

Fast communication

A nonlinear variational method for signal segmentation and reconstruction using level set algorithm

Sasan Mahmoodi^{a,*}, Bayan S. Sharif^b

^aPsychology Department, School of Biology, Henry Wellcome Building Newcastle University, Newcastle upon Tyne NE2 4HH, UK

^bSchool of Electrical, Electronic and Computer Engineering, Merz Court, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

Received 14 October 2005; received in revised form 19 June 2006; accepted 23 June 2006

Available online 18 July 2006

Abstract

A nonlinear functional is considered in this letter for segmentation and noise removal of piecewise continuous signals containing binary information contaminated with Gaussian noise. A discontinuity is defined as points in time scale that separates two signal segments with different amplitude spectra. Segmentation and noise removal of a piecewise continuous signal are obtained by deriving equations minimising the nonlinear functional. An algorithm based on the level set method is employed to implement the solutions minimising the functional. The proposed method is robust in noisy signals and can avoid local minima.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Level set method; Signal segmentation; Signal reconstruction; Nonlinear optimisation

1. Introduction

Variational methods in image and signal processing have attracted considerable attention in recent years. Segmentation using variational methods was initially introduced by Kass et al. [1] and was further developed as the Geodesic Active Contours Model and the Level-Set Method [6]. On the other hand, Mumford et al. [12] introduced a nonlinear functional for segmentation and smoothing of images which was modelled as piecewise continuous functions that are surrounded by discontinuities represented by contours. This functional was further approximated for implementation using various approaches (e.g. see

[2–6,13–15]). However, it is well known that the optimal solution of nonlinear functionals considered in [12] is not unique [15] which results in the algorithms falling into local minima. A nonlinear functional similar to the one proposed in [12] was also considered by Mahmoodi et al. [8–11] for signal and image segmentation and noise reduction. In [9], the optimal solution to this functional for continuous and discrete signals is obtained and an algorithm based on this solution is proposed. A geometrical approach is employed in [8,10] to implement this functional for signal and image segmentation which leads to a robust algorithm in a noisy environment. Two terms are introduced in the functional considered in [8–10] acting on the desired signal (or image): (1) fidelity which causes the desired signal to be as close as possible to the original noisy signal (2) smoothing the desired signal in proportion to a

*Corresponding author. Tel.: +44 191 222 6246; fax: +44 191 222 5622.

E-mail address: sasan.mahmoodi@ncl.ac.uk (S. Mahmoodi).

coefficient embedded in the functional. The competing terms operate over a time segment where the original signal is considered continuous. It is demonstrated in [9] that points corresponding to discontinuities are minimisers of the functional. Therefore, the minimisation of such a functional leads to a unique solution representing the segmented and the smoothed signals. However, the smoothing term in [8–10] introduces a low-pass filter whose bandwidth is related to the coefficient of the smoothing term. In this case, the desirable frequencies of a band pass input signal may be filtered by the smoothing term leading to the distortion of the smoothed signal. To overcome this problem, a modification to the functional considered in [8–10] is proposed in this letter to enable the algorithm to segment and reconstruct band pass signals containing high-frequency fluctuations. It is demonstrated that this method finds the global minimum by considering only the first variation of the functional. The level set method is employed to implement this functional and the results show that the proposed method is robust for signals with low SNRs. The basic principles are investigated in Section 2 whilst analysis and implementation based on the level set framework are considered in Section 3. Results are presented in Section 4 and finally conclusions are drawn in Section 5.

2. Basic principles

Discontinuities in a signal $g(t)$ are defined as points in time scale separating segments over which the corresponding signals have different amplitude spectrums. If $f_i(t)$ is a smooth continuous function of class C^n $n \geq 2$ in the i th segment, a functional is then considered as

$$E(f, S) = \sum_i \int_{x_{i-1}}^{x_i} [(f_i(x) - g(x))^2] S_i(x) dx, \tag{1}$$

where $S_i(x)$ is a windowed function representing the i th segment in which $g(x)$ has no discontinuities. In order to find the points corresponding to discontinuities, $S_i(x)$ has to be determined so that the i th segment over which there is no discontinuity, is as long as possible. In this letter, we assume that $f_i(x)$ can be written as a linear combination of finite numbers of eigenfunctions of class C^n such as sin and cos functions, i.e.:

$$f_i(x) = \sum_m (A_{im} \cos(m\omega x) + B_{im} \sin(m\omega x)), \tag{2}$$

where $g(x)$ is considered periodic with a base frequency $\omega = 2\pi/T$ and T is the entire time segment over which the whole signal is defined. Restriction imposed by (2) guarantees that f_i is of class C^n over time segment i . A similar functional has been investigated in [4,8,9]; however, a major difference in the functional considered in this letter is that the continuity of f_i for the i th time segment in functional (1) is imposed by the restriction presented in (2), whilst a smoothing term is explicitly introduced in the functional investigated in [4,8,9]. Let us now assume that $g(x)$ is an even periodic function over the entire time length of the input signal, i.e., $g(x) = g(-x)$. This assumption is used to reduce the number of calculated parameters by half without loss of generality. A half range expansion for $f_i(x)$ can then be employed [7], i.e., $B_{im} = 0$ and

