



Murdoch
UNIVERSITY

MURDOCH RESEARCH REPOSITORY

<http://dx.doi.org/10.1049/ip-vis:19971613>

Rasiah, A.I., Togneri, R. and Attikiouzel, Y. (1997) Modelling 1-D signals using Hermite basis functions. IEE Proceedings - Vision, Image, and Signal Processing, 144 (6). pp. 345-354.

<http://researchrepository.murdoch.edu.au/18660/>

Copyright © 1997 IEE

Modelling 1-D signals using Hermite basis functions

A.I. Rasiah
R. Togneri
Y. Attikouzel

Indexing terms: Hermite function, QRS detection, ECG analysis, Pattern recognition, Gaussian quadratures, LMS estimation

Abstract: The paper discusses a method for estimating the Hermite coefficients of a discrete-time one-dimensional signal. To estimate the Hermite coefficients a solution based on Gaussian quadratures is used. The paper also looks at various least mean squared (LMS) estimation methods to estimate two further parameters which are incorporated into the orthonormal Hermite basis function; a spread term and a shift term. In addition, the effects of scaling, dilation and translates of a signal on its Hermite coefficients, spread and shift terms are presented. The paper concludes with a brief discussion on the potential application of the Hermite parameters as features for use in problems requiring shape discrimination within a one-dimensional signal. It also mentions those applications where this was found to be appropriate.

1 Introduction

The objective of any feature extraction process is to obtain a pattern space that retains sufficient information, has low dimensionality and provides good inter-class separation to enhance discrimination between the various feature classes.

A typical feature extraction process involves the parameterisation of a system in terms of a mathematical model and use of its parameters as features. In this paper, the model adopted for 1-D signal representation is a series expansion of N basis functions, where the features are the coefficients of these functions, and the pattern space is a space in \mathbb{R}^N . Using this model, comparisons between the original signal and its synthesis (defined as the linear combination of basis functions) can easily be made. This provides a means to demonstrate the retention of relevant information by the features and, a way to determine the sufficiency of the number of features used.

© IEE, 1997

IEE Proceedings online no. 19971613

Paper first received 17th July 1996 and in final revised form 25th September 1997

A.I. Rasiah was with the University of Western Australia and is now with Motorola Australia Software Centre, Adelaide, Australia

R. Togneri and Y. Attikouzel are with the Centre for Intelligent Information Processing Systems, Department of Electrical and Electronic Engineering, University of Western Australia, Nedlands, WA 6907, Australia

This paper concerns itself solely with the orthonormal basis functions of the Hermite family. It outlines a deterministic method based on Gaussian quadratures for estimating the Hermite coefficients of a one-dimensional signal. In addition, two further parameters are included into the definition of the orthonormal Hermite function. They are a spread parameter and a shift parameter.

To estimate these parameters, least mean squared (LMS) estimation is used. Various gradient descent techniques were evaluated to determine a technique which would provide the fastest convergence to a solution: results are presented in this paper. The paper also examines the problem of reducing false minima in the cost function of the spread term and includes some visualisation of the search space landscape of these parameters.

The choice of the Hermite family was made for two reasons. First, the Hermite family of basis functions are orthogonal. This means that there is no redundancy between features and the assumption of independence among them can be made. Also, the Hermite polynomial has the following desirable properties:

- (i) easily computable via a three-term recurrence relation as given by the Christoffel–Darboux formula [1].
- (ii) completeness: this means that any signal can be represented to an arbitrarily high degree of accuracy by taking the number of terms in its series expansion, the truncation, to be sufficiently large.

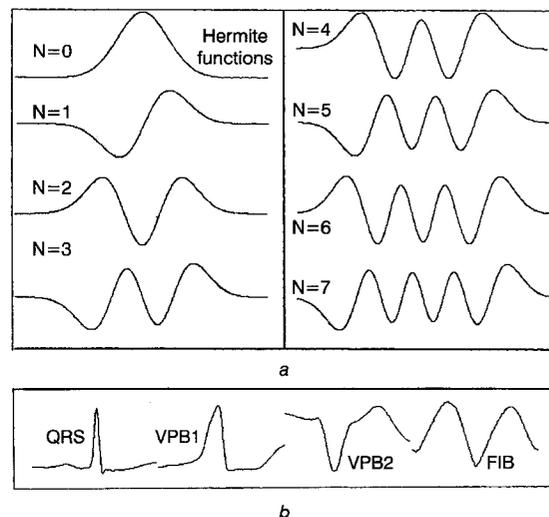


Fig. 1 *Hermite functions*
a Plots of the first eight basis functions belonging to the Hermite family of functions;
b Example of a set of 'shapes' to be modelled by Hermite functions for a specific application; namely QRS/ectopic beat detection

The second reason for selecting the Hermite function set was motivated by the form of the functions themselves. Comparing the form of the Hermite functions with typical QRS/ectopic beats, i.e. electrocardiogram (ECG) arrhythmia beats (see Figs. 1a and b), it seems reasonable to suppose that a series expansion of Hermite functions may be well suited to provide a concise representation of these signals [2–6]. In short, if a signal has a form which shares some resemblance with the form of the basis functions, then its series expansion is likely to have few terms in it, i.e. the ‘spectra’ has a ‘narrowband’ characteristic. This also means the dimensionality of the feature space is minimised.

Note that the method described in this paper has been applied to the problem of detecting and classifying QRS and ectopic beats within an ECG trace [2, 3]. Consequently, much of its description lies within the context of this application. Although the method outlined is focused on ECG waveform analysis, its suitability can be extended to signals, e.g. $u(t)$, that suffer simple scale and translation distortions as modelled by $\alpha u(\beta t - \tau)$. This paper concentrates its discussion on the methodology for determining and using Hermite coefficients. The main contribution of the paper is Section 2 where the Hermite basis is introduced and its use detailed: how to estimate the spectral coefficients, the spread factor, and the shift factor; and the effect of signal scaling, translation and dilation on the spectral coefficients. Also suggestions on estimating the number of terms to use in the series expansion are presented in the context of a specific application.

2 Methodology

2.1 Hermite basis set

The n th-order Hermite function is a weighted function of the n th-order Hermite polynomial. The Hermite polynomials are recurrence relations, and so too are their derivatives [1, 7]:

$$H_n(t) = (-1)^n \exp(t^2) \frac{d^n}{dt^n} (\exp(-t^2)) \quad (1)$$

$$H_{n+1}(t) = 2tH_n(t) - 2nH_{n-1}(t) \quad \text{where } H_0(t) = 1 \quad (2)$$

To minimise precision errors caused during the computation of factorials in the Hermite function normalisation constant, this paper uses the normalised form of the Hermite polynomials (signified by the $\hat{\cdot}$ notation). Also, a table-lookup approach is used to replace precomputable functions such as \sqrt{n} .

$$\hat{H}_0(t) = \frac{1}{\sqrt{\sqrt{\pi}}}$$

$$\hat{H}_{n+1}(t) = t\sqrt{\frac{2}{n+1}}\hat{H}_n(t) - \sqrt{\frac{n}{n+1}}\hat{H}_{n-1}(t) \quad (3)$$

$$\int_{-\infty}^{\infty} \exp(-t^2)\hat{H}_m(t)\hat{H}_n(t)dt$$

$$= \int_{-\infty}^{\infty} h_m(t)h_n(t)t \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases} \quad (4)$$

