Noise reduction, smoothing and time interval segmentation of noisy signals using an energy optimisation method

S. Mahmoodi and B.S. Sharif

Abstract: Noise reduction and time interval segmentation of a noise-contaminated piecewise continuous signal is considered by the authors as a non-linear optimisation problem. The mathematical framework of this method is presented both in continuous-time and discrete-time domains. The smoothed signal and segmented time intervals of the original noisy signal are calculated as an optimised solution for an energy functional. An algorithm similar to the level set method is developed to find the optimised solution. In this algorithm, the discontinuity points separating consecutive continuous signals are preserved while the noise is reduced. Therefore this method fundamentally exhibits a better performance compared with a traditional low-pass filter suppressing high frequency components, including discontinuity points. The results also demonstrate a better quality in noise reduction in comparison to the median and Gaussian filters.

1 Introduction

The purpose of this paper is to introduce a new method based on energy optimisation in signal processing for noise reduction and signal smoothing. Similar methods have long been developed and used in image processing and computer vision [1–7]. Image restoration, also known as the 'inverse problem', was developed by Rudin *et al.* [7] as an optimisation method based on the concept of total variation. In another development within computer vision, a method known as the 'snake' algorithm was introduced for object segmentation in images by Kass *et al.* [2] and was further developed as the Geodesic Active Contours Model and the Level-Set Method [8–12]. Alternatively, a non-linear functional was proposed by Mumford and Shah [1] and later implemented by others [4, 5, 13] to simultaneously segment and denoise images contaminated with noise. However, Mumford and Shah's model has only been developed for optimisation problems in computer vision and it is inappropriate for signals because of the fact that in their model, a contour whose length is minimised is assumed to represent the discontinuity surrounding objects in an image. This contour representation for discontinuities in images should be replaced by single points in signals. A novel method is therefore introduced in this paper to model signals as piecewise continuous functions for noise reduction and smoothing purposes based on the energy optimisation. A similar modelling approach for

signals is proposed in [14–18] in which discontinuities are initially detected using (pseudo-) spectral methods based on Fourier and Legendre series [14, 15]. The objective of this spectral method is then to recover the segmented signals from artefacts introduced by the Gibbs phenomenon, by using the Gegenbauer reconstruction algorithm [16–18], whereas in our case, noise reduction of the segmented signals is the main aim of the proposed method. Furthermore in our approach, segmentation and noise reduction are achieved simultaneously in contrast to the spectral method, where segmentation and reconstruction are performed separately. The difficulty of investigating the non-linear functional considered in this paper is the lack of differentiability in discontinuities. Hence, Euler–Lagrange equations cannot be employed in this optimisation problem.

A signal $g(t)$ is considered as a timed sequence of separate continuous signals that are subject to channel noise and degradation. $g(t)$ can be approximated as a piecewise continuous function $f(t)$ consisting of time series of at least class C^2 functions $f_i(t)$ over a time interval (t_{i-1}, t_i). These continuous functions $f_i(t)$ are optimal solutions of the following energy functional

$$E(f, S) = \frac{1}{2} \sum_i \int_{t_{i-1}}^{t_i} \left[(f_i(t) - g(t))^2 + \mu \left(\frac{\mathrm{d}f_i(t)}{\mathrm{d}t} \right)^2 \right] S_i(t) \, \mathrm{d}t \tag{1}$$

where $E(f, S)$ is the energy functional to be optimised, $f_i(t)$ the smoothed function approximating $g(t)$ over the time interval (t_{i-1}, t_i), μ the non-negative parameter, $S_i(t)$ the segmented time interval, which is a rectangular window function in the time domain and can be defined as follows

$$S_i(t) = \begin{cases} 1 & t_{i-1} < t < t_i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

In (1), the first term, $(f_i(t) - g(t))^2$, is a data fidelity term whose minimisation indicates that $f_i(t)$ approximates $g(t)$.

S. Mahmoodi is with the Psychology Division, Henry Wellcome Building, School of Biology and Pyschology, University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

B.S. Sharif is with the School of Electrical, Electronic, and Computer Engineering, Merz Court, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

E-mail: sasan.mahmoodi@ncl.ac.uk

The minimisation of the second term, which is a smoothing term, guarantees a differentiable and smoother function $f_i(t)$ than the original signal $g(t)$. Finally, $S_i(t)$ implies that this approximation is only performed for $f_i(t)$ over the time interval (t_{i-1}, t_i) where there is no discontinuity in $g(t)$. In fact, it is demonstrated in this paper that $S_i(t)$ functions, which correspond to the intervals over which $g(t)$ is continuous, minimise functional (1). The objective of this paper is to find functions $f_i(t)$ and $S_i(t)$ over each time interval i as the best solution minimising the energy functional $E(f, S)$ described in (1).

2 Optimisation of energy functional

2.1 Continuous-time signals

A pair of functions, $f_i(t)$ and $S_i(t)$, is required to optimise the energy functional in (1) for each time segment i. As $S_i(t)$ is a rectangular window function in the time domain, its general form is known and it can therefore be determined by specifying the time interval (t_{i-1}, t_i) minimising the energy functional. $f_i(t)$ is assumed to be a continuous function with continuous first derivative over the time interval (t_{i-1}, t_i). Methods of calculus of variations are used in this section to derive the differential equations leading to solutions $f_i(t)$ and $S_i(t)$ [19, 20].