$$A_{im} = \int_{x_{i-1}}^{x_i} f_i(x) \cos(m\omega x) dt. \tag{3}$$

Our minimisation problem is now reduced to determine the coefficients A_{im} and function $S_i(x)$ minimising functional (1). In order to find the most optimised A_{im} , functions $S_i(x)$ are initially considered fixed. In the next section, equations leading to these coefficients are obtained in the level set framework. On the other hand, to find the most optimised $S_i(x)$, variations of functional (1) should vanish by changing $S_i(x)$. This requires that we consider the appropriate changes in $f_i(x)$ as well.

3. Analysis and implementation

The proposed model by Chan et al. [4] in the level set formulation involves minimising functional (4):

$$E(f_i, f_o, \varphi) = \int_0^T \left[(f_i(x) - g(x))^2 + \mu \left(\frac{df_i}{dx} \right)^2 \right] H(\varphi(x)) dx + \int_0^T \left[(f_o(x) - g(x))^2 + \mu \left(\frac{df_o}{dx} \right)^2 \right] \times (1 - H(\varphi(x))) dx + \nu \int_0^T \left| \frac{dH(\varphi(x))}{dx} \right| dx, \tag{4}$$

where φ is a Lipschitz function whose zero level set ($\varphi(x) = 0$) known as front, represents discontinuities. $H(\cdot)$ in (4) represents the regularised form of Heaviside function. The associated Euler–Lagrange equations obtained by minimising functional (4)

with respect to f_i, f_o and φ are derived as:

$$\mu \frac{d^2 f_i}{dx^2} = f_i - g \quad \text{for } \varphi > 0, \tag{5}$$

$$\mu \frac{d^2 f_o}{dx^2} = f_o - g \quad \text{for } \varphi < 0, \tag{6}$$

$$\begin{aligned} \frac{\partial \varphi}{\partial t} = \delta_d(\varphi) & \left[v \frac{\partial}{\partial x} \left(\frac{\partial \varphi / \partial x}{|\partial \varphi / \partial x|} \right) - (g - f_i)^2 \right. \\ & \left. + (g - f_o)^2 - \mu \left(\frac{df_i}{dx} \right)^2 + \mu \left(\frac{df_o}{dx} \right)^2 \right], \tag{7} \end{aligned}$$

where $\delta_d(\varphi)$ is the regularised Dirac delta function. Functional (4) is the one-dimensional Mumford–Shah functional in the level set framework. The Mumford–Shah functional and hence Chan–Vese model is based on the calculation of the first variation as discussed above. The solutions of the above differential equations are the segmented, and smoothed signals, i.e., discontinuities are the solutions of differential equation (7). Let us now assume a point at $x = x_p$ corresponding to $\varphi(x_p) = 0$ which is not in a neighbourhood of any discontinuity. In this point, $f_i(x_p) = f_o(x_p)$ and $df_i/dx|_{x=x_p} = df_o/dx|_{x=x_p}$, since we assume that $g(x)$ is piecewise continuous of the class $C^n, n > 2$. It should also be noted that at such a point in the front $\frac{\partial \varphi}{\partial x} / |\frac{\partial \varphi}{\partial x}| = 1$ or -1 and hence $\partial(\frac{\partial \varphi}{\partial x} / |\frac{\partial \varphi}{\partial x}|) / \partial x = 0$. Therefore, at $x = x_p$ Eq. (7) vanishes. This analysis indicates that the point at $x = x_p$ is a local minimum of functional (4). In fact, the reason that the Chan–Vese model is trapped in local minima in piecewise continuous cases is due to an intrinsic feature in their model. This feature is introduced by treating Eq. (7) globally to detect discontinuities, whereas this equation should be treated as an equilibrium equation which is approximated in a neighbourhood of discontinuities [9]. The model proposed in [8,9] considers the second variation of a functional similar to (4) with respect to the fronts. It is proved in [9] that discontinuities are the minimisers of the functional and points away from any discontinuities that are saddle points of the functional and therefore should be discarded. In this short letter, however, we demonstrate that it is also possible to avoid local minima by considering only the first variation in the cases where a signal is composed of a finite number of frequency components so that each time segment contains a number of complete cycles of each frequency. The advantage of using only the first

variation is that less numerical complexity is introduced, since the second variation calculations are not required.