Therefore, $\{h_n(t)\}$ represents the orthonormal basis set of Hermite functions as shown in Fig. 1a. By replacing

t with its scaled version $(t - \tau)/\sigma$, a generalised form of the n th-order orthonormal Hermite function, $h_{n,\sigma}(t, \tau)$ is defined, where σ is the ‘spread’ of the function and τ is the ‘shift’ of the function.

$$h_{n,\sigma}(t, \tau) = \frac{1}{\sqrt{s^n n! \sigma \sqrt{\pi}}} \exp\left(-\frac{(t - \tau)^2}{2\sigma^2}\right) H_n\left(\frac{t - \tau}{\sigma}\right)$$

$$= \frac{1}{\sqrt{\sigma}} \exp\left(-\frac{(t - \tau)^2}{2\sigma^2}\right) \hat{H}_n\left(\frac{t - \tau}{\sigma}\right) \quad (5)$$

2.2 Evaluating the Hermite coefficients

To determine the Hermite coefficients of signal $u(t)$, the following integral is evaluated:

$$a_n(\sigma, \tau) = \int_{-\infty}^{\infty} u(t)h_{n,\sigma}(t, \tau)dt \quad (6)$$

Substituting $x = (t - \tau)/(\sigma\sqrt{2})$ into eqn. 6 we obtain

$$a_n(\sigma, \tau) = \int_{-\infty}^{\infty} \sqrt{2\sigma}u(x\sigma\sqrt{2} + \tau)\hat{H}_n(x\sqrt{2})\exp(-x^2)dx \quad (7)$$

To evaluate this integral, the Gauss–Jacobi integration theorem is used [1, 8]

$$\int_a^b f(x)\rho(x)dx = \sum_{i=0}^{M-1} w_i f(x_i) + \frac{M!\sqrt{\pi}}{2^M(2M)!} f^{(2M)}(\xi)$$

$$(-\infty < \xi < \infty) \quad (8)$$

where the set of weights w_i is known.

If $\rho(x)$ is defined as an exponential weighting function, $\exp(-x^2)$, then the optimum grid points are the roots of the $(M + 1)$ th Hermite function (i.e. the M th-order Hermite function). If these points are examined, we find that they are unevenly spaced, i.e. the outermost points are further apart than the points near $t = 0$. To compute the roots of H_M and its corresponding weights, the algorithm by Secrest and Stroud is used [9], p.154.

The first task is to exploit the form of eqn. 8 by rewriting the integrand of eqn. 7 as the product of two functions, $f(x)$ and $\rho(x)$, where $\rho(x)$ is defined as an exponential weighting function:

$$\rho(x) = \exp(-x^2) \quad (9)$$

$$f_{n,\sigma}(x, \tau) = \sqrt{2\sigma}\hat{H}_n(x\sqrt{2})u(x\sigma\sqrt{2} + \tau) \quad (10)$$

If we assume $u(x)$ is a polynomial such that $f_{n,\sigma}(x, \tau)$ is also a polynomial of at most degree $(2M - 1)$, then the spectral coefficients are

$$a_n(\sigma, \tau) = \sum_{i=0}^{M-1} w_i f_{n,\sigma}(x_i, \tau) \quad (11)$$

If no such assumption is made, then the equality of eqn. 11 becomes an approximation, i.e. the spectral coefficients are estimates.

2.3 Estimating the Hermite spectral coefficients for a discrete time signal

The previous subsection has shown how the spectral coefficients for the Hermite basis set can be determined for continuous-time signals. Since most signals are represented in a time sampled form, it would be useful if

eqn. 11 could be used accordingly. Unfortunately, there is no integer-based quadrature technique since the manner in which the spectral coefficients are determined makes it necessary to know the value of the signal, i.e. $u(x)$, at the grid points; which have noninteger values and thus lie between the sample points of the signal. To get around this problem, interpolation is done; e.g. zero-order holding or linear interpolation. Experimental results on the signals in Fig. 1b have shown improvements in the fit (i.e. squared error) between a signal and its synthesised form when linear interpolation is used instead of zero-order holding. Note that because of interpolation, the equality in eqn. 11 is replaced by approximation.

If the grid values (or roots) of H_M are examined, one finds M values $\{x_0, x_1, \dots, x_{M-1}\}$, such that $x_0 > x_1 > \dots > x_{M-1}$. Note that

- (i) for M odd; $f_{n,\sigma}(0, \tau) = 0$ for n odd; $x_{M/2} = 0$ and $x_{M/2+1+i} = -x_i$ where $i \in \{0, 1, \dots, M/2 - 1\}$.
- (ii) for M even; $x_{M/2+i} = -x_i$ where $i \in \{0, 1, \dots, M/2 - 1\}$.

With this in mind, eqn. 11 is rewritten as

$$a_n(\sigma, \tau) = \sum_{i=0}^{M/2-1} w_i [f_{n,\sigma}(x_i, \tau) + f_{n,\sigma}(-x_i, \tau)]$$

$$n = 0, 1, \dots, N - 1 \quad (12)$$

However, the following expression is used in the case where M is odd and n even:

$$a_n(\sigma, \tau) = w_{M/2} f_{n,\sigma}(0, \tau) + \sum_{i=0}^{M/2-1} w_i [f_{n,\sigma}(x_i, \tau) + f_{n,\sigma}(-x_i, \tau)]$$

$$(13)$$

Also from eqn. 10, and given that $u(t)$ is a continuous-time form of a discrete time signal of finite duration, i.e. $(2L + 1)$ samples, the range of values for t for which $u(t) = 0$ is, $|t| < L$ or equivalently $-(\tau + L)/(\sigma\sqrt{2}) > x > (L - \tau)/(\sigma\sqrt{2})$; given the substitution used in eqn. 7. Referring to eqns. 12 and 13 we need not evaluate the entire summation over the interval $i = 0$ to $(M/2 - 1)$. Instead the minimum value of k is found such that $x_k \leq (\text{abs}(\tau) + L)/(\sigma\sqrt{2})$. Then the summation interval of eqns. 12 and 13 becomes $i = k$ to $(M/2 - 1)$. Fig. 2 demonstrates the modelling of a typical discrete-time, finite duration signal for various values of N where N is the number of Hermite coefficients. It also shows that the Hermite series representation is not par-

ticularly efficient when modelling a typical discrete-time, finite duration 'wideband' signal, i.e. a signal which has both low and high frequency components. However, in the following Section 2.4.4, an approach is considered to selectively model the relevant component within such signals.

2.4 Estimating the spread/shift value

In the previous subsection, a procedure for computing the Hermite coefficients was outlined, where the spread and shift term have fixed real values. This Section investigates methods to estimate the spread value, where such methods can be similarly applied to the estimation of the shift value. Since the theory for choosing an optimum spread is still incomplete, LMS techniques are used instead. After each estimate of σ , the coefficients are recalculated and the iterations continue until a solution is reached. Note that if the spread value is chosen poorly then a large number of series expansion terms are required to produce a satisfactory fit. Also, this relationship between σ and N is a significant one as emphasised in [10]. In this subsection various optimisation techniques [11] are considered to estimate the spread for a fixed N . They, however, do not guarantee the spread to be optimal, i.e. finding the global minimum of the cost function.