We initially consider an arbitrary time interval (t_{i-1}, t_i) and find the solutions $f_i(t)$ and $S_i(t)$ to optimise energy functional (1). These solutions can then be applied to any time interval in (1). In the time interval (t_{i-1}, t_i), the energy functional $E_i(f_i, S_i)$ is considered as

$$E_i(f_i, S_i) = \frac{1}{2} \int_{t_{i-1}}^{t_i} \left[(f_i(t) - g(t))^2 + \mu \left(\frac{df_i}{dt} \right)^2 \right] S_i(t) \, dt \quad (3)$$

$S_i(t)$ is initially considered fixed and $f_i(t)$ is varied. In this case, functional (3) is convex and therefore its minimiser $f_i(t)$ exists [20]. Equation (3) can therefore be rewritten as

$$E_i(f_i, S_i) = \frac{1}{2} \int_{t_{i-1}}^{t_i} \left[(f_i(t) - g(t))^2 + \mu \left(\frac{df_i}{dt} \right)^2 \right] dt \quad (4)$$

Let δf_i represent a class C^2 function, then we calculate the variations of $E_i(f_i, S_i)$ by varying $f_i(t)$ as much as δf_i

$$E_i(f_i + m\delta f_i, S_i) - E_i(f_i, S_i)$$

$$= \frac{1}{2} \left(\int_{t_{i-1}}^{t_i} \left[(f_i + m\delta f_i - g)^2 + \mu \left(\frac{d(f_i + m\delta f_i)}{dt} \right)^2 \right] dt \right.$$

$$\left. - \int_{t_{i-1}}^{t_i} \left[(f_i - g)^2 + \mu \left(\frac{df_i}{dt} \right)^2 \right] dt \right)$$

$$\delta E_i = \frac{1}{2} \left(\int_{t_{i-1}}^{t_i} [(f_i + m\delta f_i - g)^2 - (f_i - g)^2] \, dt \right.$$

$$\left. + \mu \int_{t_{i-1}}^{t_i} \left[\left(\frac{d(f_i + m\delta f_i)}{dt} \right)^2 - \left(\frac{df_i}{dt} \right)^2 \right] dt \right)$$

$$\delta E_i = \frac{1}{2} \int_{t_{i-1}}^{t_i} \left[(2f_i m\delta f_i + m^2 \delta f_i^2 - 2gm\delta f_i) \right.$$

$$\left. + \mu \left(m^2 \left(\frac{d\delta f_i}{dt} \right)^2 + 2m \left(\frac{df_i}{dt} \right) \left(\frac{d\delta f_i}{dt} \right) \right) \right] dt$$

$$\frac{\delta E_i}{\delta f_i} = \lim_{m \to 0} \frac{\delta E_i}{m} = \int_{t_{i-1}}^{t_i} (f_i - g)\delta f_i \, dt + \mu \int_{t_{i-1}}^{t_i} \left(\frac{df_i}{dt} \right) \left(\frac{d\delta f_i}{dt} \right) dt$$

(5)

Integrating by part and assuming that f_i changes only over the time interval (t_{i-1}, t_i) and remains fixed at t_{i-1} and t_i

$$\frac{\delta E_i}{\delta f_i} = \int_{t_{i-1}}^{t_i} \left[(f_i - g) - \mu \frac{d^2 f_i}{dt^2} \right] \delta f_i \, dt = 0 \quad (6)$$

As $\delta f_i \neq 0$ over the time interval (t_{i-1}, t_i), in order that (6) becomes zero with the optimised f_i, we should have

$$\mu \frac{d^2 f_i(t)}{dt^2} = (f_i(t) - g), \quad t_{i-1} < t < t_i \quad (7)$$

The solution to differential equation (7) would optimise the energy functional (1) over the time interval (t_{i-1}, t_i). The time interval (t_{i-1}, t_i) optimising energy functional (1) can be specified by varying the time interval and optimising functional (1) by considering variations of f_i. This is achieved by varying t_i and calculating f_i variations in a neighbourhood of t_i. The conclusion would then be applicable to t_{i-1} as well. Fig. 1 shows how f_i and f_{i+1} change if t_i moves to t_i' where $t_i' = t_i \pm \delta t_i$.

Let us assume that t_i corresponds to a discontinuity and study the behaviour of f by moving t_i backward and forward as much as δt_i. This is to say that variations of f are considered in a neighbourhood (say I) of t_i'. We further define f^+ and f^- corresponding to $t_i' = t_i + \delta t_i$ and $t_i' = t_i - \delta t_i$, respectively, as can be seen in Fig. 1.

$$f^+ = \begin{cases} f_i^+ & t \in I \text{ and } t < t_i + \delta t_i \\ f_{i+1}^+ & t \in I \text{ and } t > t_i + \delta t_i \\ \text{unchanged} & t \notin I \end{cases}$$

$$f^- = \begin{cases} f_i^- & t \in I \text{ and } t < t_i - \delta t_i \\ f_{i+1}^- & t \in I \text{ and } t > t_i - \delta t_i \\ \text{unchanged} & t \notin I \end{cases}$$
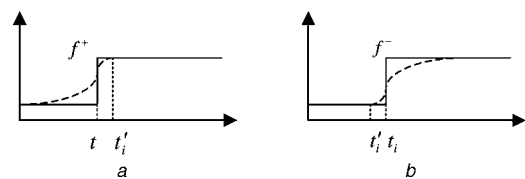

Fig. 1 *Variations of f by varying t_i*

S^+ and S^- are defined in a similar fashion. Variations in E are then calculated as

$$\delta E^+ = E^+(f, t_i + \delta t_i) - E(f, t_i)$$

$$= \frac{1}{2}\sum_i \int_0^{+\infty}\left[(f^+ - g)^2 + \mu\left(\frac{df^+}{dt}\right)^2\right]S_i^+(t)\,dt$$

$$- \frac{1}{2}\sum_i \int_0^{+\infty}\left[(f - g)^2 + \mu\left(\frac{df}{dt}\right)^2\right]S_i(t)\,dt \qquad (8)$$

$$\delta E^- = E(f, t_i) - E^-(f, t_i - \delta t_i)$$

$$= \frac{1}{2}\sum_i \int_0^{+\infty}\left[(f - g)^2 + \mu\left(\frac{df}{dt}\right)^2\right]S_i(t)\,dt$$

$$- \frac{1}{2}\sum_i \int_0^{+\infty}\left[(f^- - g)^2 + \mu\left(\frac{df^-}{dt}\right)^2\right]S_i^-(t)\,dt \qquad (9)$$

$$\delta E = \delta E^+ + \delta E^- = \frac{1}{2}\int_I\left[(f^+ - g)^2 + \mu\left(\frac{df^+}{dt}\right)^2\right]dt$$

$$- \frac{1}{2}\int_I\left[(f^- - g)^2 + \mu\left(\frac{df^-}{dt}\right)^2\right]dt \qquad (10)$$

In order to optimise the energy functional (1) with respect to t_i, (10) should be set equal to zero, while $\delta t_i \to 0$, that is

$$\frac{\delta E}{\delta t_i} = \left[(f^+ - g)^2 + \mu\left(\frac{df^+}{dt}\right)^2\right]$$

$$- \left[(f^- - g)^2 + \mu\left(\frac{df^-}{dt}\right)^2\right] = 0 \qquad (11)$$

In a neighbourhood of the discontinuity, t_i satisfying (11) corresponds to the discontinuity point. This concept has been used to develop an algorithm for discontinuity detection in Section 3.