Let us now consider functional (1) in the level set framework:

$$\begin{aligned} E(A_i, A_o, \varphi) &= \alpha \int_0^T \left[\left(\sum_m A_{im} \cos(m\omega x) - g(x) \right)^2 \right] H(\varphi(x)) dx \\ &+ \beta \int_0^T \left[\left(\sum_m A_{om} \cos(m\omega x) - g(x) \right)^2 \right] \\ &\times (1 - H(\varphi(x))) dx, \tag{8} \end{aligned}$$

where φ is a Lipschitz function whose front ($\varphi(x) = 0$) represents discontinuities. $H(\cdot)$ in (8) represents the regularised Heaviside function and therefore $H(\varphi)$ and $1 - H(\varphi)$ for $\varphi(x) > 0$ represent intervals over which signals are characterised with different amplitude spectrums. Functional (8) in the level set formulation is equivalent to functional (1) with the restriction indicated in (2). Optimal values for the coefficients in (2) are obtained by vanishing the derivative of E with respect to A_{ip} for $p \geq 0$ (i.e., $\partial E / \partial A_{ip} = 0$).

$$\begin{aligned} & \int_0^T \cos(p\omega x) g(x) H(\varphi(x)) dx \\ &= \int_0^T \sum_m A_{im} \cos(p\omega x) \cos(m\omega x) H(\varphi(x)) dx. \tag{9} \end{aligned}$$

A similar linear system to (9) can be derived for A_{om} . Functional (8) is also minimised with respect to the signed distance function φ . This therefore leads to the following Euler–Lagrange equation:

$$\begin{aligned} \frac{\partial \varphi}{\partial t} = \delta_d(\varphi) & \left[-\alpha \left(g - \sum_m A_{im} \cos(m\omega x) \right)^2 \right. \\ & \left. + \beta \left(g - \sum_m A_{om} \cos(m\omega x) \right)^2 \right], \tag{10} \end{aligned}$$

where $\delta_d(\varphi)$ is the regularised Dirac delta function. Differential equation (10) coupled with a linear system such as (9) for different time intervals should be solved iteratively. φ and hence discontinuities are iteratively evolved to converge to the desired solution. Upon convergence, the discontinuities corresponding to $\varphi(x) = 0$ represent the segmented signal (segmentation) and Fourier coefficients A_{im} and A_{om} are used to reconstruct the signal (noise

reduction). For the simplicity of the analysis in this short letter, we assume that signal $g(x)$ consists of only one frequency component with two different amplitudes. As demonstrated in the next section, the numerical simulations indicate that the proposed algorithm in this letter can be applied to signals with

any number of frequency components. However, a mathematical argument for signals containing multiple frequencies is not presented here. It is noted that the proposed method in this letter is only applicable to signals with time segments containing a number of complete cycles of a frequency or a set