Consider the discrete-time signal (interpolated) $u(t)$, and its synthesised form, $u_{N,\sigma}(t)$. The error-sum (cost) function, $E_N(\sigma)$, is defined as the squared error between these two signals. To minimise this error, two well-known methods, namely the step-search approach and a gradient descent approach have been tried; see the Appendix for details.

2.4.1 Step-search approach: The step-search approach to selecting σ that minimises the error sum involves computing the error sum at various values of σ , within the range $\sigma_{min} \leq \sigma \leq \sigma_{max}$, as σ moves in constant $\Delta\sigma$ increments; and choosing the value of σ with the least error sum. This method requires that the search range of σ be known *a priori* together with its step size. Although this approach is computationally expensive, especially if the step size is small, the solution it provides is a 'ball-park' estimate of the global minima.

2.4.2 Gradient descent approach: This approach is computationally less expensive than the step-search method. Two methods that adopt this approach, i.e. steepest-descent and Newton's method, are presented below; see the Appendix for equations.

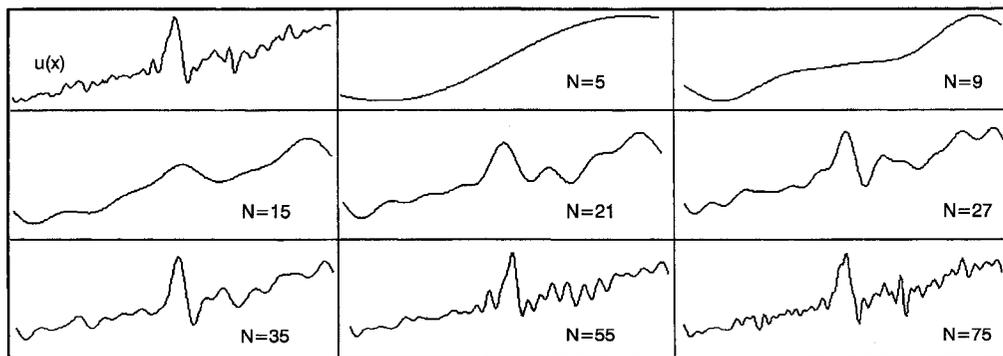


Fig. 2 Modelling a typical signal, $u(x)$ comprised of a high-frequency (spike) and a low-frequency component (linear trend) as a series of N orthonormal Hermite basis functions

Top left panel: signal $u(x)$ (216 samples). Remaining panels show the synthesised signal $u_{N,\sigma}(x)$ for various values of N .

Steepest descent method. The steepest descent method involves adjusting the value of σ after each iteration as given by eqn. 14; see also eqns. 30, 31 and 32 in Section 7.1. See [4, 5] for a similar proposition. After each adjustment of σ , the coefficients are recomputed with this new value of σ using the method outlined in earlier sections:

$$\sigma_{k+1} = \sigma_k - \lambda \nabla_k \quad (14)$$

where $\lambda \geq 0$ and ∇_k is the error gradient

Newton's method. Newton's method is similar to steepest descent. However, unlike steepest descent, there is no step-size parameter λ . Instead, the second derivative is used and the adjustment of σ is given by eqn. 15; see also eqns. 33, 34 and 35. This method converges rapidly but is sensitive to the initial value of σ . It works especially well for problems where the 'error surface' is quadratic in form.

$$\sigma_{k+1} = \sigma_k - \frac{\nabla_k}{\nabla_{kk}} \quad (15)$$

where ∇_{kk} is the derivative of the error gradient.

2.4.3 Combined approach: Given the strengths and weaknesses of the step-search and gradient descent methods, it seems only appropriate to use both approaches to increase the likelihood of finding an optimum solution for the spread. The combined method would involve the following stages.

Stage 1. Step-search method. Here a sparse step-search is made to minimise the computational expense and provide a 'ball-park' estimate of the global minimum.

Stage 2. Steepest descent method. Since the initial value provided by stage 1 is only an estimate and could well be far from the solution, the steepest descent method is employed to provide a better estimate. It is preferred over Newton's method because it is far better behaved in the presence of a poor initial estimate.

Stage 3. Newton's method. Newton's method is used in the final stage because of its fast convergence. Also, it is assumed that the initial spread value provided by stage 2 is a good enough estimate to guarantee the nonchaotic behaviour of this final stage.

The maximum number of iterations is fixed when implementing such an algorithm in a time constrained application. This is because the maximum computation time allocated for the algorithm to complete its task has to be known *a priori* to guarantee known behaviour. Therefore, gradient descent algorithms usually have as their stopping criterion the following conditions: (i) $\nabla_k \approx 0$, or (ii) $k > \text{MAX_ITERATIONS}$; whichever ever is satisfied first.

2.4.4 Problems with local minima: To illustrate problems inherent in selecting an optimum σ when using gradient descent techniques, consider the following example. Note that this example seeks to demonstrate how the presence of more than one local minima can cause methods like steepest descent to fail; and how, through 'preprocessing' the signal, the 'landscape' of the cost function can be changed to favour the intended minima.

To begin, a test case is generated, i.e. signal $s(x)$ as defined by eqn. 16, which contains a high-frequency component (spike) and a low-frequency component (linear trend). Next $s(x)$ and $w(x)$ are multiplied

together to create a signal $u(x) = w(x).s(x)$ that provides more than one dominant local minima, i.e. local minima with comparable error sum values. This is done to provide a scenario that emphasises the difficulties in picking the 'right' minima.

$$s(x) = \begin{cases} \frac{x}{L} + 0.92 \exp\left(-\frac{x^2}{2}\right) & -L \leq x \leq L \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

$$w(x) = \begin{cases} 1.0 & 0 \leq |x| \leq \frac{L}{2} \\ \frac{1}{2} + \frac{1}{2} \cos\left[\pi \frac{2x-L}{L}\right] & \frac{L}{2} \leq |x| \leq L \end{cases} \quad (17)$$

Figs. 3-5 demonstrate that selection of the value of the spread σ_1 at the first 'local minimum' isolates the 'high-frequency' signal in $s(x)$, i.e. the spike. Alternatively, by selecting the next spread value, σ_2 , at the next 'local minimum', a signal of 'lower frequency' i.e. the linear trend component, is isolated. The relation between the spread and frequency is an intuitive one, i.e. the smaller the spread value, the higher the frequency. The spread parameter is also a useful feature for measuring the 'width' of a signal's shape. Note that this particular example typifies the problem of detecting QRS complexes in ECG signals with severe baseline wander.