In addition to the first variation, it is also important to consider the second variation of functional (1) with respect to t_i. Using (8) and (9), the second variation can be written as

$$\delta^2 E = \delta E^+ - \delta E^- = \frac{1}{2}\int_I\left[(f^+ - g)^2 + \mu\left(\frac{df^+}{dt}\right)^2\right]dt$$

$$+ \frac{1}{2}\int_I\left[(f^- - g)^2 + \mu\left(\frac{df^-}{dt}\right)^2\right]dt$$

$$- \int_I\left[(f - g)^2 + \mu\left(\frac{df}{dt}\right)^2\right]dt \qquad (12)$$

where I is the neighbourhood that includes the variations of f^+ and f^-.

If the neighbourhood I includes a discontinuity over which f^+ and f^- are smoothed, then this implies that in such a neighbourhood, the following properties are correct: the slope of f^+ and f^- are higher than f and the terms $(f^+ - g)^2$ and $(f^- - g)^2$ are larger than $(f - g)^2$, as shown in Fig. 1. As a result, according to (12), $\delta^2 E > 0$, which indicates that a discontinuity point is a minimiser of functional (1). Let us now assume that neighbourhood I is away from any discontinuity so that in this neighbourhood, the original signal g, and hence f and its first

derivative, is continuous. Hence

$$f^+ = f^- = f$$
$$\frac{df^+}{dt} = \frac{df^-}{dt} = \frac{df}{dt}$$

It is seen from (11) that $\delta E = 0$ for a point away from any discontinuity. Furthermore, according to (12), $\delta^2 E = 0$. Therefore any point away from any discontinuity is a saddle point of our functional and is not a solution to the minimisation problem. To detect discontinuities as minimisers of functional (1), the following minimisation condition must therefore be met for any point in question [20], that is

$$\delta E = 0$$

$$\delta^2 E > 0$$

At this stage, it is interesting to examine the behaviour of (4) and (7) by varying coefficient μ. If $\mu \to 0$, then the first term in (4) becomes dominant, implying that f_i follows $g(t)$ and its fluctuations. However if $\mu \to \infty$, the first term in (4) can be ignored, and according to (7), f_i is heavily smoothed and therefore approximated as a line. In Section 3, the implementation of the algorithm based on (7) and (11) is described to minimise (3). This leads to a noise removal method, preserving discontinuities that characterise the noiseless original signal. Before implementation, the equivalence of (7), (11) and (12) and the minimisation condition discussed in this section for discrete-time signals will be derived in Section 2.2.

2.2 Discrete-time signals

In this section, discontinuities in discrete-time signals are defined based on a threshold-based approach, and hence, we prove that they are the minimisers of the proposed functional. The formulation based on minimising functional (1) for continuous-time signals can be demonstrated to be applicable to discrete-time signals with a few minor changes.

Functional (1) can be reformulated in discrete time as follows

$$E(f, S) = \sum_i \sum_n [(f_i(n) - g(n))^2$$

$$+ \mu(f_i(n) - f_i(n-1))^2]S_i(n) \qquad (13)$$

Equation (7) can be simply discretised into a second-order difference equation, that is

$$\mu(f_i(n+1) - 2f_i(n) + f_i(n-1)) = f_i(n) - g(n) \qquad (14)$$

where $g(n)$ are the samples from continuous-time signal $g(t)$, and $f_i(n)$ are calculated iteratively using (14).

The concept of discontinuity discussed in Section 2.1 for continuous-time signals is defined for discrete-time signals, and an equivalent equation to (11) for discontinuity detection in discrete signals is derived. To define a discontinuity in a discrete-time signal $g(n)$ at sample $n + 1$, we use

$$|g(n+1) - g(n)| = M > T \qquad (15)$$

where T is the threshold to discriminate between a sample representing discontinuity and a sample corresponding to continuity. At sample i where the sampled data are considered continuous, the following property applies
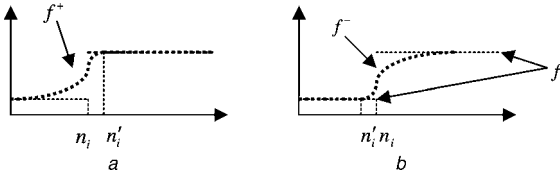
$$|g(i) - g(i-1)| = m \le T \qquad (16)$$

Fig. 2 *Variations of f(n) by changing time intervals with one sampling period*

Let us first assume that our sampled data are characterised by one single discontinuity defined by (15) and (16), as depicted in Fig. 2

$$g(n+1) - g(n) = M > T$$

This can further be generalised to a signal with any number of discontinuities. Let us assume that the earlier inequality applies at $n+1$ and n, so that samples $n+1$ and n are included in the intervals introduced by $S_i(n)$ and $S_{i+1}(n)$ respectively. If $S_i(n)$ and $S_{i+1}(n)$ are moved forward by one sampling period, then $f_i(n)$ and $f_{i+1}(n)$ are changed to $f_i^+(n)$ and $f_{i+1}^+(n)$. However f^+ would differ from f in a neighbourhood (say I) of the discontinuity point. Let us now calculate the variations in the energy term (13) because of these variations in $S_i(n)$ and $S_{i+1}(n)$

$$\Delta E^+ = \sum_{n \in I} [(f^+(n) - g(n))^2 + \mu(\Delta f^+)^2]$$
$$- \sum_{n \in I} [(f(n) - g(n))^2 + \mu(\Delta f)^2] \quad (17)$$

where $\Delta f^+ = f^+(n) - f^+(n-1)$ and $\Delta f = f(n) - f(n-1)$.