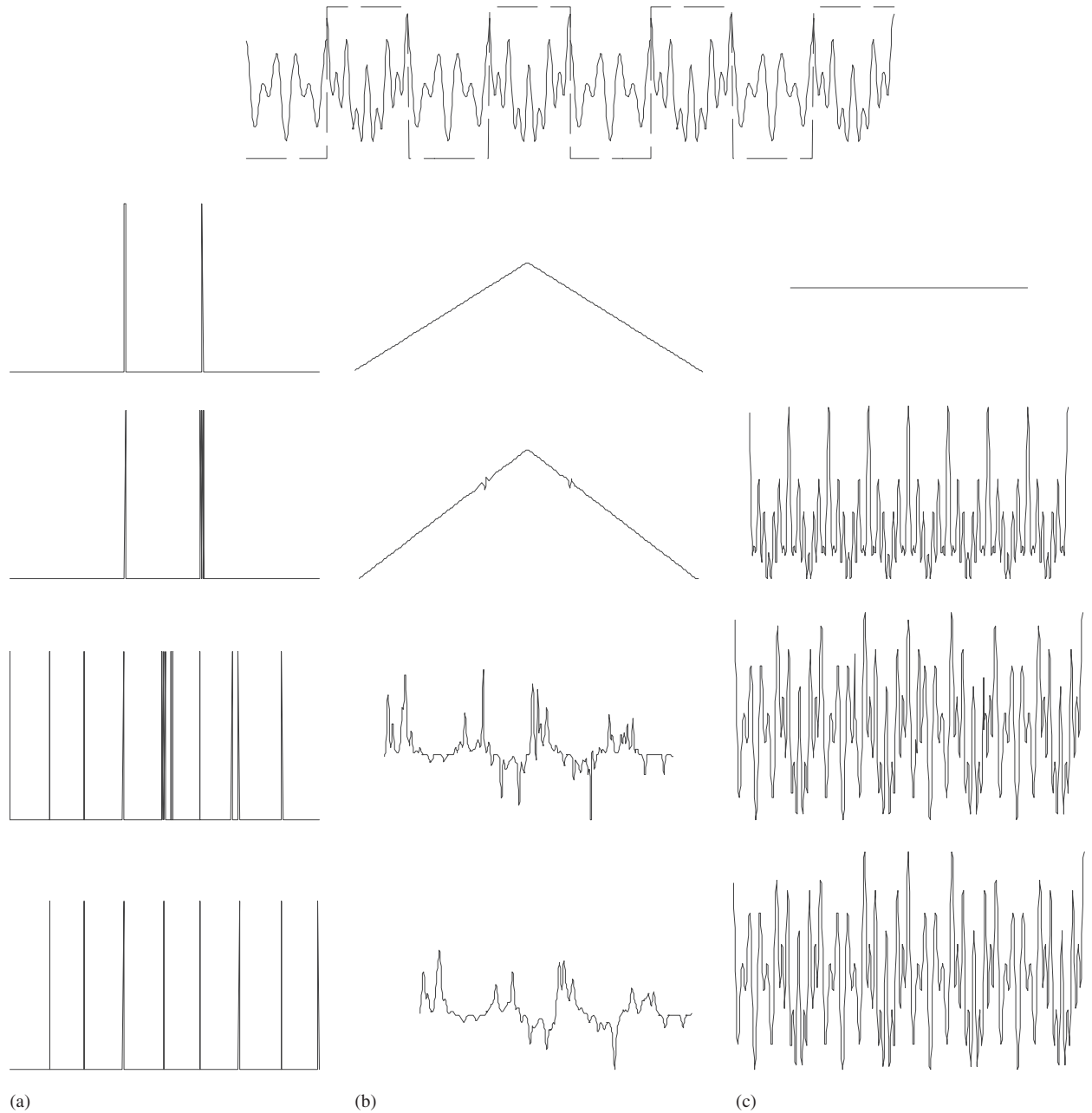
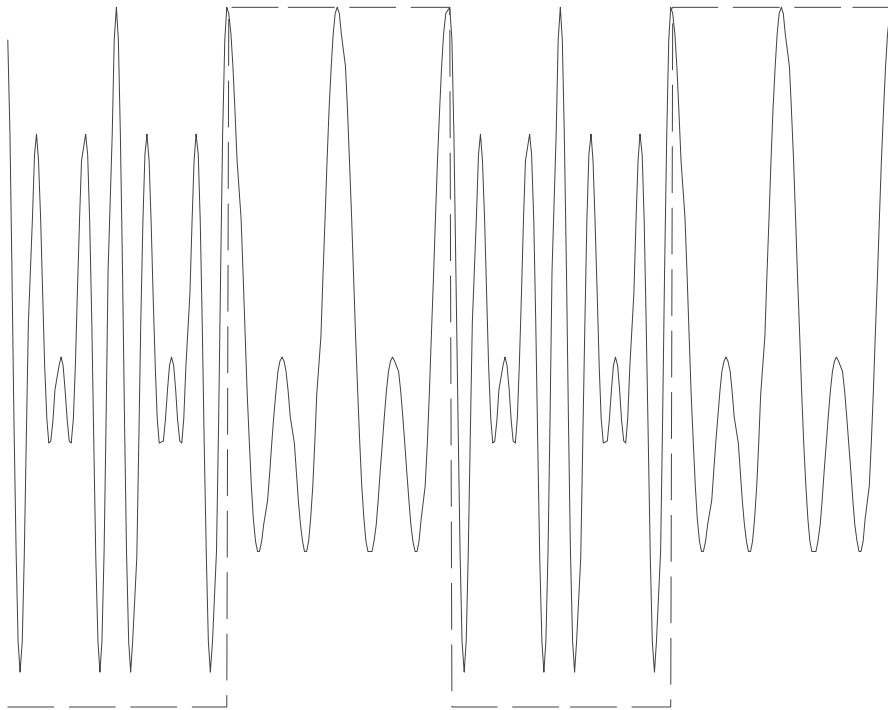
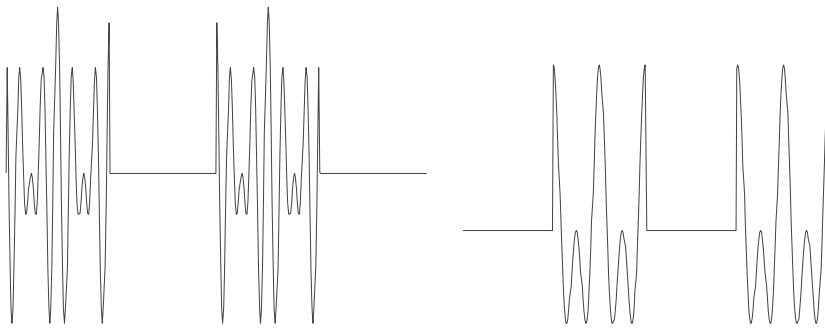


Fig. 1. Original signal (first row) front evolution in iterations 1 (second row) 2 (third row) 8 (fourth row) 12 (fifth row). Columns (a)–(c) show the segmented signal, the level set function φ and the reconstructed signal in each iteration, respectively. The bandwidth of the low-pass filter is set to $\pi/3$.