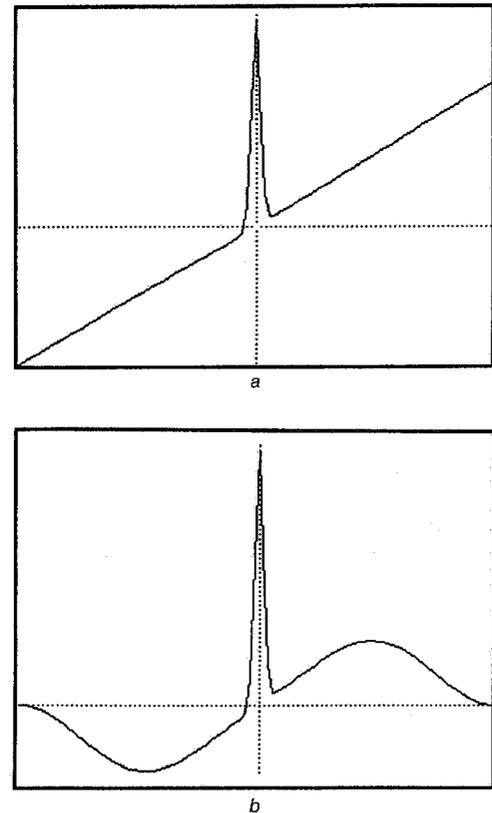


Fig. 3 Example of more than one 'optimum'
a Typical signal, $s(x)$, with a high-frequency spike and a low-frequency linear trend
b Signal windowed, $u(x) = s(x).w(x)$, to provide comparable local minima

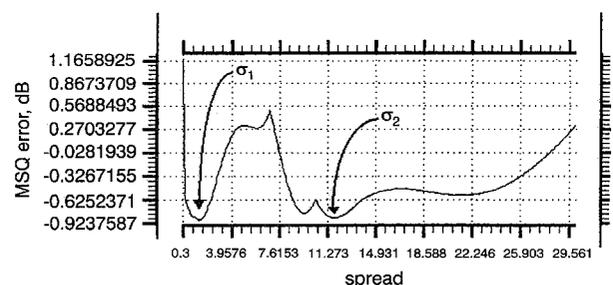


Fig. 4 Error sum for various values of σ for the signal from Fig. 3b

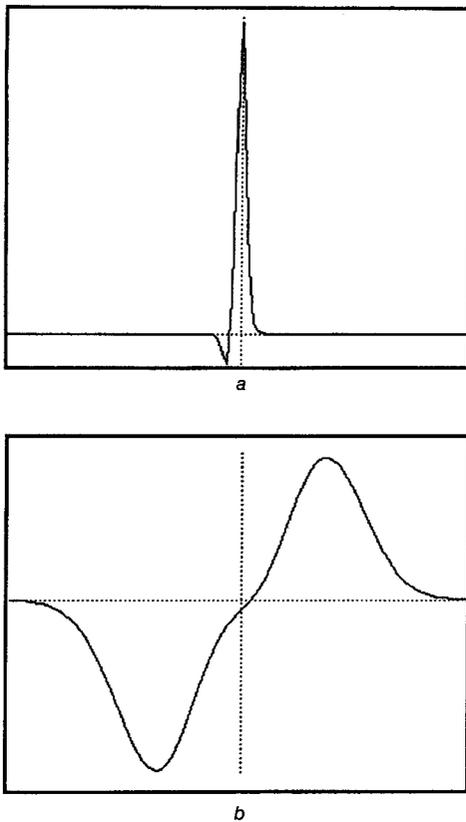


Fig. 5 Signal reconstructions ($N = 5$) based on Fig. 4
a High-frequency spike, $\mu_{N,\sigma_1}(x)$, reconstructed by selecting the lower of the two minima, σ_1
b Low-frequency signal, $\mu_{N,\sigma_2}(x)$, reconstructed by selecting the higher of the two minima, σ_2

Returning to the problem of detecting a beat in an ECG, consider the scenario where the ECG waveform is processed as a series of running snapshots of partial ECG waveforms as viewed through a sliding window that traverses the full ECG waveform in time. Clearly only when the beat to be recognised occupies a central position in the window, is the beat considered detected. This means that the signal information at the centre of the window is more relevant than that at the ends. In other words the fit of the signal segment located in the central portion of the window should be better than the segments at the ends. To reflect this in the overall fit, the error contributed by the samples at the end of the window have less weight than those in the middle; hence the error term is redefined as

biased error sum,

$$E_N(\sigma) = \sum_{x=-L}^L \exp\left(-\frac{x^2}{2m}\right) (u(x) - u_{N,\sigma}(x))^2 \quad (18)$$

where m is equal to the sample width of the signals, i.e. $\approx 0.6L$.

From Fig. 6*a* note that the biased error sum alters the cost function 'landscape' such that only a single minima is left, thus making the selection of an optimum spread simpler.

The previous attempt to remove the undesirable local minima works only for signals whose relevant information all lie within the regions of the window of similar size. Therefore if there are signals with vastly different widths (i.e. frequencies), erroneous local minima are usually the minima with the largest spread values and can be attributed to some linear trend component within the signal. As an alternative to using a biased

error sum, the unbiased error sum is used. It is defined in terms of $u(x)$ whose linear trend component has been removed. To carry out this removal, a straight line is fitted through $u(x)$ using linear regression, and then $u(x)$ is redefined by removing this linear component within it. Fig. 6*b* shows that the error sum after removal of the linear trend effectively creates a single minima, thus making the selection of an optimum spread simpler.

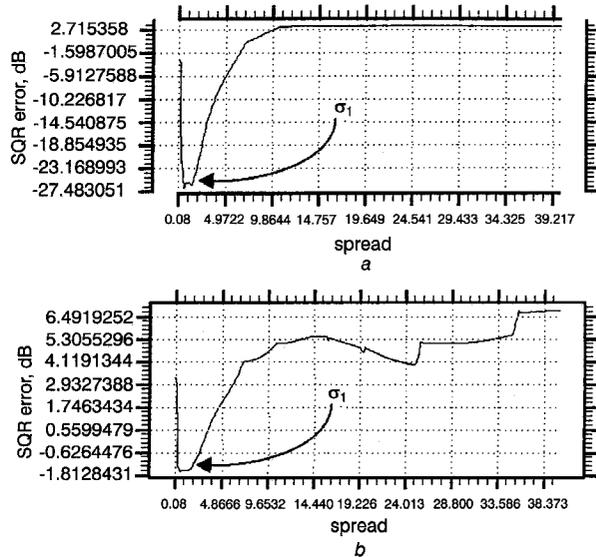


Fig. 6 Error sum
a Biased
b Unbiased, but with linear trend removed from $u(x)$
 $N = 5$ in all cases

2.5 Overview of results

To determine a suitable combined estimation approach, experiments were carried out using five test signals. These were signals QRS, FIB, VPB1 and VPB2 of Fig. 1*b*, and a signal with arbitrarily chosen Hermite parameters as shown in Fig. 7*a*. They were represented as sampled signals of 200 samples each (i.e. $-100 \leq t \leq 100$) and taken from the synthesised form of the signal as defined by their Hermite parameters.

Using the estimation techniques outlined earlier, various permutations of the combined approach were tried; see the second column of Table 1. In the first stage of this approach, a step-search is adopted to estimate a starting point for use by the subsequent gradient descent stage. This is done by imposing a uniform grid of σ and τ values across a fixed range of values and computing the error sum at every point in the grid. Then the (σ, τ) pair with the smallest error sum is chosen. Note that for certain applications, this grid need not be uniform, but random or user defined to minimise the number of grid points searched. Also, the Hermite coefficients are recalculated for every point in the grid and after every adjustment to σ and τ during the gradient descent optimisation process. Therefore, if a signal $s(t)$ has an error-sum minimum that lies at $(\sigma, 0)$, then the error-sum minimum of the signal $s(t - \tau)$ will lie at (σ, τ) , see Figs. 7*a* and *b*.

