settling time and the filters design parameters.

Reducing the value on M will reduce the difference between the methods. This is illustrated in the following examples (all filters are tuned to have the same W=0.0071).

Filter	k_s (M=10%)	k_s (M=1%)	η
H_{BW}	9	22	ω_o=0.29
H_{IEA}	9	22	ρ=182
H_{DES}	12	21	λ=0.74

Table 5.5. Settling times for M=10% and M=1%.

As seen from the table, reducing M from 10% to 1% gives an opposite ranking of the filters.

In Figure 5.8 the step responses for the three filters are shown. All filters have W=0.0071. The two filters H_{BW} and H_{IEA} have practically the same response.

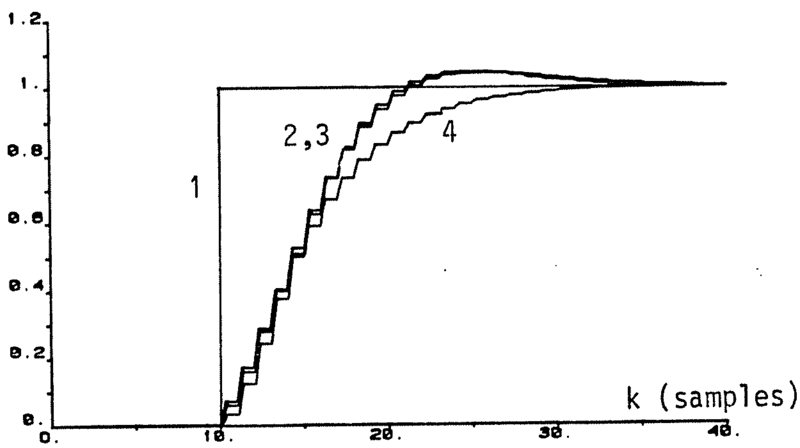


Figure 5.8. Step responses for the filters. The true derivative (1), the filter H_{BW} (2), H_{IEA} (3) and H_{DES} (4).

The compromise between the noise sensitivity and the trackability to step changes might be hard to trade-off. A natural solution is to (if possible) decrease the sampling interval T. Assume that a change in the sampling interval not affects the noise variance. It follows then directly from (5.25) and (5.26) that $k_s \sim T^{1/3}$ for a constant W. When comparing the behaviour for different sampling intervals, it is natural to measure the settling time in continuous-time. We then get

$$t_s \sim T^{4/3} \tag{5.27}$$

where $t_s = k_s T$.

To further improve the trackability it may be fruitful to use adaptive filter techniques. One approach is to use a detector for significant changes in the derivative. Different types of detectors applied to differentiating filters are studied in Carlsson (1987).

Poles of the filters

The poles of the three filters as a function of the design parameters are shown in Figure 5.9.

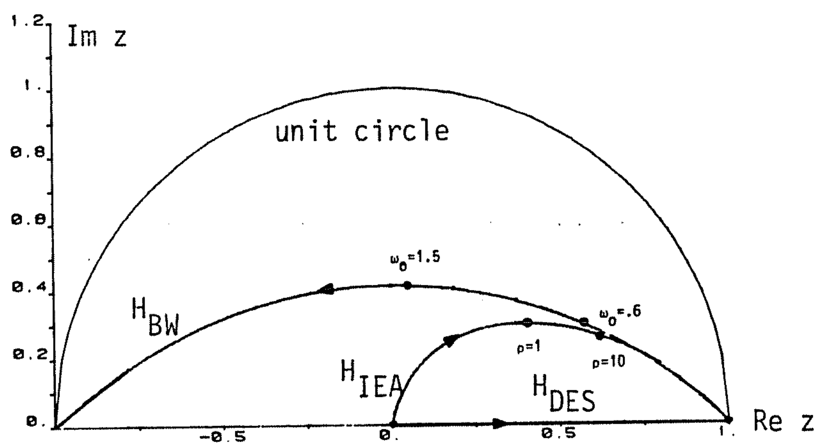


Figure 5.9. Poles of the filters as a function of the design parameters.