(a)



(b)



(c)

Fig. 2. Original noiseless signal (a) segmented and reconstructed signal using the proposed method in this letter (b) segmented and reconstructed signals using the methods of differential equations considered previously in literature with $\mu = 10$ and $\nu = 10$ (e.g., [4]) (c).

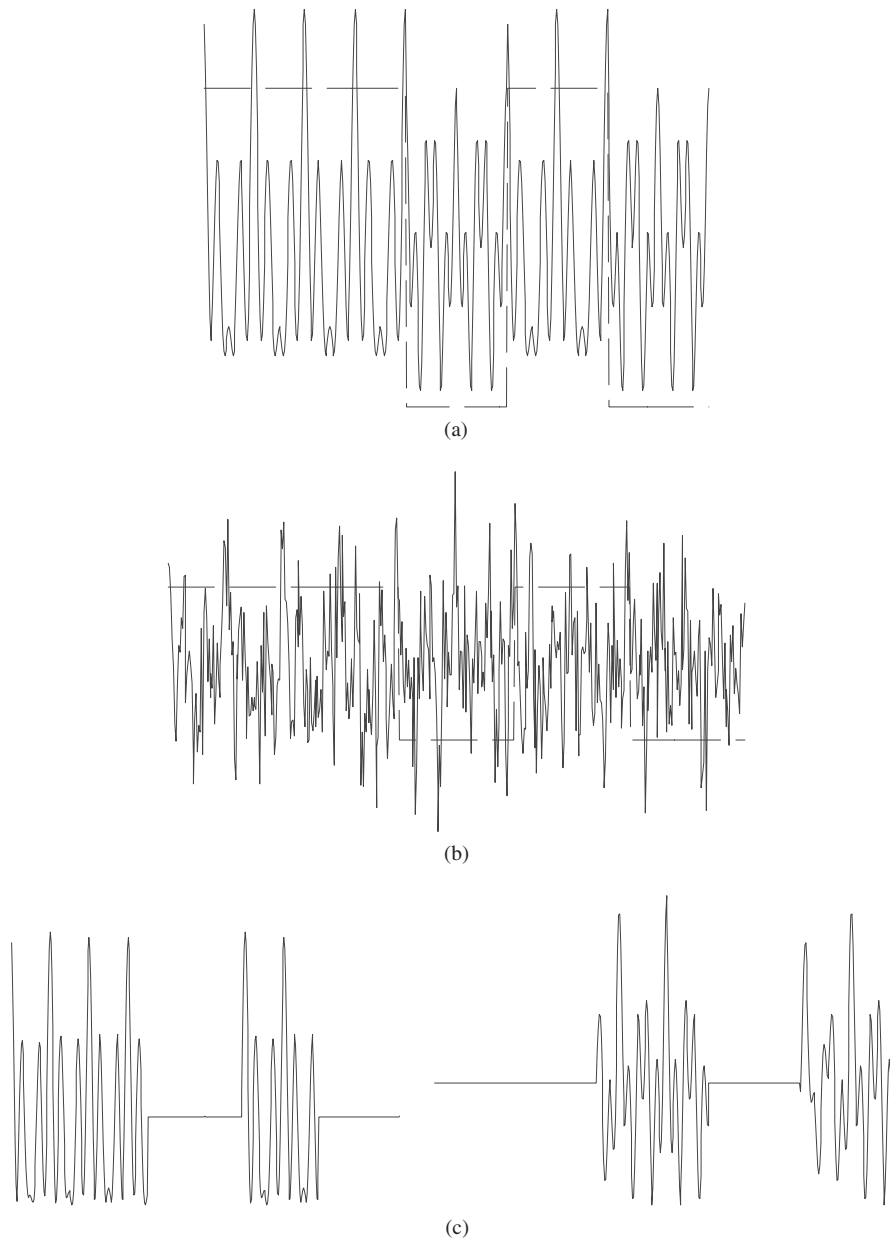


Fig. 3. Original noiseless signal (a) noisy signal with SNR = 1 (b) and segmented and reconstructed signals with the bandwidth of the low-pass filter = $\pi/8$ (c).

of frequencies. A signal with a single carrier frequency is usually considered in communications. In this case, Eq. (10) is written as

$$\frac{\partial \varphi}{\partial t} = \delta_d(\varphi) [-\alpha(g - A_i \cos(\omega_0 x))^2 + \beta(g - A_o \cos(\omega_0 x))^2]. \quad (11)$$