It can be proved that in the neighbourhood of a discontinuity, $\Delta E^+ > 0$. To do this, we rewrite (17) as

$$\Delta E^+ = \sum_{n \in I} [(f^+(n) - g(n))^2 - (f(n) - g(n))^2]$$
$$+ \sum_{n \in I} \mu[(\Delta f^+)^2 - (\Delta f)^2] \quad (18)$$

Using the boundary condition at $n+1$, $f^+(n+1) = g(n+1)$, (14) at n can be written as

$$f^+(n) = \frac{\mu g(n+1) + \mu f^+(n-1) + g(n)}{2\mu + 1}$$

In contrast, as a boundary condition for $f(n)$, we have

$$f(n) = g(n)$$

Therefore

$$f^+(n) - f(n) = \frac{\mu(g(n+1) - g(n)) + \mu(f^+(n-1) - g(n))}{2\mu + 1}$$
$$= \frac{\mu(M + f^+(n-1) - g(n))}{2\mu + 1}$$

For $\mu > 0$, $f^+(n) - f(n)$ must be positive. Otherwise, for $f^+(n) - f(n) \leq 0$, it is concluded that $f^+(n) \leq g(n)$. As $f^+(n+1) = g(n+1)$, then

$$f^+(n+1) - f^+(n) \geq g(n+1) - g(n)$$

The earlier mentioned inequality cannot be true, as the objective of the functional is to minimise f^+ fluctuations in comparison to g. Therefore

$$f^+(n) - f(n) > 0 \quad (19)$$

If interval I starts from sample m and the discontinuity is at sample n so that $n > m$ and $f^+(m) = f(m)$ as can be seen from Fig. 2, then (14) can be written for both f^+ and f as

$$f^+(m+1) = \frac{\mu(f^+(m+2) + f^+(m)) + g(m+1)}{2\mu + 1}$$

$$f(m+1) = \frac{\mu(f(m+2) + f(m)) + g(m+1)}{2\mu + 1}$$

Then

$$f^+(m+1) - f(m+1) = \frac{\mu}{2\mu + 1}(f^+(m+2) - f(m+2))$$
$$(20)$$

$$f^+(m+2) - f(m+2) = \frac{\mu}{2\mu + 1}[(f^+(m+3) - f(m+3))$$
$$+ (f^+(m+1) - f(m+1))] \quad (21)$$

$$f^+(i) - f(i) = \frac{\mu}{2\mu + 1}[(f^+(i+1) - f(i+1))$$
$$+ (f^+(i-1) - f(i-1))],$$
$$m+1 \leq i \leq n \quad (22)$$

As implied from (20)–(22), sequences $f^+(i) - f(i)$ for $m+1 \leq i \leq n$ must have the same sign, that is, they are all either positive or negative. According to inequality (19), as $f^+(n) - f(n) > 0$ (in the case of $g(n+1) - g(n) = M > 0$), therefore

$$f^+(i) - f(i) > 0, \quad m+1 \leq i \leq n \quad (23)$$

It is easy to conclude that

$$f^+(i) - g(i) > f(i) - g(i), \quad m+1 \leq i \leq n \quad (24)$$

According to (23), and (20), it can be further seen that

$$f^+(m+2) - f(m+2) > f^+(m+1) - f(m+1) \quad (25)$$

Using (21) and (25), it is easy to prove that

$$f^+(m+3) - f(m+3) > f^+(m+2) - f(m+2) \quad (26)$$

and in general

$$f^+(i) - f(i) > f^+(i-1) - f(i-1),$$
$$m+1 \leq i \leq n \quad (27)$$

Therefore

$$f^+(i) - f^+(i-1) > f(i) - f(i-1),$$
$$m+1 \leq i \leq n \quad (28)$$

Using inequalities (24) and (28), it can be seen in (18) that $\Delta E^+ > 0$. In contrast, ΔE^- can be calculated using the same method for ΔE^+ in (18)

$$\Delta E^- = \sum_{n \in I} [(f(n) - g(n))^2 - (f^-(n) - g(n))^2]$$
$$+ \sum_{n \in I} \mu[(\Delta f)^2 - (\Delta f^-)^2] \quad (29)$$

where f^- is obtained by moving $S_i(n)$ and $S_{i+1}(n)$ backward by one sample and is defined in the same fashion as f^+. Using the same method for ΔE^+, it can be proven that $\Delta E^- < 0$.

This implies that

$$\Delta E^+ \Delta E^- < 0 \quad (30)$$

With the same method, it can be seen that inequality (30) is applicable in the neighbourhood of a discontinuity satisfying $g(n+1) - g(n) = M < -T < 0$. This inequality indicating a zero crossing in ΔE at discontinuities for discrete-time signals is equivalent to (11) for continuous-time signals. The minimisation condition discussed in Section 2.1 therefore requires that at any point satisfying (30), the following inequality is met

$$\Delta E^+ - \Delta E^- = T > 0$$

In the above equation, T is the threshold that can be empirically set by the user to avoid over- or under-segmentation in the presence of noise.

In the next section, we use the concepts discussed in Sections 2.1 and 2.2 to develop an algorithm for discontinuity detection.