The filter $H_{DES}(z;\lambda)$ has a real double pole at $z=\lambda$ $(0\leq\lambda<1)$. The filters $H_{BW}(z,\omega_o)$ and $H_{IEA}(z,\varrho)$ have complex poles (which causes a slight overshoot). The poles approximately coincide for low values on ω_o and high values on ϱ (say, $\omega_o<0.1$ and $\varrho>20$). In the asymptotic case, $\varrho\to 0$ and $\lambda\to 0$ the filters H_{IEA} and H_{DES} tend to a first-order difference filter.

Frequency responses

In Figure 5.10 the amplitude and phase of the frequency responses of the filters are shown. All filters are tuned to have the same noise transmission $W=0.0071$. This corresponds to a settling time $k_s=9$ for H_{BW} and H_{IEA} and $k_s=12$ for H_{DES}. Due to the zero at $z=-1$, the filter H_{BW} has slightly better damping for high frequencies. This filter also have the largest differentiation band. At $\omega=0$ the filters have an ideal phase responses. The phase error rapidly increases for higher frequencies.

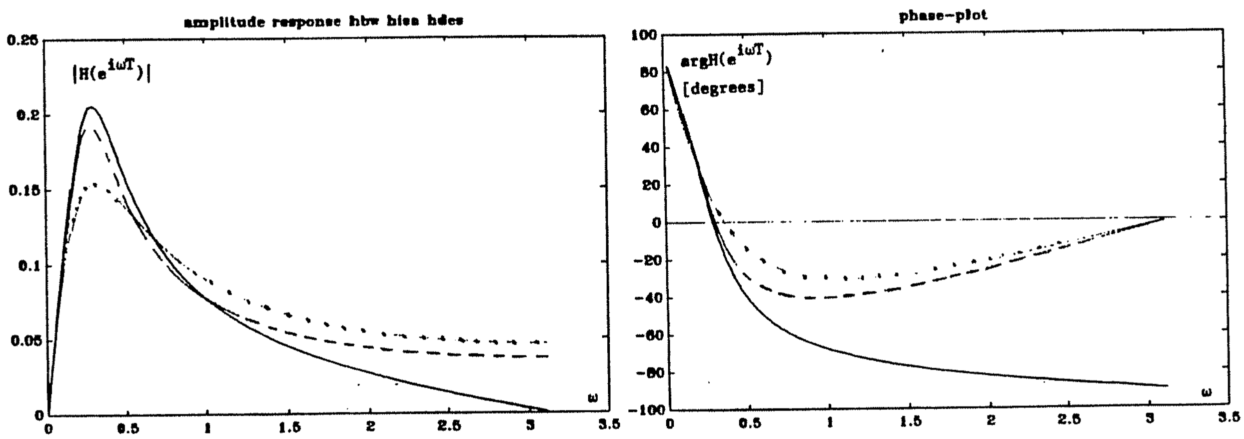


Figure 5.10. Magnitude and phase of the frequency responses, H_{BW} (---), H_{IEA} (---), H_{DES}(···).

Comparisons

The table below summarizes the relative merits of the three filters. The conclusions are based on the results from the analysis and simulation experiments described above.

Filter	H_{BW}	H_{IEA}	H_{DES}	Remarks
k_s(M=10%)	1	1	2	
k_s (M=1%)	2	2	1	small differences
overshoot	2	2	1	
filter order	2	1	1	
damping for high frequencies	1	2	3	

Table 5.6. The relative merits of the filters (ranking in order of performance).

What filter to select naturally depends on how the different properties above are weighted. However, the simple H_{BW} filter seems to be a good candidate for "the best choice".

CONCLUSIONS

The design and analysis of discrete-time differentiating filters have been investigated. Both frequency domain and time domain methods were handled. Many factors must in practice be taken into account when choosing an appropriate filter. Some ways to characterize the properties of differentiating filters were discussed

A number of methods for designing differentiating filters were surveyed. Some relations and comparisons between different approaches were also given.

In a simulation study some methods were illustrated and compared. The regularized spline approach with generalized cross-validation have shown good results for a number of stochastic signals when the noise was white and the variance not too high. Comparisons with the optimal accuracy and a frequency domain method were also given. For high noise levels and non-white measurement noise the derivative obtained from cross-validated splines shown a significant detoriation compared with the optimal differentiating Kalman filter.