Coefficients A_i and A_o are then calculated as

$$A_i = \frac{\int_0^T \cos(\omega_0 x) g(x) H(\varphi) dx}{\int_0^T \cos^2(\omega_0 x) H(\varphi) dx}, \quad (12)$$

$$A_o = \frac{\int_0^T \cos(\omega_0 x) g(x) (1 - H(\varphi)) dx}{\int_0^T \cos^2(\omega_0 x) (1 - H(\varphi)) dx}. \quad (13)$$

In a signal with only one discontinuity, we assume a point P at $x = x_p$ which is not in a neighbourhood of a discontinuity at $x = x_d$ so that $x_d > x_p$. It is also assumed that point P is in the time domain over which the sinusoidal signal has amplitude G_i . The other part of signal is characterised with a sinusoidal signal with amplitude $G_o \neq G_i$. From the above equations, A_i and A_o are calculated as

$$A_i = \frac{\int_0^{x_p} \cos(\omega_0 x)(G_i \cos(\omega_0 x)) dx}{\int_0^{x_p} \cos^2(\omega_0 x) dx} = G_i \quad (14)$$

$$\begin{aligned} A_o &= \frac{\int_{x_p}^{x_d} \cos(\omega_0 x)(G_i \cos(\omega_0 x)) dx + \int_{x_d}^T \cos(\omega_0 x)(G_o \cos(\omega_0 x)) dx}{\int_{x_p}^T \cos^2(\omega_0 x) dx} \\ &= \frac{G_i I_0 + G_o I_1}{I_0 + I_1} \end{aligned} \quad (15)$$

or

$$A_o = G_o + (G_i - G_o) \frac{I_0}{I_0 + I_1}, \quad (16)$$

where $I_0 = \int_{x_p}^{x_d} \cos^2(\omega_0 x) dx$ and $I_1 = \int_{x_d}^T \cos^2(\omega_0 x) dx$. Therefore, Eq. (11) can be written as

$$\frac{\partial \varphi}{\partial t} = \delta_d(\varphi) \left[\beta \left(\frac{(G_i - G_o)^2 I_0^2}{(I_0 + I_1)^2} \cos^2(\omega_0 x) \right) \right]. \quad (17)$$

The right-hand side of Eq. (17) corresponds to a force driving the front ($\varphi = 0$) towards the discontinuity. This driving force exists as long as the front does not correspond to the discontinuity. However, in a discontinuity $x_p = x_d$ hence $I_0 = 0$ implying $A_i = G_i$ and $A_o = G_o$. Therefore, Eq. (11) representing the first variation vanishes only in a discontinuity. In the presence of noise, Eq. (11), still vanishes, although the front might be slightly shifted with respect to the discontinuity in the original signal. This analysis demonstrates that a point corresponding to discontinuity is a minimiser of functional (8). This property prevents the algorithm from falling into a local minimum. The above analysis applies to a signal with any number of discontinuities, i.e., if a segment of the input signal is in the ‘wrong’ part of φ ($\varphi > 0$), it produces a negative term in Eq. (10) and hence creates two new discontinuity points (or zero crossings in φ). This argument is also true for the case where the ‘wrong’ part of φ corresponds to $\varphi < 0$. This fact is due to the well-known property of the level set formulation based on which the algorithm can be adjusted to the topology of the

input data. It is also noted that the minimisation of functional (8) leads to the segmentation and reconstruction of signals with binary information where the original signal consists of a time sequence of two types of signals with different amplitude spectra. The described evolution procedure occurs only if the signed distance function is correctly initialised. This is because the right side in iterative equation (10) should be comparable to the initial signed distance function to be able to change the level set function for the evolution of the

algorithm. Throughout this letter, the signed distance function $\varphi(x)$ is initialised as

$$\varphi(x) = \frac{1}{4} - \frac{2}{T} \left| x - \frac{T}{2} \right|, \quad (18)$$

where T is the total length of the input signal. The above signed distance function initially corresponds to only two discontinuities. However during the evolution of the algorithm, the number of discontinuities changes to match the number of discontinuities in the input signal $g(x)$. In this letter α and β in Eq. (10) are set to 1. These coefficients can, however, be set to a value corresponding to the maximum absolute amplitude of the input signal. For signals with low absolute amplitude, higher values for these coefficients should be chosen to allow Eq. (10) to evolve. The algorithm proposed in this letter, is summarised in pseudo-code below. The signed distance function φ is firstly initialised by using Eq. (18). Differential equation (10) is then solved to update $\varphi(x)$ and hence evolve the fronts representing discontinuities. In each iteration, a one-dimensional Gaussian filter is applied to $\varphi(x)$ to regularise the differential equation (10), which improves the robustness of our algorithm.