3 Implementation

Before discussing the details of algorithm development, practical issues concerning discretising (7) and (11) need to be addressed. The desired solution, $f(t)$, can be considered as a piecewise continuous function minimising an energy term such as

$$E(f, S) = \sum_i \left[\int \mu \left(\frac{\mathrm{d}f_i}{\mathrm{d}t} \right)^2 S_i(t)\, \mathrm{d}t \right.$$
$$\left. + \sum_n (f_i(n) - g(n))^2 S_i(n) \right] \qquad (31)$$

where $f(n)$ and $S_i(n)$ are samples of $f(t)$ and $S_i(t)$ with the same sampling rate used for $g(n)$. Equations (7) and (11) would then change, respectively, to

$$\mu \left. \frac{\mathrm{d}^2 f_i}{\mathrm{d}t^2} \right|_{t=n} = f_i(n) - g(n) \qquad (32)$$

$$\Delta E^+ \Delta E^- < 0 \qquad (33)$$

where ΔE^+ and ΔE^- are calculated as

$$\Delta E^+ = \sum_{n \in I} [(f^+(n) - g(n))^2 - (f(n) - g(n))^2]$$
$$+ \mu \int_{t \in I} \left[\left(\frac{\mathrm{d}f^+}{\mathrm{d}t} \right)^2 - \left(\frac{\mathrm{d}f}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t$$

$$\Delta E^- = \sum_{n \in I} [(f(n) - g(n))^2 - (f^-(n) - g(n))^2]$$
$$+ \mu \int_{t \in I} \left[\left(\frac{\mathrm{d}f}{\mathrm{d}t} \right)^2 - \left(\frac{\mathrm{d}f^-}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t$$

where I is a neighbourhood of discontinuity.

Discrete estimates of $\mathrm{d}f/\mathrm{d}t$ and $\mathrm{d}^2 f/\mathrm{d}t^2$ in (32) and (33) can then be implemented by using different methods such as spline transform [21–24] or Taylor's series (finite difference) [5, 19, 25, 26]. Up-sampling and down-sampling can also be defined for interpolation and magnification in both methods. Smoothing spline [21] is a linear case of energy term (31), where the original signal is assumed not to contain any discontinuity. Therefore a modification for the energy term used in smoothing spline [21] such as (31) is required to enable it to solve this non-linear optimisation problem. However, this would then lead to a similar energy term to that given in (31). Spline transform is computationally more expensive because of the requirements for control-point calculations; however, it is smoother than

the finite-difference method. The comparison of numerical accuracy of these two methods remains to be investigated. In this paper, finite difference is used for discretisation where the desired solution f is sampled at the same sampling rate used for the original signal $g(t)$. It is important to note that similar to smoothing spline, higher derivatives could be added to (1) or (31). In the case of signals represented as piecewise continuous functions separated by discontinuity points, this would then lead to slightly more smoothed signals albeit at the expense of added numerical complexity. However, including higher derivative terms could detect points in which the desired signal is C^n, where n is the highest order of derivative in the functional. In this case, the desired signal can be approximated by piecewise continuous polynomials with degree n whose parameters are calculated by minimising the functional.

An algorithm is proposed to implement the non-linear minimisation method considered in Section 2. This algorithm is based on (7) or its equivalence in the discrete-time domain (30) and minimisation condition discussed in Section 2. It should be noted that this algorithm is similar to the level set method used in computer vision for segmentation [4, 5, 8]. The main difference between the proposed method and the level set method is that in our case, the smoothed function f is used instead of signed distance function which is widely used in the level set framework.

The idea is that a point representing the discontinuity is assumed to be moving along the signal. At any sample, signal is broken in two parts and for each part, a smoothed function f is calculated using (7) or (14). Functional (1) is then obtained for the left and right neighbourhoods (I) of the point, respectively, and the difference ΔE can be considered as 'geometrical forces' acting on the point to move it in a direction of the stronger force. If the point crosses a discontinuity, (30) is satisfied, that is, the direction of the stronger force reverses. However, before the direction of the point movement reverses due to the stronger force, the second variation $(\Delta E)_n - (\Delta E)_{n-1}$ is evaluated to ensure that it is greater than the user-defined threshold T, that is

$$(\Delta E)_n - (\Delta E)_{n-1} > T$$

where n is the iteration number. If this condition is met, the process continues until the point settles down in a discontinuity in a way similar to the Pendulum movement however with non-linear forces. In this implementation, neighbourhood I is considered as a 10 sample window whose centre is the moving point. As soon as a discontinuity is detected, a new point is assigned to continue moving from the recently detected discontinuity to find the next one. This process repeats until no other discontinuity is detected. There are some issues which need to be addressed with regard to this algorithm. As 'geometrical forces' are non-linear, they sharply increase in a close neighbourhood of a discontinuity and it is therefore necessary to avoid setting the speed of our point proportional to forces acting on it. Instead, the speed is set to a constant value. This constant is halved, every time the moving point crosses a discontinuity, which therefore guarantees a smooth convergence of the point on discontinuity. Fig. 3 shows some iterations of this algorithm applied on a signal with SNR = 11.66, where SNR is the ratio of the power of signal to the power of noise. As can be seen in Fig. 3, the moving point depicted by a square is affected by 'geometrical forces' and finally converges to discontinuities one by one. The detected discontinuities are presented by circles in Fig. 3. Once a discontinuity is detected, a time interval $S_i(t)$ and its
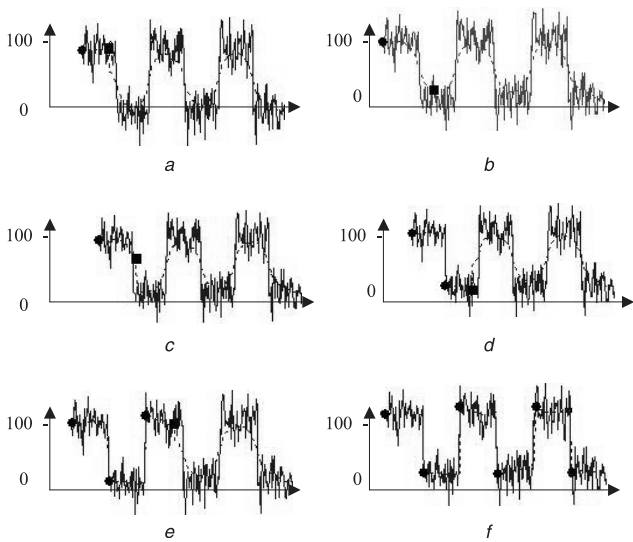
Fig. 3 *Noisy signal with SNR = 11.66 and its smoothed signal using our algorithm ($\mu = 80$, $T = 246$)*

a $n = 2$
b $n = 4$
c $n = 6$
d $n = 15$
e $n = 18$
f $n = 36$

corresponding smoothed function $f_i(t)$ are obtained simultaneously. Finally, it should be noted that the performance of the proposed algorithm will eventually degrade under very low SNR values, which can lead to over or under segmentation.