The problem to track step changes in the true derivative has been studied. Three different approaches were suggested and analysed. An approximative relation between the settling time to step changes and the noise sensitivity was given.

ACKNOWLEDGEMENTS

We are very grateful to Mr Karl-Gustav Wallén for competent programming help. We also thank Miss Ylva Johansson for fast and skilful typing.

REFERENCES

Abraham B and J Ledolter (1983)
Statistical Methods for Forecasting
John Wiley and Sons

Ahlén A (1984)
An input estimation approach to differentiation of noisy data
Report UPTEC 8450R, Department of Technology, Uppsala University, Uppsala, Sweden

Ahlén A and M Sternad (1985)
Optimal input estimation: A polynomial approach
Report UPTEC 85103R, Department of Technology, Uppsala University, Uppsala, Sweden

Ahlén A and M Sternad (1987)
Optimal deconvolution based on polynomial methods
Submitted for publication

Ahlén A (1986)
Identifiability of an input estimation problem
Report UPTEC 8614R, Department of Technology, Uppsala University, Uppsala, Sweden

Allum J H J (1975)
A least mean squares CUBIC algorithm for on-line differential of sampled analog signals
IEEE Trans Computers, vol C-24, no 6, pp 585-590

Anderssen R S and P Bloomfield (1974a)
Numerical differentiation procedures for non-exact data
Numerische Mathematik, vol 22, pp 157-182

Anderssen R S and P Bloomfield (1974b)
A time series approach to numerical differentiation
Technometrics, vol 16, no 1, pp 69-75

Anderson B D O and J B Moore (1979)
Optimal Filtering
Prentice-Hall, Englewood Cliffs.

Antoniou A and C Charalambous (1981)
Improved design procedure for equiripple digital differentiators
IEEE Int Symposium on Circuits and Systems, Chicago, USA, vol 2, pp 540-545

Åström K J (1970)
Introduction to Stochastic Control Theory
Academic Press, New York

Åström K J and B Wittenmark (1984)
Computer Controlled Systems
Prentice-Hall, Englewood Cliffs.

de Boor C (1978)
A Practical Guide to Splines
Springer Verlag, New York

Brown R G (1962)
Smoothing, Forecasting and Prediction of Discrete Time Series
Prentice-Hall, Englewood Cliffs.

Carlsson B (1987)
Double exponential smoothing with adaptive forgetting applied to differentiation
Report UPTEC 8715R, Department of Technology, Uppsala University, Uppsala, Sweden

Conte S and C de Boor (1972)
Elementary Numerical Analysis
McGraw-Hill, New York

Craven P and G Wahba (1979)
Smoothing noisy data with spline functions
Numerische Mathematik, vol 31, pp 377-403

Gradshteyn I S and I M Ryzhik (1980)
Table of Integrals, Series and Products
Academic Press, New York

Gustafsson L and H Lanshammar (1977)
ENOCH - An integrated system for measurement and analysis of human gait
Report UPTEC 7723R, Department of Technology, Uppsala University, Uppsala, Sweden

Harris F J (1978)
On the use of windows for harmonic analysis with the discrete Fourier transform
Proc IEE, vol 66, no 1, pp 51-83

Hofstetter E, A Oppenheim and J Siegel (1971)
A new technique for the design of non-recursive digital filters
Proc 5th Annu Princeton Conf Inform Sci and Systems, pp 64-72

Kaiser J F (1966)
Digital Filters
In Systems Analysis by Digital Computer, F F Kuo and J F Kaiser, eds, John Wiley & Sons Inc, pp 218-285

Kailath T (1980)
Linear Systems
Prentice Hall, Englewood Cliffs

Koenig D M (1966)
Numerical differentiation of experimental data
Chemical Engineering, sep, pp 186-189

Lanczos C (1956)
Applied Analysis
Prentice-Hall, Englewood Cliffs.