Table 1 summarises the results of the experiments using signal Fig. 7*a* and is typical. If the combined approaches (as given by their row number) are rated from the least to the most number of iterations taken to reach the solution; we obtain {13, 7, 14, 8, 12, 6} for signal Fig. 7*a*, {11, 14, 9, 12, 8, 10, 6, 5} for QRS, {11, 7, 13/14, 12, 5, 9, 8} for FIB, {11, 14, 7, 13, 5, 6/8/10/

Table 1: Results of various optimisation procedures for estimating the Hermite coefficients, σ and τ of a signal (200 samples) as defined by the first row's parameters

Row	Description	a_0	a_1	a_2	a_3	a_4	σ	τ	Iterations	Sqr. Err.
0	actual parameters	0.638990	-0.215757	0.115461	0.250657	-0.430597	5.30	33.0	n.a.	n.a.
1	fixed σ and τ	0.637912	-0.212890	0.112609	0.245640	-0.421328	5.30	33.0	-	0.000129
2	SS ₀	-0.118669	0.354021	0.472938	0.401858	0.255709	4.0	20.0	100	0.123111
3	SS ₁	0.697764	0.069484	-0.054234	0.414469	-0.001563	4.0	30.0	400	0.043237
4	SS ₂	0.678336	0.049505	-0.164955	0.433701	-0.012664	5.0	30.0	1600	0.030190
5	SS ₀ \Rightarrow SD	0.040127	0.572624	0.496171	0.190828	0.008323	4.83565	21.205	100+1000	0.100142
6	SS ₁ \Rightarrow SD	0.639133	-0.211001	0.114017	0.252562	0.415916	5.26824	32.9406	400+342	0.000195
7	SS ₀ \Rightarrow SDm	0.634383	-0.209197	0.104241	0.252602	-0.419617	5.31852	32.9591	100+46	0.000187
8	SS ₁ \Rightarrow SDm	0.639247	-0.209536	0.104993	0.251952	-0.419788	5.31694	32.9629	400+44	0.000176
9	SS ₀ \Rightarrow CG	-0.076746	0.360708	0.546929	0.391232	0.139814	5.12297	19.1261	100+1000	0.009945
10	SS ₁ \Rightarrow CG	0.639264	-0.210716	0.113571	0.253219	-0.415621	5.26821	32.936	400+217	0.00020
11	SS ₀ \Rightarrow Nwt	-0.075087	0.366233	0.547403	0.386901	0.136195	5.10387	19.19	100+1000	0.099445
12	SS ₁ \Rightarrow Nwt	0.639768	-0.208934	0.106940	0.255050	-0.416852	5.29602	32.9338	400+72	0.000186
13	SS ₀ \Rightarrow SDm \Rightarrow CG	0.637520	-0.214199	0.117177	0.244124	-0.420675	5.28179	33.0038	100+20+3	0.000154
14	SS ₀ \Rightarrow SDm \Rightarrow Nwt	0.639052	-0.211207	0.114501	0.252184	-0.415999	5.26707	32.9427	100+20+28	0.000196

Procedures

SS₀ = Step-search across a 10 \times 10 uniform grid; SS₁ = Step-search across a 20 \times 20 uniform grid; SS₂ = Step-search across a 40 \times 40 uniform grid; SD = Steepest descent; $\lambda_\sigma = 1.0$; $\lambda_\tau = 10.0$; SDm = Steepest descent w/momentum; $\lambda_\sigma = 1.0$; $\lambda_\tau = 10.0$; $\mu_\sigma = 0.5$; $\mu_\tau = 1.0$. CG = Conjugate gradient; $\lambda_\sigma = 1.0$; $\lambda_\tau = 10.0$. Nwt = Newton's method.

Stopping criterion

(1) Sqr. Err. \leq 0.0002; or (2) Number of iterations (by stages > 1) \geq 1000

12, 9} for VPB1 and {11, 14, 7, 12, 10, 5, 8, 9, 13, 6} for VPB2. Therefore, given the desired approach must take as few iterations as possible and must always reach the solution, the approach SS₀ \Rightarrow SDm \Rightarrow Nwt (row 14) was selected and used by the applications described in [2, 3].

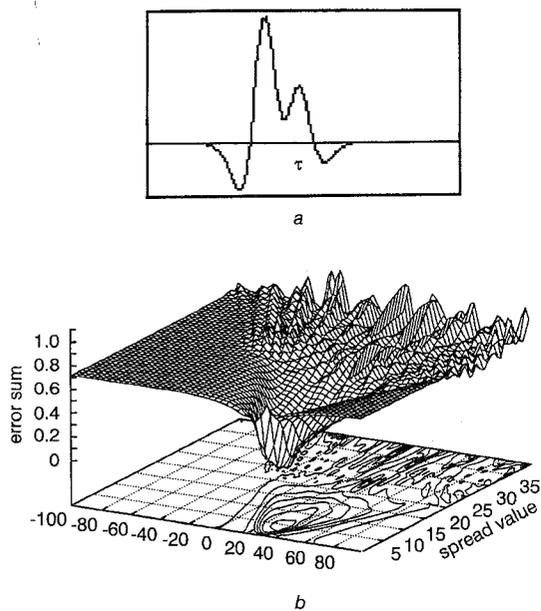


Fig. 7 Error-sum surface for a 'shifted' signal with respect to σ and τ
a $s(t - \tau)$
b Error sum surface of $s(t - \tau)$

2.6 Effect of signal scaling, translation and dilation on Hermite coefficients

In this Section the effects of signal scaling, translation and dilation on its Hermite spectral coefficients, spread term and shift term are determined. The influence of these three common transformations is particularly important because they are a first-order approximation

of the kinds of typical distortions experienced by signals. Consider a signal $u(t)$ whose Hermite spectral coefficients are $\{b_0, b_1, \dots, b_{N-1}\}$ for some spread σ_u and shift τ_u , which have been determined such that the squared error of the fit is minimised, i.e.

$$u(t) = u_{N, \sigma_u}(t) = \sum_{i=0}^{N-1} b_i h_{i, \sigma_u}(t, \tau_u) \quad (19)$$

Let signal $v(t)$ be defined as a scaled version of signal $u(t)$, with Hermite spectral coefficients $\{c_0, c_1, \dots, c_{N-1}\}$ and spread σ_v and shift τ_v or equivalently described with Hermite spectral coefficients $\{a_0, a_1, \dots, a_{N-1}\}$ and spread σ_{vv} .

$$v(t) = \alpha u(\beta t - \tau) \quad \text{where } \alpha, \beta \text{ and } \tau \text{ are scalars} \quad (20)$$

$$\begin{aligned} v(t) &= v_{N, \sigma_v}(t) \\ &= \sum_{i=0}^{N-1} a_i h_{i, \sigma_{vv}}(t) \\ &= \sum_{i=0}^{N-1} c_i h_{i, \sigma_v}(t, \tau_v) \end{aligned} \quad (21)$$

The parameters τ_v , σ_v and c_n can be rewritten in terms of τ_u , σ_u and b_n as follows,

$$\tau_v = \frac{\tau + \tau_u}{\beta} \quad \sigma_v = \frac{\sigma_u}{\beta} \quad \text{and} \quad c_n = \alpha \frac{b_n}{\sqrt{\beta}} \quad (22)$$

where $\beta > 0$ and $n \in \{0, 1, \dots, N-1\}$

Also, if $\tau = 0$ in eqn. 20 then,

$$\sigma_{vv} = \frac{\sigma_u}{\beta} \quad \text{and} \quad a_n = \alpha \frac{b_n}{\sqrt{\beta}} \quad (23)$$

where $\beta > 0$ and $n \in \{0, 1, \dots, N-1\}$

In summary, the relationships of eqns. 22 and 23 provide a means for estimating the Hermite coefficients, spread and shift for a distorted signal, $v(t)$, using the known parameters of the undistorted signal, $u(t)$, and the distortion, as given by parameters τ , β and α . In addition, by comparing the Hermite model parameter estimates of $u(t)$ with that of $v(t)$, the values of the distortion parameters can be determined.