4 Results

In the first part of this section, we discuss the behaviour of the algorithm, its input and output signals. The second part compares the results of this algorithm with those of the low-pass and median filters [27], and finally, an algorithm is proposed to find an optimised μ for a given noisy signal.

Initially, a signal consisting of six discontinuities is considered as can be seen in Fig. 4 (top of the figure). Gaussian noise with standard deviation 20 is then added to result in a noisy signal as shown in Fig. 4 (middle). The algorithm is finally applied to the noisy signal for noise reduction with
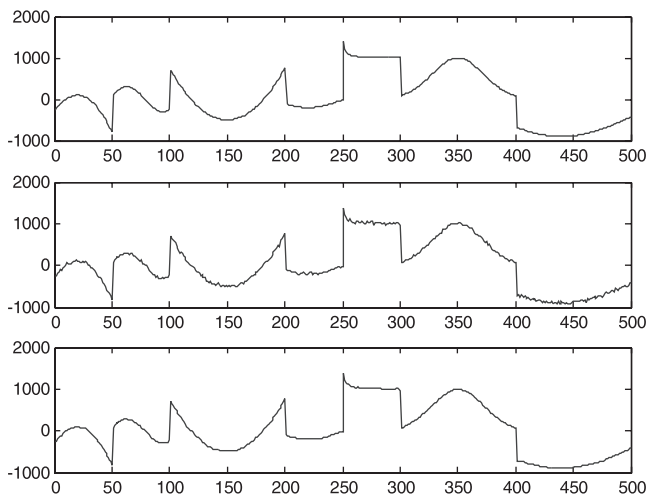

Fig. 4 *Original signal (top), signal with Gaussian noise (middle), enhanced signal using the proposed method with $\mu = 5$ and $T = 460$ (bottom)*
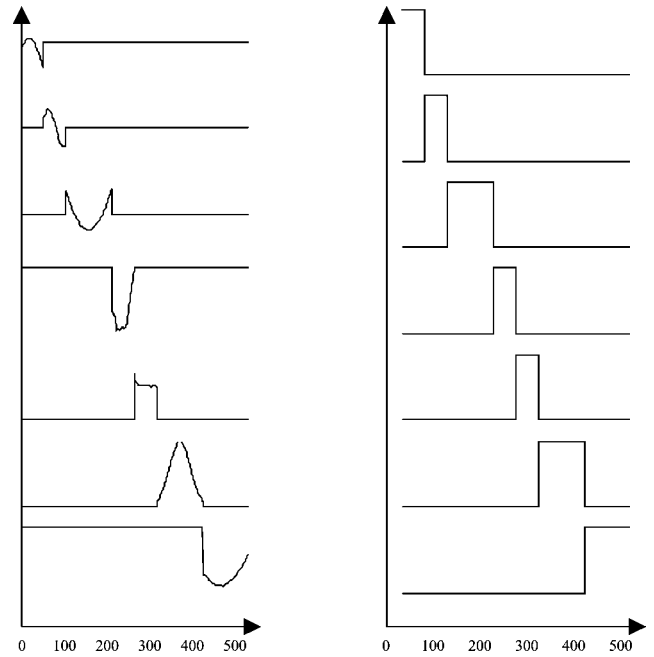

Fig. 5 *f_i (left) and S_i (right) functions for $i = 1$ to 7 (top to bottom)*

$\mu = 5$, and Fig. 4 (bottom) shows $f(t)$ representing the enhanced signal as the final result.

However, it is interesting to investigate f_i and S_i functions separately, as these functions, which $f(t)$ is composed of, are immediate outputs of the algorithm. Fig. 5 illustrates f_i and S_i individually for the signal considered in Fig. 4. As indicated in Fig. 5, S_is representing the time intervals in which f_is are valid can be regarded as the segmented time intervals of f_is. It is important to note that no prior knowledge about the length of time intervals of f_is is required for this segmentation. As mentioned earlier and also depicted in Figs. 4 and 5, f_is are detected because of the discontinuities in the original signal therefore this method can be considered to preserve these discontinuities.

Fig. 6 depicts a typical square wave signal and its noisy version with an additive Gaussian noise (SNR = 5.42). The algorithm described in this paper is applied to the noisy signal of Fig. 6. As shown in Fig. 7, the enhanced signals using different values of μ (=10, 20, 50, 100 from top to bottom (left) in Fig. 7, respectively) in (1) and (7) are calculated. As seen from the figure, the higher the μ, the smoother the enhanced signal is, while preserving discontinuities specific to the original noiseless signal. A
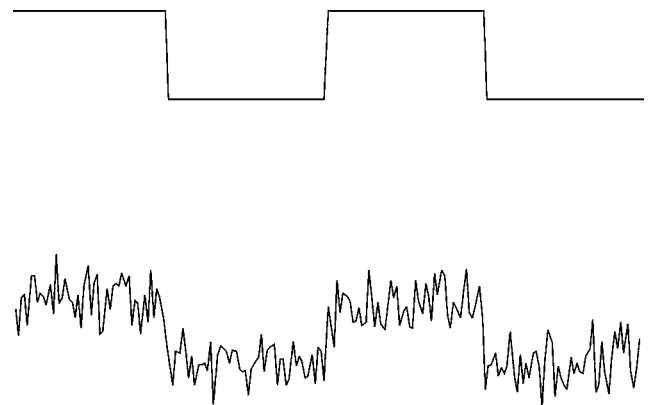