Lanshammar H (1982)
On precision limits for derivatives numerically calculated from noisy data
J Biomechanics, vol 15, no 6, pp 459-470

Lavery J E (1979)
The perils of differentiating engineering data numerically
Chemical Engineering, jan, pp 121-123

Ljung L and T Glad (1984)
Velocity estimation from irregular noisy position measurements
Proc IFAC World Congress, Budapest, vol VI, pp 275-279

Ljung L and T Söderström (1983)
Theory and Practice of Recursive Identification
MIT Press, Cambridge, Mass.

McClellan J H and T W Parks (1973)
A unified approach to the design of optimum FIR linear-phase digital filters
IEEE Trans Circuit Theory, vol CT-20 (6), pp 697-701

McClellan J H, T W Parks and L R Rabiner (1973)
A computer program for designing optimum FIR linear phase digital filters
IEEE Trans Audio Electroacoustics, vol AU-21, pp 506-526

Morgera S D (1982)
Digital filtering and prediction for communication system time synchronization
Canadian Comm & Energy Conf, Montreal, Canada, 13-15 Oct, pp 351-356

Morozov V A (1984)
Methods for Solving Incorrectly Posed Problems
Springer Verlag, New York

Oppenheim A V and R W Schafer (1975)
Digital Signal Processing
Prentice-Hall, Englewood Cliffs.

Peterka (1984)
Predictor based self-tuning control
Automatica, vol 20, no 1, pp 39-50

Rahenkamp C A and B V K Vijaya Kumar (1986)
Modifications to the McClellan, Parks, and Rabiner computer program for designing higher order differentiating FIR filters
IEEE Trans Acousstics, Speech and Signal Processing, vol ASSP-34, no 6, pp 1671-1674

Rabiner L R, B Gold and C A McGonegal (1970)
An approach to the approximation problem for nonrecursive digital filters
IEEE Trans Audio Electroacoustics, vol AU-18, pp 83-106

Rabiner L R and K Steiglitz (1970)
The design of wide-band recursive and nonrecursive digital differentiators
IEEE Trans Audio Electroacoustics, vol AU-18, pp 204-209

Rabiner L R, J H McClellan and T W Parks (1975)
FIR digital filter design techniques using weighted Chebyshev Approximation
Proc IEEE, vol 63, no 4, pp 595-610

Söderström T (1980)
On the design of digital differentiating filters
Report UPTEC 8017R, Department of Technology, Uppsala University, Uppsala, Sweden

Söderström T (1982)
Estimation of time derivative from noisy data
Report UPTEC 8290R, Department of Technology, Uppsala University, Uppsala, Sweden

Söderström T (1985)
Some optimization problems for stochastic systems
Report UPTEC 8530K, Department of Technology, Uppsala University, Uppsala, Sweden

Spriet J and J Bens (1979)
Optimal design and comparison of wide-band digital on-line differentiators
IEEE Trans Acoustics, Speech and Signal Processing, Vol ASSP-27, no 1, pp 46-52

Steiglitz K (1970)
Computer-aided design of recursive digital filters
IEEE Trans Audio Electroacoustics, vol AU-18, pp 123-129

Usui S and I Amidror (1982a)
Digital low-pass differentiation for biological signal processing
IEEE Trans on Biomedical Engineering, vol BME-29, no 10, pp 686-693

Usui S and I Amidror (1982b)
A new relationship between the Lanczos and the MMSE algorithms for low-pass differentiation
The Trans of the IECE of Japan, vol 65, no 3, pp 175-176

Wahba G and S Wold (1975)
Periodic splines for spectral density estimation; the use of cross-validation for determining the degree of smoothing
Communication in Statistics, vol 4, pp 125-141

Woltring H J (1985)
On optimal smoothing and derivative estimation from noisy displacement data in Biomechanics
Human Movement Science, vol 4, pp 229-245

Woltring H J (1986)
A Fortran package for generalized, cross-validatory spline smoothing and differentiation
Adv Eng Software, vol 8, no 2, pp 102-113

Woltring H J, A de Lange, J M G Kauer and R Huiskes (1986)
Instantaneous helical axis estimation via natural, cross-validated splines
5th ESB Congress, Berlin (West), Sept 8-10.

Young P (1984)
Recursive Estimation and Time-Series Analysis
Springer-Verlag, Berlin