The results of eqns. 22 and 23 are particularly useful because they provide a method for describing a signal by appropriately defining the feature vector in terms of the signal's Hermite model parameters. For example, if there is to be no discrimination between a signal $u(t)$ and its distortion $\alpha u(\beta t)$, then the norm of its Hermite coefficients is taken:

$$\Gamma(u(t)) = \vec{a} = \frac{[a_0 \ a_1 \ \dots \ a_{N-1}]^T}{\max\{|a_0, a_1, \dots, a_{N-1}|\}} = \Gamma(\alpha u(\beta t)) \quad (24)$$

where $\Gamma(u(t))$ defines the feature vector of signal $u(t)$ and represents a feature extraction process on $u(t)$.

Similarly, if there is a need to discriminate between a signal's scale distortions in time but not in magnitude then,

$$\Gamma(u(t)) = [\sigma_u \ \vec{a}]^T = \Gamma(\alpha u(t)) \neq \Gamma(\alpha u(\beta t)) \quad (25)$$

where $\beta \neq 1$

Hence by choosing the features in the manner similar to that described by examples eqns. 24 and 25, the type of discrimination between signals can be predefined.

2.7 Order estimation

Evaluating the order of our series expansion is very much dependent on the 'patterns' to be classified. Since these patterns are known (and it is assumed that there are K exemplars of their typical form), a value for the order, N , is chosen such that there is no significant improvement in the fit for the order greater than N . To do this, some measure of the fit is required. Two such 'measures' are considered i.e. the squared error and the correlation coefficient.

Considering the squared error as the first measure of the fit.

$u_k(x)$ = the x th sample of the k th exemplar, where the exemplar is $2L + 1$ samples long

$u_{k,n,\sigma}(x)$ = the x th sample of the n th order synthesised k th exemplar

$\epsilon_k(n)$ = the 'measure-of-fit' of the k th exemplar for a n th-order fit

$$\text{squared error, } \epsilon_k(n) = \sum_{x=-L}^L (u_k(x) - u_{k,n,\sigma}(x))^2 \quad (26)$$

$$\vec{\epsilon}_k = [\epsilon_k(0), \epsilon_k(1), \dots, \epsilon_k(N-1)]^T$$

$$E(n) = \frac{1}{K} \sum_{k=0}^{K-1} \frac{\vec{\epsilon}_k(n)}{\|\vec{\epsilon}_k(n)\|_\infty} \quad (27)$$

The maximum value for the normalised averaged error, $E(n)$ is unity for $n \in \{0, 1, \dots, N-1\}$. In general, $n = 0$ has the worst fit, which implies $E(0) = 1$. Let the order be n when $E(n+1) \geq E(n)$. This condition for picking n does not always work and depends on the convergent properties of the exemplars i.e. the inequality is satisfied for large values of n only.

Next consider the correlation coefficient as the measure-of-fit [13] i.e.

$$\text{linear correlation coefficient, } \epsilon_k(n) = \frac{s_{uv}(k)}{s_u(k)s_v(k)}$$

where $u = u_k(x)$ and $v = u_{k,n,\sigma}(x)$ (28)

As long as it is of no consequence that the synthesised signal is a linear function of the original signal (i.e. it could be scaled and/or have its mean shifted), it is safe to interpret eqn. 28 as a measure of fit to use.

When the average correlation coefficient $E(n)$ equals unity, there is a perfect fit. Similar to the least squares approach, n is chosen such that $E(n+1) \leq E(n)$. Once again, depending on the convergent properties of the exemplars, this inequality may only be satisfied for large n . Although the basis functions satisfy completeness, there will be a point beyond which there will be no 'significant' improvement in the fit. At this point the truncation error is small and susceptible to precision errors. These errors give rise to an overall error sum that oscillates along a decreasing long-term trend (see Fig. 8). Since this tends to occur for large n , the inequality is modified such that the order is chosen as the smallest value of n that satisfies, $E(n) \geq E_{TH}$ for a predetermined threshold E_{TH} .

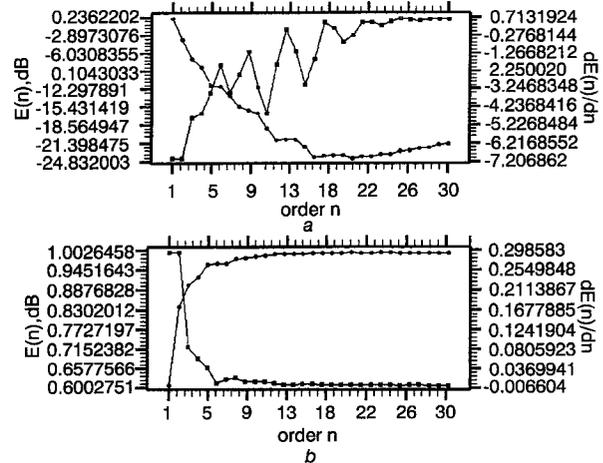


Fig. 8 Plots of $E(n)$ and $dE(n)/dn$ versus N for various definitions of $E(n)$
a $E(n)$ defined as sum of squared errors, $E(1) = 0.24$
b $E(n)$ defined as correlation coefficient, $E(1) = 0.60$

This inequality does not particularly suit the least squares approach because E_{TH} is arbitrarily chosen. The correlation coefficient, however, lends itself to the 'threshold' inequality because it is a measure that is not influenced by the scalar variations of $u(t)$, and it is intuitively more meaningful.

To estimate a suitable n , we can either be conservative or liberal in our choice. We define a liberal selection by picking n for which the correlation coefficient ≥ 0.95 and conservative, by picking n for which the correlation coefficient ≥ 0.99 . From Fig. 8, these values of n are 5 and 11 for the liberal and conservative approach, respectively.

3 Applications

The representation of a 1-D signal as a series expansion of Hermite basis functions and the subsequent use of the coefficients of this series expansion as features towards signal analysis has been successfully applied to the area of ECG waveform analysis. The suitability of

Hermite coefficients as morphological features in QRS detection was first suggested by Sornmo *et al.* [6]. However, no method for estimating these coefficients was proposed. Since then, Laguna *et al.* [4, 5] have also investigated the use of Hermite coefficients together with an associated spread term for QRS and ectopic beat detection. However, they did suggest a method for estimating the Hermite coefficients. Theirs was a scheme based on the adaptive linear combiner [13], called the adaptive Hermite model estimation system (AHMES). It used LMS estimation to estimate both the coefficients and the spread term. While the AHMES approach is well suited to tracking an ECG signal, it is unclear if the coefficients and spread term model the 'shape' of the beats and if they can indeed be interpreted as shape features.

Unlike the AHMES system, where the approach seeks to minimise the squared error of each 'incoming' sample, the authors have proposed an approach in [2, 3] which seeks to minimise the squared error of each 'incoming' signal segment. In other words, the cost function is the sum of the squared errors of all the samples taken over a user-defined period within which the shape of signal to be modelled lies. Note that the length of this user-defined period is a heuristic choice such that the 'shapes' being looked for in a signal all fit within this period. Hence the length of this period is determined by the shape with the largest 'width'.