1. Initialise φ^0 for $n = 0$

Loop :

2. Compute A_{im} and A_{om}

3. Solve differential equation for φ to obtain φ^{n+1}

4. Apply Gaussian low pass filter to φ^{n+1}

If convergence is not reached,

Then $n = n + 1$ and **go** to step 2;

Otherwise end the **Loop** and **go** to step 5

5. Reconstruct the signal using A_{im} and A_{om}

4. Results

A synthetic signal with eight discontinuities is shown in Fig. 1 (first row). The level set function φ , segmented and reconstructed signals in four different iterations of the algorithm are depicted in the second to the fifth row of Fig. 1. The dashed curves in all figures show the different segments of the signals. In this example, the bandwidth of the Gaussian low-

pass filter is set to $\pi/3$. Fig. 2 compares the proposed method for segmentation and reconstruction of signals with a method in which the optimised functional contains an explicit smoothing term (see e.g. [4]). The original signal is shown in Fig. 2a. The proposed method in this letter is applied to the original signal and the segmented and reconstructed signals are presented in Fig. 2b. The segmented and reconstructed signals using the traditional methods in

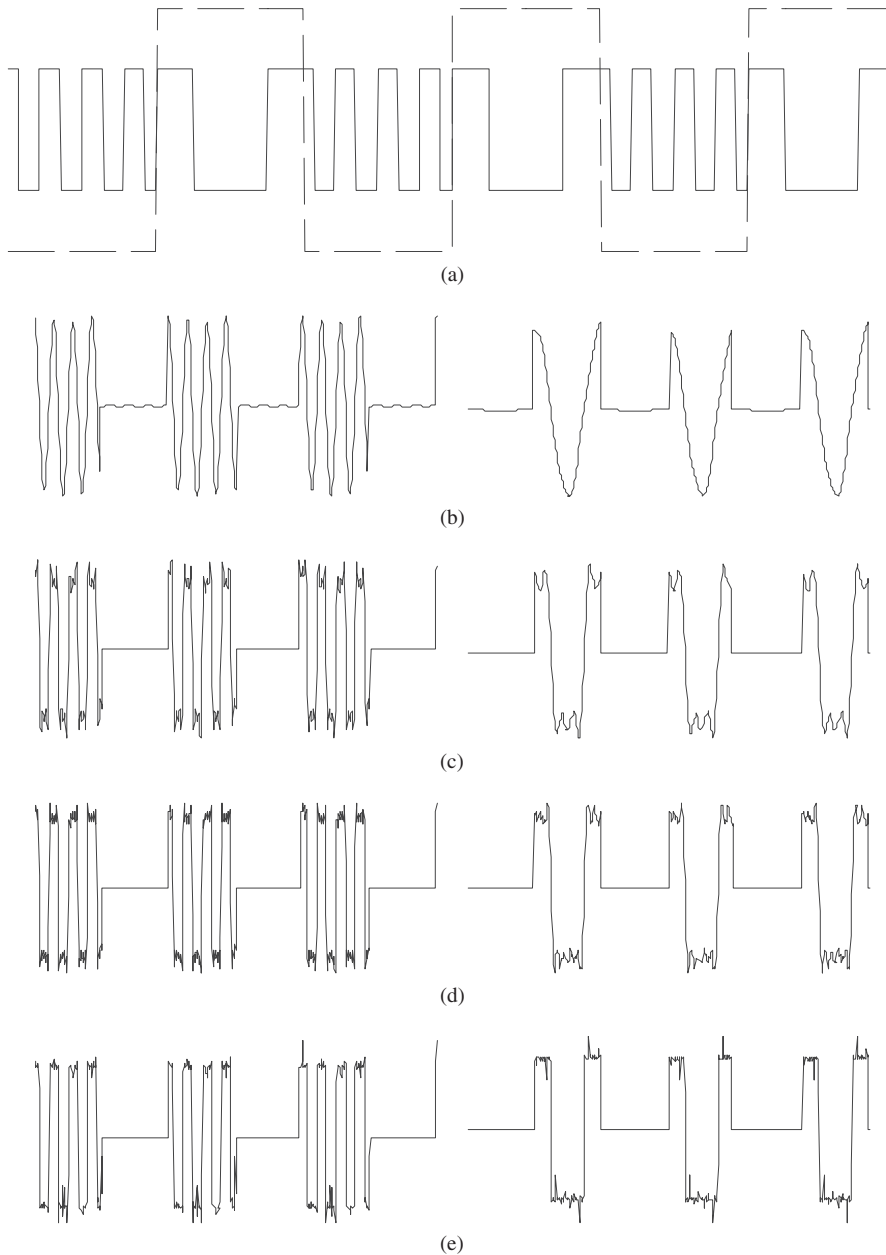


Fig. 4. Original signal (a) segmented and reconstructed signals with one (b) three (c) five (d) and 30 (e) harmonics, the bandwidth of the low-pass filter is set to $\pi/7$.