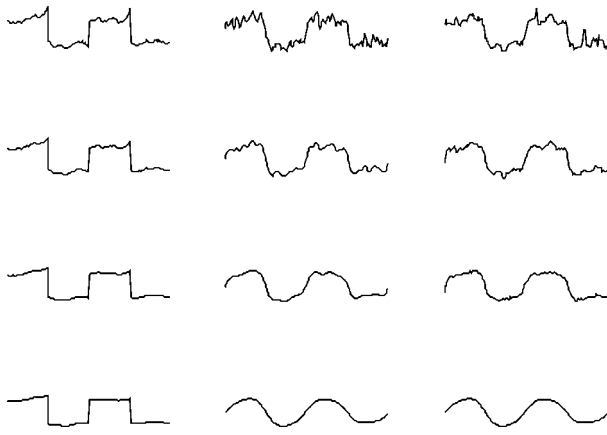**Fig. 6** *Square wave signal and its noisy version (SNR = 5.42)*

Fig. 7 *Enhanced signals using the optimisation method (left) with μ = 10, 20, 50 and 100 and T = 530 (top to bottom) and a Gaussian low-pass filter (middle) with bandwidths 4π/5, 2π/5, π/5, and π/10 (top to bottom) and a median filter (right) with orders 5, 10, 20 and 50 (top to bottom)*

Gaussian low-pass filter with different bandwidths is applied to the noisy signal of Fig. 6 and the results for bandwidths $4\pi/5$, $2\pi/5$, $\pi/5$ and $\pi/10$ are shown in Fig. 7 (middle). As can be observed from this figure, a linear system such as Gaussian filter reduces the noise but smoothes the discontinuities. The same kind of problem is observed in median filter applied to the noisy signal of Fig. 6. A high-order median filter reduces the noise; however, it leaves the discontinuity points fuzzy, and a low-order median filter preserves the discontinuity points but fails to reduce the noise in the signal (Fig. 7, right). The main advantage of the proposed non-linear method compared to common linear and non-linear filters such as Gaussian and median filters discussed above is that it simultaneously segments and smoothes the original signal. Therefore it can preserve the discontinuities.

The enhanced signals, shown at the bottom row of Fig. 7, achieved using the three methods are compared in more detail in Fig. 8. As observed from this Figure, the enhanced signal obtained using the optimisation method with $\mu = 100$ best approximates the original noiseless signal compared to the Gaussian low-pass and median filters.

The difference signal between the original noisy signal and the enhanced signals in Fig. 8 is depicted in Fig. 9. Errors for different methods used in this paper can also be calculated by taking the sum of the absolute difference between the enhanced signals and the original noiseless
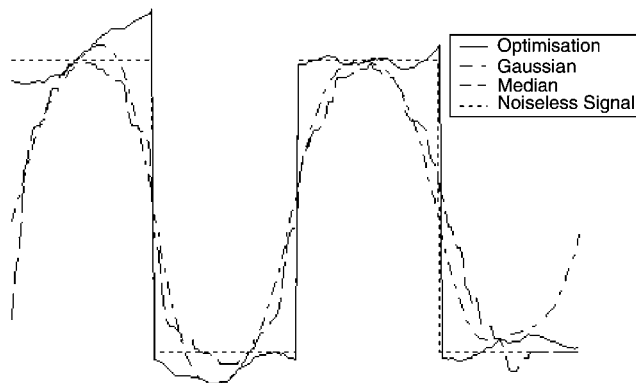

Fig. 8 *Enhancement of noisy signal of Fig. 7 using median filter with order 50, a Gaussian low-pass filter with bandwidth π/10 and the optimisation method described in this paper with μ = 100, and the original noiseless signal plotted for comparison*

Fig. 9 *Noise added to the original noiseless signal (a), the difference signal between the original noisy signal and the enhanced signal using the optimisation method with μ = 100 (b), using Gaussian low-pass filter with bandwidth π/10 (c), using median filter with order 50 (d)*

Table 1: Total error, error per sample of the enhanced signals (Fig. 8) and variance of the difference signals (Fig. 9) for the three methods discussed in this paper

Method	Total error	Error per sample	Variance of difference signal in Fig. 9
Optimisation	978.09	4.8905	1019.7
Low-pass filter	3135.9	15.6795	1313.4
Median filter	2768.2	13.8410	1343.6

signal to quantify the comparison. The errors are further divided by the number of samples to indicate the average error per sample for the different methods considered in this study. Further the variance of each difference signal is calculated. Table 1 shows the results. As observed from the table, the variance of difference signal obtained from the optimisation method is closest to the actual variance (actual variance = 1089).

Finally, we propose an adaptive algorithm to automatically find the optimised μ for a signal contaminated with homogeneous Gaussian noise with variance σ^2. Once the signal is segmented (and simultaneously smoothed) with an initial μ using the proposed algorithm discussed in Section 3, the mean-square error between smoothed signal and original noisy signal is then calculated and μ is changed accordingly so that the mean-squared error term approaches the variance of noise, that is, if with initial μ, the mean-squared error term is less than the variance of
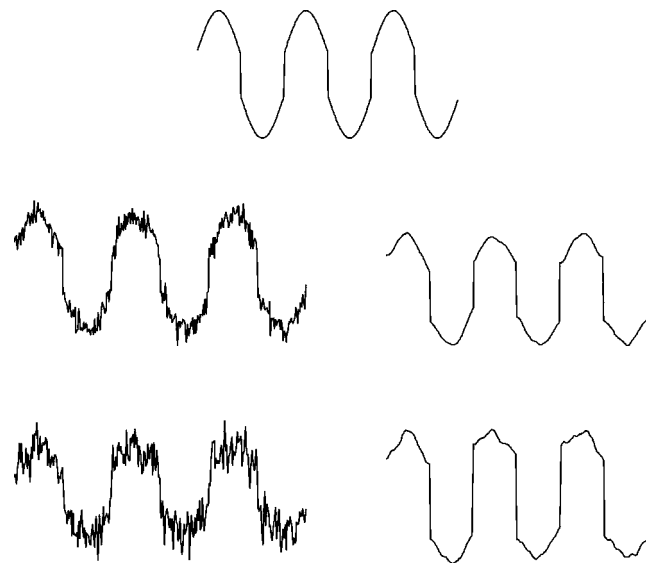

Fig. 10 *Original noiseless signal (top), noisy signal with SNR = 37.8, smoothed signal with μ = 12 obtained using the adaptive algorithm (middle) and noisy signal with SNR = 9.9 and smoothed signal with μ = 34 obtained using the adaptive algorithm (bottom)*

noise, μ increases; otherwise it decreases, until the mean-squared error term approaches the variance of noise. The typical results of this adaptive method are presented in Fig. 10. As can be seen from the figure, different values of SNR result in different values in μ for the most optimised solution. It should be noted that the proposed method in this paper iteratively detects and smoothes noisy signals, so that the number of iterations depends on the number of samples as well as on the discontinuities. Therefore this algorithm is slower than the conventional low-pass Gaussian and median filters.