It is worth noting that the first Hermite basis function has the form of a Gaussian distribution and all subsequent Hermite basis functions are derivatives of this form. This could suggest that the Hermite series expansion may be well suited to modelling 'distributions'.

4 Conclusions

The selection of an appropriate orthogonal basis function is very much dependent on the shape of the signals to be modelled. This choice involves finding an orthogonal set of basis functions which most conservatively models the types of 'shapes' intended for recognition. More often than not, the close morphological similarity of these functions to the 'shapes' provides a good indication of their suitability.

In this regard, this paper has purposely considered the orthonormal Hermite basis set given, the suitability of the Hermite parameters in modelling the one-dimensional shapes found in signals such as ECG traces. This paper has outlined a methodology to estimate the Hermite-spectral coefficients of a one-dimensional signal and its associated spread and/or shift term values.

It is worth noting that the deterministic approach to estimating the Hermite spectral coefficients of a 1-D signal, as described in this paper, allows the method to constrain the use of LMS estimation techniques to just the two (one) variables, i.e. the spread and (or) shift term. As a consequence, the visualisation of the search space 'landscape' for these two terms is now a practical proposition. From such visualisation, the problems in trying to estimate an optimum spread and shift term value, like false minima and nonconvergence were addressed. In addition, heuristic approaches to estimating the number of coefficients to use were suggested, and special attention was given to determining the influence of common signal distortions such as scalar, dilatatory and translatatory effects, on the Hermite coefficients, spread and shift parameters.

5 Acknowledgments

We wish to thank the PhD and Scholarship Committee of the University of Western Australia for supporting this work through the University Research Scholarship Award.

6 References

- 1 ABRAMOWITZ, M., and STEGUN, L.A.: 'Handbook of mathematical functions' (Applied Mathematics Series, vol. 55, Dover Publications, 1968)
- 2 RASIAH, A.I., and ATTIKIIOUZEL, Y.: 'Syntactic recognition of common cardiac arrhythmias'. *IEEE Eng., Biology and Medicine Society 16th annual international conference*, November 1994, Baltimore, USA
- 3 RASIAH, A.I., and ATTIKIIOUZEL, Y.: 'A syntactic approach to the recognition of common cardiac arrhythmias within a single ambulatory ECG trace', *Australian Computer J.*, 1994, **26**, (3), pp. 102-112
- 4 LAGUNA, P., CAMINAL, P., THAKOR, N.V., and JANE, R.: 'Adaptive QRS shape estimation using Hermite model'. 11th IEEE annual conference of *Eng., Medicine and Biology Society*, 1989, pp. 683-684
- 5 LAGUNA, P., JANE, R., and CAMINAL, P.: 'Adaptive feature extraction for QRS classification and ectopic beat detection', *Proceedings Computers in Cardiology*, 1992, IEEE Computer Society Press, pp. 613-616
- 6 SORNMO, L., BORJESSON, P.O., NYGARDS, M., and PAHLM, O.: 'A method for evaluation of QRS shape features using a mathematical model for the ECG', *IEEE Trans.*, 1981, **BME-28**, (10), pp. 713-717
- 7 GRADSHTEYN, I.S., and RYZHIK, I.M. 'Table of integrals, series and products' (Academic Press, 1980), pp. 1033, corrected and enlarged edition
- 8 BOYD, J.P.: 'Chebyshev and Fourier spectral methods' in BREBIA, C.A., and ORSZAG, S.A. (Eds.): 'Lecture Notes in Engineering 49' (Springer-Verlag, 1989)
- 9 PRESS, W.H., VETTERLING, W.T., TEUKOLSKY, S.A., and FLANNERY, B.P.: 'Numerical recipes in C. The art of scientific computing' (Cambridge University Press, 1992), Chap. 4, pp. 147-161, 2nd edn
- 10 BOYD, J.P.: 'The asymptotic coefficients of Hermite functions', *J. Computat. Phys.*, 1984, **54**, pp. 382-410
- 11 FLETCHER, R.: 'Practical methods of optimization' (Wiley, 1987), 2nd edn
- 12 SPIEGEL, M.R.: 'Probability and statistics' (Schaum's Outline Series, McGraw-Hill, 1988)
- 13 WIDROW, B., GLOVER, J.R., MCCOOL, J.M., KAUNITZ, J., WILLIAMS, C.S., HEARN, R.H., ZEDLER, J.R., DONG, E., and GOODLIN, R.C.: 'Adaptive noise cancelling: principles and applications', *IEEE Proc.*, 1975, **63**, (12), pp. 1692-1716

7 Appendix

7.1 Estimating the spread value

$u(x)$ = discrete time, finite duration signal for $x \in \{-L, \dots, L\}$

$u_{N,\sigma}(x)$ = synthesised signal from N Hermite coefficients with a spread of σ

$\epsilon_k(\sigma, x)$ = the error between $u(x)$ and $u_{N,\sigma}(x)$ at the k th iteration

error sum,

$$E_N(\sigma) = \sum_{x=-L}^L (u(x) - u_{N,\sigma}(x))^2 = \sum_{x=-L}^L (\epsilon_k(\sigma, x))^2 \quad (29)$$

$$\nabla_k = \frac{\partial E_N(\sigma)}{\partial \sigma} = 2 \sum_{x=-L}^L \epsilon_k(\sigma, x) \frac{\partial \epsilon_k(\sigma, x)}{\partial \sigma} = 0 \quad (30)$$

$$\frac{\partial \epsilon_k(\sigma, x)}{\partial \sigma} = - \sum_{n=0}^{N-1} a_n \frac{\partial h_{n,\sigma}(x)}{\partial \sigma} \quad (31)$$

$$\begin{aligned} \frac{\partial h_{n,\sigma}(x)}{\partial \sigma} &= \frac{1}{2\sigma\sqrt{\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ &\times \left[\sqrt{(n+1)(n+1)} \hat{H}_{n+2}\left(\frac{x}{\sigma}\right) \right. \\ &\quad \left. - \sqrt{n(n-1)} \hat{H}_{n-2}\left(\frac{x}{\sigma}\right) \right] \quad (32) \end{aligned}$$

$$\begin{aligned} &-(2n+1) \hat{H}_n\left(\frac{t-\tau}{\sigma}\right) \\ &+ \sqrt{n(n-1)} \hat{H}_{n-2}\left(\frac{t-\tau}{\sigma}\right) \quad (37) \end{aligned}$$

$$\begin{aligned} \nabla_{kk} &= \frac{\partial^2 E_N(\sigma)}{\partial \sigma^2} \\ &= 2 \sum_{x=-L}^L \left[\left(\frac{\partial \epsilon_k(\sigma, x)}{\partial \sigma} \right)^2 + \epsilon_k(\sigma, x) \frac{\partial^2 \epsilon_k(\sigma, x)}{\partial \sigma^2} \right] \quad (33) \end{aligned}$$

$$\frac{\partial^2 \epsilon_k(\sigma, x)}{\partial \sigma^2} = - \sum_{n=0}^{N-1} a_n \frac{\partial^2 h_{n,\sigma}(x)}{\partial \sigma^2} \quad (34)$$

$$\begin{aligned} \frac{\partial^2 h_{n,\sigma}(x)}{\partial \sigma^2} &= \frac{1}{2} \sigma^{-\frac{5}{2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ &\times \left[\frac{1}{2} \sqrt{(n+4)(n+3)(n+2)(n+1)} \hat{H}_{n+4}\left(\frac{x}{\sigma}\right) \right. \\ &\quad - \sqrt{(n+2)(n+1)} \hat{H}_{n+2}\left(\frac{x}{\sigma}\right) - (n^2+n+1) \hat{H}_n\left(\frac{x}{\sigma}\right) \\ &\quad + \sqrt{n(n-1)} \hat{H}_{n-2}\left(\frac{x}{\sigma}\right) \\ &\quad \left. + \frac{1}{2} \sqrt{n(n-1)(n-2)(n-3)} \hat{H}_{n-4}\left(\frac{x}{\sigma}\right) \right] \quad (35) \end{aligned}$$

