Robust denoising of electrophoresis and mass spectrometry signals with minimum description length principle

Janne Ojanen[a,*], Timo Miettinen[b], Jukka Heikkonen[a], Jorma Rissanen[a,c]

[a] *Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9400 (Tekniikantie 14), FIN-02150 Espoo, Finland*
[b] *Finnish Genome Center, University of Helsinki, P.O. Box 63 (Haartmaninkatu 8), FIN-00014 Helsinki, Finland*
[c] *Helsinki Institute for Information Technology, Tammasaarenkatu 3, FIN-00180 Helsinki, Finland*

Received 26 March 2004; revised 25 May 2004; accepted 3 June 2004

Available online 19 June 2004

Edited by Veli-Pekka Lehto

Abstract The need for high-throughput assays in molecular biology places increasing requirements on the applied signal processing and modelling methods. In order to be able to extract useful information from the measurements, the removal of undesirable signal characteristics such as random noise is required. This can be done in a quite elegant and efficient way by the minimum description length (MDL) principle, which treats and separates 'noise' from the useful information as that part in the data that cannot be compressed. In its current form the MDL denoising method assumes the Gaussian noise model but does not require any ad hoc parameter settings. It provides a basis for high-speed automated processing systems without requiring continual user interventions to validate the results as in the conventional signal processing methods. Our analysis of the denoising problem in mass spectrometry, capillary electrophoresis genotyping, and sequencing signals suggests that the MDL denoising method produces robust and intuitively appealing results sometimes even in situations where competing approaches perform poorly.
© 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Minimum description length; MDL denoising; Wavelet; Mass spectrometry; Capillary electrophoresis; Sequencing; Genotyping

1. Introduction

Signals in mass spectrometry and capillary electrophoresis measurements result from very complex physical processes. Since the phenomena influencing the characteristics of the signal are generally unknown or far too complex for analytic treatment, building a model on physical grounds is impossible. For these reasons statistical modelling is required. In this it stands reason to expect that the effects of undesirable noise in the measured data must be somehow modelled and removed if we want to extract information about the studied data generating machinery. Also, to fully utilize the high-throughput capabilities of the measuring instruments such as DNA sequencers the data analysis process cannot rely on continual manual editing, and it is highly desirable to do the work by automatic methods.

The denoising problem has been approached from many different directions. Typically one assumes that the underlying signal is contaminated by a certain type of additive or multiplicative noise, characterized by parameters such as the variance. The presence of these parameters introduce uncertainty to the statistical inference and may make an automated process unreliable. The difficulty stems from the fact that these parameters cannot be estimated from the data in a reliable manner, because it is not known which part in the data is the noise! In this paper, we describe the minimum description length (MDL) modelling principle [1], which can be used to denoising in cases where additive type of noise is justified, i.e., the measured signal is the sum of the informative signal and noise. We have chosen to analyze the performance of the MDL denoising method on signals produced by popular instruments such as MALDI-ToF (matrix-assisted laser desorption ionization-time of flight) mass spectrometer and capillary array electrophoresis DNA analyzer. Mass spectrometers (MS) are successfully used for automated detection and characterization of biomolecules, such as oligonucleotides, peptides, proteins, synthetic polymers, and carbohydrates. Capillary electrophoresis (CE) instruments are also routinely used in laboratories for various tasks like DNA sequencing, genotyping, and forensic and carbohydrate analysis. Although the MS and CE instruments are sensitive the resulting signal-to-noise ratio can be poor in many applications due to low reaction volumes. An efficient denoising method enables smaller details to be extracted reliably in high-throughput applications, where cost effectiveness demands minimization of the reaction volumes. In addition to the real life signals we also use synthetic signals to shed further light on the performance of the MDL denoising approach.

2. Materials and methods

2.1. Materials

Two typical sequencing and microsatellite genotyping electropherograms, recorded with MegaBACE 1000, were used to compare a number of denoising methods on capillary electrophoresis data. A special characteristic of such signals is the channel cross-talk phenomenon caused by overlapping spectra of the fluorescent dye labels. The microsatellite electropherogram was recorded from an assay of 16 markers from the ABI Linkage Mapping Set II. The sequencing signal was recorded from a DNA fragment of 491 bases (A: 113, T: 124, C: 120 and G: 133). The color separation matrices for both applications were determined from calibration runs. The MS spectrogram was

* Corresponding author.
E-mail address: janne.ojanen@hut.fi (J. Ojanen).

0014-5793/$22.00 © 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2004.06.022

recorded with Bruker Autoflex(tm) MALDI-ToF mass spectrometer from a typical protein identification measurement.

We also compared different denoising methods on synthetic signals with known noise properties. Since now the actual true form of the signal is known beforehand, the efficiency of each method can easily be assessed quantitatively. The generated synthetic signals were designed to have characteristics similar to the real sequencing signal data. They were generated by the equation

$$\tilde{\mathbf{Y}} = \mathbf{X}\mathbf{Y} + \varepsilon + \beta \tag{1}$$

where the m rows of $\tilde{\mathbf{Y}}$ of length n give the generated data signals, while the rows of $\mathbf{Y}$ consists of m noise free signals with length n. $\mathbf{X}$ is an $m \times m$ mixing matrix. The mixing matrix is called spectral overlap matrix in fluorescent dye electrophoresis applications. The diagonal elements of $\mathbf{X}$ are the relative weights of the signals, while the non-diagonal elements introduce cross-talk effect between signals. The next to last term ε, an $m \times n$ matrix, consist of additive Gaussian noise components, and β represents the baselines.