5 Conclusions

The optimisation method discussed in this paper presents a novel scheme for noise reduction and time interval segmentation. The study in this paper indicates that this method can achieve a significant noise reduction while preserving discontinuities specific to the signal. The results obtained in this work show that the proposed optimisation method performs better than the traditional methods such as low-pass and median filters.

The finite-difference method has been used in this study for implementation because of its numerical efficiencies. Other interpolation schemes such as spline transform could also be employed; however, the spline transform is more computationally expensive due to requirements for control-point calculations. Implementation of this optimisation problem using spline transform and numerical comparison between spline transform and finite difference could therefore be a subject for future study. Implementation of this method for segmentation and noise reduction of 2D and 3D images can be another interesting subject for future research.

6 Acknowledgments

7 References

1 Mumford, D., and Shah, J.: 'Optimal approximations by piecewise smooth functions and associated variational problems', *Comm. Pure. Appl. Math.*, 1989, **42**, (4), pp. 577–688
2 Kass, M., Witkin, A., and Terzopoulos, D.: 'Snakes: active contour models', *Int. J. Comput. Vis.*, 1987, **1**, pp. 321–331
3 Cohen, L.D.: 'On active contour models and balloons', *CVGIP, Image Underst.*, 1991, **53**, (2), pp. 211–218
4 Tsai, A., Yezzi, A., and Willsky, A.S.: 'Curve evolution implementation of the Mumford–Shah functional for image segmentation, denoising, interpolation and magnification', *IEEE Trans. Image Process.*, 2001, **10**, (8), pp. 1169–1186
5 Chan, T.F., and Vese, L.A.: 'Active contours without edges', *IEEE Trans. Image Process.*, 2001, **10**, (2), pp. 266–277
6 Mahmoodi, S., Sharif, B.S., Chester, E.G., Owen, J.P., and Lee, R.E.J.: 'Skeletal growth estimation using radiographic image processing and analysis', *IEEE Trans. Inf. Technol. Biomed.*, 2000, **4**, (4), pp. 292–297
7 Rudin, L.I., Osher, S., and Fatemi, E.: 'Nonlinear total variation based noise removal algorithms', *Physica D*, 1992, **60**, pp. 259–268
8 Sethian, J.A.: 'Level set methods: evolving interfaces in geometry, fluid mechanics, computer vision and material science' (Cambridge University Press, 1996)
9 Caselles, V., Kimmel, R., and Sapiro, G.: 'Geodesic active contours'. Proc. 5th Int. Conf. on Computer Vision, IEEE Computer Society Press, 1995, pp. 694–699
10 Caselles, V., Kimmel, R., and Sapiro, G.: 'Geodesic active contours', *Int. J. Comput. Vis.*, 1997, **22**, (1), pp. 61–79
11 Sapiro, G.: 'Geometric partial differential equations and image analysis' (Cambridge University Press, 2001)
12 Aubert, G., and Kornprobst, P.: 'Mathematical problems in image processing: partial differential equations and calculus of variations' (Springer-Verlag, New York, 2002)
13 Richardson, T.J.: 'Scale independent piecewise smooth segmentation of images via variational methods'. PhD dissertation, Dept. Elect. Eng. Comput. Sci., MIT, 1989
14 Gelb, A.: 'Detection of edges in spectral data', *Appl. Comput. Harmonic Anal.*, 1999, **7**, pp. 101–135
15 Gelb, A., and Tadmor, E.: 'Detection of edges in spectral data. II. Nonlinear enhancement', *SIAM J. Numer. Anal.*, 2000, **38**, (4), pp. 1389–1408
16 Gelb, A.: 'A hybrid approach to spectral reconstruction of piecewise smooth functions', *J. Sci. Comput.*, 2001, **15**, pp. 293–322
17 Gelb, A., and Tadmor, E.: 'Spectral reconstruction of piecewise smooth functions from their discrete data', *Math. Model. Numer. Anal.*, 2002, **36**, (2), pp. 155–175
18 Archibald, R., and Gelb, A.: 'A method to reduce the Gibbs ringing artefact in MRI scans while keeping tissue boundary integrity', *IEEE Med. Imaging*, 2002, **21**, (4), pp. 305–319
19 Kreyszig, E.: 'Advanced engineering mathematics' (John Wiley, 1999)
20 Manderscheid, U.B.: 'Introduction to the calculus of variations' (Chapman & Hall, 1991)
21 Unser, M., Aldroubi, A., and Eden, M.: 'B-spline signal processing: Part I – Theory', *IEEE Trans. Signal Process.*, 1993, **41**, (2), pp. 821–833
22 Unser, M., Aldroubi, A., and Eden, M.: 'B-spline signal processing: Part II – Efficient design and applications', *IEEE Trans. Signal Process.*, 1993, **41**, (2), pp. 834–848
23 Rogers, D., and Adams, J.A.: 'Mathematical elements for computer graphics' (McGraw-Hill Publishing Company, 1990)
24 Unser, M., Aldroubi, A., and Eden, M.: 'Fast B-spline transforms for continuous image representation and interpretation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991, **13**, pp. 277–285
25 Ortega, J.M.: 'A introduction to numerical methods for differential equations' (Pitman Publications, 1981)
26 Anderson, J.D. Jr.: 'Computational fluid dynamics: the basics with applications' (McGraw-Hill International Editions, 1995)
27 Proakis, J.G., and Manolakis, D.G.: 'Digital signal processing principles, algorithms, and applications' (Prentice-Hall International Limited, 1996)