the literature (e.g. [4]) are depicted in Fig. 2c. As can be seen from this figure, the Chan–Vese algorithm converges to a local minimum, i.e., segmented and reconstructed signals do not correspond to the original signal. In this example, $\mu = 10$ and $\nu = 10$. If μ increases, the reconstructed signal becomes smoother and hence more distorted, whereas if the value of μ is reduced, the algorithm over-segments the signal. This problem becomes more severe when noise is added to the original signal. In the third example, a synthetic noiseless signal depicted in Fig. 3a is contaminated with a Gaussian noise to produce the noisy signal shown in Fig. 3b with $\text{SNR} = 1$. The segmented and reconstructed signals are demonstrated in Fig. 3c. A signal containing several discontinuities with two different frequencies is shown in Fig. 4. In this example, we aim to show that the proposed method can segment and reconstruct the signal based on discontinuities in frequency rather than those in the signal itself. Fig. 4a shows the original signal. As shown in this figure, the signal is composed of two square waves with different frequencies. The segmentation is achieved based on the first harmonic. Having segmented the signal, more harmonics can be employed for reconstruction. The segmented and reconstructed signals using only the first frequency harmonic is shown in Fig. 4b, and using three, five and 30 harmonics in Figs. 4c–e, respectively. It is interesting to note that in this example discontinuous signals are reconstructed by using a set of continuous eigenfunctions. The fluctuations and oscillations in the reconstructed signals in Fig. 4e are due to Gibbs phenomenon. Segmentation and reconstruction of signals of the type shown in Fig. 4a are not possible by minimising any functional containing an explicit smoothing term of the desired reconstructed signal, since the desired signal contains discontinuities.

5. Conclusion

A new method is proposed in this letter to find the global minimum in a minimisation problem leading to signal segmentation and noise reduction of a given noisy signal. A functional is considered for this purpose and continuity of each segment of the signal is implicitly imposed in comparison with the previous functionals in which continuity was explicitly presented. Numerical comparison has been made between the proposed method and the traditional method. The level set method is employed to implement this functional. The algorithm

presented in this letter converges very easily, is robust in the presence of noise and can reconstruct desired signals containing discontinuities.

Acknowledgements

Authors would like to thank the anonymous reviewers for their constructive and interesting comments.

References

- [1] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, *Internat. J. Comput. Vis.* 1 (1988) 321–331.
- [2] T.F. Chan, B.Y. Sandberg, L.A. Vese, Active contours without edges for vector-valued images, *J. Vis. Commun. Image Representation* 11 (2000) 130–141.
- [3] T.F. Chan, L.A. Vese, Active contours without edges, *IEEE Trans. Image Process.* 10 (2) (2001) 266–277.
- [4] T.F. Chan, L.A. Vese, Level set algorithm for minimising the Mumford–Shah functional in image processing, in: *Proceedings IEEE Workshop on Variational and Level Set Methods in Computer Vision*, 2001, pp. 161–168.
- [5] M. Hintermuller, W. Ring, An inexact-CG-type active contour approach for the minimization of the Mumford–Shah functional, *J. Math. Imaging Vis.* 20 (2004) 19–42.
- [6] G. Aubert, P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and Calculus of Variations*, Springer, New York, 2002.
- [7] E. Kreyszig, *Advanced Engineering Mathematics*, Wiley, New York, 1999.
- [8] S. Mahmoodi, B.S. Sharif, Signal segmentation and denoising algorithm based on energy optimisation, *Signal Processing* 85 (9) (2005) 1845–1851.
- [9] S. Mahmoodi, B.S. Sharif, Noise reduction, smoothing, and time interval segmentation of noisy signals using an energy optimisation method, *IEE Proc. Vis. Image Signal Process.* 153 (2) (2006) 101–108.
- [10] S. Mahmoodi, B.S. Sharif, Nonlinear optimisation method for image segmentation and noise reduction using geometrical intrinsic properties, *Image Vis. Comput.* 24 (2006) 202–209.
- [11] S. Mahmoodi, B.S. Sharif, Unsupervised texture segmentation using a nonlinear energy optimisation method, *J. Electron. Imaging* 15 (3) (2006) in press.
- [12] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, *Comm. Pure Appl. Math.* 42 (4) (1989) 577–688.
- [13] A. Tsai, A. Yezzi, A.S. Willsky, Curve evolution implementation of the Mumford–Shah functional for image segmentation, denoising, interpolation and magnification, *IEEE Trans. Image Process.* 10 (8) (2001) 1169–1186.
- [14] L.A. Vese, Multiphase object detection and image segmentation, in: S. Osher, N. Paragios (Eds.), *Geometrical Level Set Methods in Imaging, Vision, and Graphics*, Springer, Berlin, 2003, pp. 175–194.
- [15] L.A. Vese, T.F. Chan, A multiphase level set framework for image segmentation using the Mumford and Shah model, *Internat. J. Comput. Vis.* 50 (3) (2002) 271–293.