The synthetic signal we used had $m = 4$ channels, each of length $n = 3000$ samples. In constructing Y a total of 100 Gaussian shape peaks were generated. Each has the shape of a Gaussian density function of maximum height or amplitude I and width defined by variance $\sigma^2 = 2$. The separation d between any two peaks, i.e., the distance between their centers, were generated randomly by sampling a Gaussian density function with mean 30 and variance 5. Similarly the amplitudes of the peaks were generated by sampling another Gaussian density function of mean 1500 and variance 300. Finally the peaks were randomly assigned to different channels. The resulting signal is similar to a sequencing signal caused by four different fluorescent dye labels. The spectral overlap matrix used was

$$X = \begin{pmatrix} 1.000 & 0.087 & 0.184 & 0.117 \\ 0.210 & 1.000 & 0.001 & 0.004 \\ 0.134 & 0.020 & 1.000 & 0.557 \\ 0.003 & 0.026 & 0.258 & 1.000 \end{pmatrix}$$

The signal baselines were set to the constant 200 for all the four channels to which Gaussian noise $\varepsilon_{ij} \sim N(0, \sigma^2)$ was added.

2.2. Methods

Introduction to MDL principle. Modelling is a process to find regular features that the data generating physical machinery has imposed on the data and which collectively define a *model.* A fundamental difficulty in modelling is to select the complexity of the model to be fitted, which in traditional statistics has been dealt with by intuitive ad hoc means, because the idea of complexity has not been formally defined and hence could not be measured. Several such approaches have been suggested for the model complexity selection including for example bootstrapping [2], cross-validation [3], and Bayesian statistics [4]. It is clear that complexity has to do with the number of parameters in the models, but since the influence of the various parameters is not equal the number of parameters alone is not an adequate measure of model complexity.

A drastically different approach to modelling is the MDL principle, which is based on information theory and coding. The basic idea, however, can be explained with only primitive notions of coding. Consider the familiar problem of curve fitting, where the data appear as a cloud of points (y_i, x_i) on a two-dimensional plane, and we would like to pass a 'smooth' polynomially defined curve through the data. The lower degree polynomial we fit the smoother the curve, but what do we mean by an optimally smooth curve? It will clearly have to do with how close the polynomially calculated curve is to the data, which means that we need a distance measure between the curve and the data points, say the quadratic. The sum of the squared differences alone cannot be taken as the criterion to minimize, because the higher degree polynomial we fit, the smaller the sum of the squared differences and the less smooth the curve. Somehow we must 'penalize', as the jargon goes, the too high a degree of polynomial, the overly complex model.

In MDL principle this is done as follows: imagine that we try to describe or encode the observed data efficiently. Clearly, the smaller the squared difference $(y_i - f(x_i))^2$ between the polynomially computed point and the data point is the fewer bits we need to describe the difference $y_i - f(x_i)$ and hence y_i if we know the calculated value $f(x_i)$, which means that we must encode also the polynomial function $f(\cdot)$ used. It can be done by writing its parameters, each a binary number, to a certain precision, and adding all the bits needed. The higher the precision the more bits we need. We can optimize the precision as well as the number of parameters such that the total code length for the data is minimized. The result is the polynomial of optimum degree, which gives the optimally smooth curve in the code length sense.

The MDL principle can be justified on several grounds. For instance, for models as probability distributions it can be shown to provide the maximum probability to the observed data among large families of such models, thereby extending the classical maximum likelihood principle to include the number of parameters as well. The actual codes need not be constructed, because all we need is the code length. More precisely, the code length for the data, given the best model, is the negative logarithm of the probability (also called likelihood) on the data, maximized over the parameters, while the code length of the parameter estimates themselves depends on the precision with which the parameters are written. The optimal precision can be shown to be about $1/\sqrt{n}$ per parameter, where n denotes the number of data points, and it is seen to grow with the number of data points. If we include only the fractional bits in the binary numbers, which for large data sets form the bulk of the number of bits required, the code length for k parameters becomes about $(k/2)\log n$. The total code length, then, to a first order approximation is the sum of the negative logarithm of the maximized likelihood and the term $(k/2)\log n$. The MDL principle calls for minimization of this expression over the number of the parameters and even the structure in which the parameters appear.

The denoising problem addressed in this paper is not unlike the curve fitting problem. Instead of polynomials we write the smooth curve as a linear combination of wavelet basis functions, and the structure then depends on which regressor functions are included in the linear combinations. The quadratic distance measure implies a Gaussian family of density functions for which optimal coding techniques are well known. We can also define the elusive idea of noise as that part in the data that cannot be compressed with the models available, which appears to capture exactly the intuitive idea of noise.