7.2 Estimating the shift value

As demonstrated in Section 2.6, the Hermite coefficients are translation variant, i.e. the Hermite coefficients for a signal $u(t-\tau)$ differ for different values of τ . When a shift term, τ , is introduced into the definition of the Hermite basis function, as given by eqn. 5, then least-squares estimation can be used to estimate a value for τ , such that the Hermite functions of eqn. 4 are fitted about $t = \tau$.

In the previous Section, the equations for estimating a spread value for a fixed shift value (i.e. $x = t - \tau$) were presented. Now the remaining cases are considered: (i) estimating the shift value for a fixed spread value; and (ii) estimating the spread and shift value. Note that after each adjustment in value to the spread and/or shift parameter, the Hermite coefficients are recomputed.

7.2.1 For constant σ : Using eqns. 36 to 39, the same LMS techniques are used to estimate the shift value.

$$\begin{aligned} \frac{\partial}{\partial \tau} h_{n,\sigma}(t, \tau) &= \frac{1}{\sigma\sqrt{2\sigma}} \exp\left(-\frac{(t-\tau)^2}{2\sigma^2}\right) \\ &\times \left[\sqrt{n+1} \hat{H}_{n+1}\left(\frac{t-\tau}{\sigma}\right) - \sqrt{n} \hat{H}_{n-1}\left(\frac{t-\tau}{\sigma}\right) \right] \quad (36) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \tau^2} h_{n,\sigma}(t, \tau) &= \frac{1}{2\sigma^2\sqrt{\sigma}} \exp\left(-\frac{(t-\tau)^2}{2\sigma^2}\right) \\ &\times \left[\sqrt{(n+1)(n+2)} \hat{H}_{n+2}\left(\frac{t-\tau}{\sigma}\right) \right. \end{aligned}$$

$$\nabla_{\tau k} = \frac{\partial E_N(\tau)}{\partial \tau} = 2 \sum_{x=-L}^L \epsilon_k(\tau, x) \frac{\partial \epsilon_k(\tau, x)}{\partial \tau} \quad (38)$$

$$\begin{aligned} \nabla_{\tau kk} &= \frac{\partial^2 E_N(\tau)}{\partial \tau^2} \\ &= 2 \sum_{x=-L}^L \left[\left(\frac{\partial \epsilon_k(\tau, x)}{\partial \tau} \right)^2 + \epsilon_k(\tau, x) \frac{\partial^2 \epsilon_k(\tau, x)}{\partial \tau^2} \right] \quad (39) \end{aligned}$$

7.2.2 For variable σ : Once again, the techniques of the previous Section are employed with some notable changes and additions. First, the second derivative of the error sum function is now defined by eqn. 44 and its inverse, i.e. the Hessian matrix, is required for Newton's approach. Next, two more gradient descent techniques are introduced: (i) steepest descent with a momentum term; and (ii) conjugate gradient.

$$v = [\sigma \quad \tau]^T \quad \lambda = \begin{bmatrix} \lambda_\sigma & 0 \\ 0 & \lambda_\tau \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_\sigma & 0 \\ 0 & \mu_\tau \end{bmatrix} \quad (40)$$

$$\begin{aligned} &\frac{\partial^2 h_{n,\sigma}(t, \tau)}{\partial \sigma \partial \tau} \\ &= \frac{\partial^2 h_{n,\sigma}(t, \tau)}{\partial \tau \partial \sigma} \\ &= \frac{1}{2\sigma^2\sqrt{2\sigma}} \exp\left(-\frac{(t-\tau)^2}{2\sigma^2}\right) \\ &\times \left[\sqrt{(n+1)(n+2)(n+3)} \hat{H}_{n+3}\left(\frac{t-\tau}{\sigma}\right) \right. \\ &\quad - (n+2)\sqrt{n+1} \hat{H}_{n+1}\left(\frac{t-\tau}{\sigma}\right) \\ &\quad - (n-1)\sqrt{n} \hat{H}_{n-1}\left(\frac{t-\tau}{\sigma}\right) \\ &\quad \left. + \sqrt{n(n-1)(n-2)} \hat{H}_{n-3}\left(\frac{t-\tau}{\sigma}\right) \right] \quad (41) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 E_N(\sigma, \tau)}{\partial \sigma \partial \tau} &= 2 \sum_{x=-L}^L \left[\frac{\partial \epsilon_k(\sigma, \tau, x)}{\partial \sigma} \frac{\partial \epsilon_k(\sigma, \tau, x)}{\partial \tau} \right. \\ &\quad \left. + \epsilon_k(\sigma, \tau, x) \frac{\partial^2 \epsilon_k(\sigma, \tau, x)}{\partial \sigma \partial \tau} \right] \quad (42) \end{aligned}$$

$$\nabla v_k = \left[\frac{\partial E_N(\sigma, \tau)}{\partial \sigma} \quad \frac{\partial E_N(\sigma, \tau)}{\partial \tau} \right]^T \quad (43)$$

$$\nabla v_{kk} = \begin{bmatrix} \frac{\partial^2 E_N(\sigma, \tau)}{\partial \sigma^2} & \frac{\partial^2 E_N(\sigma, \tau)}{\partial \sigma \partial \tau} \\ \frac{\partial^2 E_N(\sigma, \tau)}{\partial \sigma \partial \tau} & \frac{\partial^2 E_N(\sigma, \tau)}{\partial \tau^2} \end{bmatrix} \quad (44)$$

The motivation for steepest descent with momentum comes from having found that a typical error-sum surface has significant portions that are 'flat'. Therefore the introduction of a momentum term may prove use-

ful in reducing the convergence time. We have also found it useful to include eqn. 47 which drops the momentum to zero when an 'overshoot' (as given by the predetermined $\Delta E_{\text{THRESHOLD}}$) occurs (see Fig. 7b).

$$s_{k+1} = \mu s_k - \lambda \nabla v_k \quad (45)$$

$$v_{k+1} = v_k + s_{k+1} \quad (46)$$

$$\text{if}((E_N(\sigma_{k+1}, \tau_{k+1}) - E_N(\sigma_k, \tau_k)) > \Delta E_{\text{THRESHOLD}}) \\ s_k = 0 \quad (47)$$

With conjugate gradient, if the error-sum surface is of 'quadratic' form, then a faster convergence can be expected:

$$\mu_{k+1} = \frac{[\nabla v_{k+1} - \nabla v_k]^T \nabla v_{k+1}}{\|\nabla v_k\|^2} \quad (48)$$

$$s_{k+1} = \mu_{k+1} s_k - \nabla v_{k+1} \quad (49)$$

$$v_{k+1} = v_k + \lambda s_{k+1} \quad (50)$$