There are other important consequences of the principle. Since the optimal code length for data depends only on the broad properties restricting the data the results are automatically robust: random accidental data points do not affect the result in an essential manner. One cannot shorten the code length by trying to base the code design on random properties, because such properties having no pattern will take a lot of bits to describe.

Conventional filtering methods. A number of different methods have been conventionally applied in denoising problems. Usually noise is described as the high-frequency component of the signal spectrum, and Fourier transform based low-pass filtering methods are used to remove frequencies higher than a suitably selected threshold. However, since the threshold determines the noise characteristics there is a clear danger of prejudice: we remove what we have defined the noise to be.

A common approach to the denoising problem is to employ discrete wavelet transform (DWT) [5]. Like the Fourier transform, DWT can be used to express any function as a linear combination of wavelet basis functions. In the Fourier transform the basis functions are sines and cosines with different frequencies, and such basis functions are very good for periodic signals. The wavelet basis functions capture better non-periodic signals with abruptly changing local variations. Wavelet functions are generated by scaling and translating a single function $\psi(t)$, called mother wavelet. With dyadic dilations and translations of the mother wavelet we then get a family

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{2}$$

where j is the scale index and k the location index.

A wavelet expansion of an arbitrary $n \times 1$ signal y is given by a very simple algorithm, which in effect performs a linear operation of the kind

$$c = \mathbf{W}y \tag{3}$$

where $\mathbf{W}$ is an $n \times n$ matrix of wavelet functions. The signal can be recovered by the inverse transform

$$y = \mathbf{W}^{-1}c \tag{4}$$

There are a number of different wavelet bases, the rows of $\mathbf{W}^{-1}$, defined by the mother wavelet. Even though a certain basis may be preferable for specific applications, any properly defined basis will do. In this paper we have used the symmlet3 wavelets [6], which give the DWT. The matrix $\mathbf{W}$ is orthonormal, which means that $\mathbf{W}^{-1}$ is the transpose $\mathbf{W}^{\mathrm{T}}$.

The coefficient representation (Eq. (3)) can be used in signal processing in such a way that certain coefficients are manipulated before signal reconstruction. In denoising the manipulation amounts to setting to zero the coefficients representing the noise according to a particular criterion, and the inverse transformation (Eq. (4)) of the manipulated data then gives the desired smooth version $\hat{y}$ of the original signal y.

The conventional methods we have selected for comparison are Wiener filtering [7,8], Median filtering [9], and a wavelet based thresholding method by Donoho and Johnstone [10]. A Wiener filter is a stationary linear filter, which is optimal in the mean square error sense. Median filtering, again, is a non-linear signal processing technique for smoothing signals while preserving edges.

The thresholding method by Donoho and Johnstone (further on referred to as DJ method) employs a wavelet transform, and it removes wavelet coefficients with absolute values smaller than a suitably selected threshold

$$\theta = \hat{\sigma}\sqrt{2\ln n} \tag{5}$$

where $\hat{\sigma}^2$ is an estimate of the noise variance and n is the length of the signal. This popular so-called hard thresholding method is designed to remove Gaussian white noise.

The DJ method requires an estimate of the noise variance. This leads to circular reasoning since the way the noise variance is estimated determines the threshold and hence the noise. In other words, the noise gets defined in terms of the main characteristic of the noise. The noise variance $\hat{\sigma}^2$ is usually estimated from the wavelet coefficients with the median absolute deviation (MAD) estimator

$$\hat{\sigma}^2 = \text{median}(\bar{c})/0.6745 \tag{6}$$

where $\bar{c} = \{|c_1|, |c_2|, \ldots, |c_n|\}$ are the absolute values of wavelet coefficients on the first resolution level, and the factor 0.6745 results from a calibration process. Notice that the initial estimate of the noise variance (6) is not equal to the estimated variance of the noise that eventually gets defined through the thresholding process (5).

MDL denoising method. The denoising problem is a special case of the linear quadratic regression problem, in which a wavelet transform creates the coefficients. We rewrite Eq. (4) as

$$y_t = c_1 w_{t,1} + \cdots + c_n w_{t,n}, \quad t = 1, \ldots, n \tag{7}$$

where the wavelet transform algorithm gives the n coefficients c_i, and the n^2 elements $w_{n,t}$ define the regressor matrix $\mathbf{W}$. Then we form the 'smooth' curve $\hat{y}$ as a linear combination of some of the basis functions

$$\hat{y}_t = c_{(1)} w_{t,(1)} + \cdots + c_{(k)} w_{t,(k)}, \quad t = 1, 2, \ldots, n \tag{8}$$

leaving

$$e_t = y_t - \hat{y}_t \tag{9}$$

as noise. Here we denote by (i) the index of the ith largest coefficient in absolute value.

Indeed, one can show that the subset of the regressor functions that minimize the MDL criterion

$$(n-k)\log\frac{\sum_{t=1}^{n} e_t^2}{n-k} + k\log\frac{\sum_{t=1}^{n}\hat{y}_t^2}{k} + \log(k(n-k)) \tag{10}$$

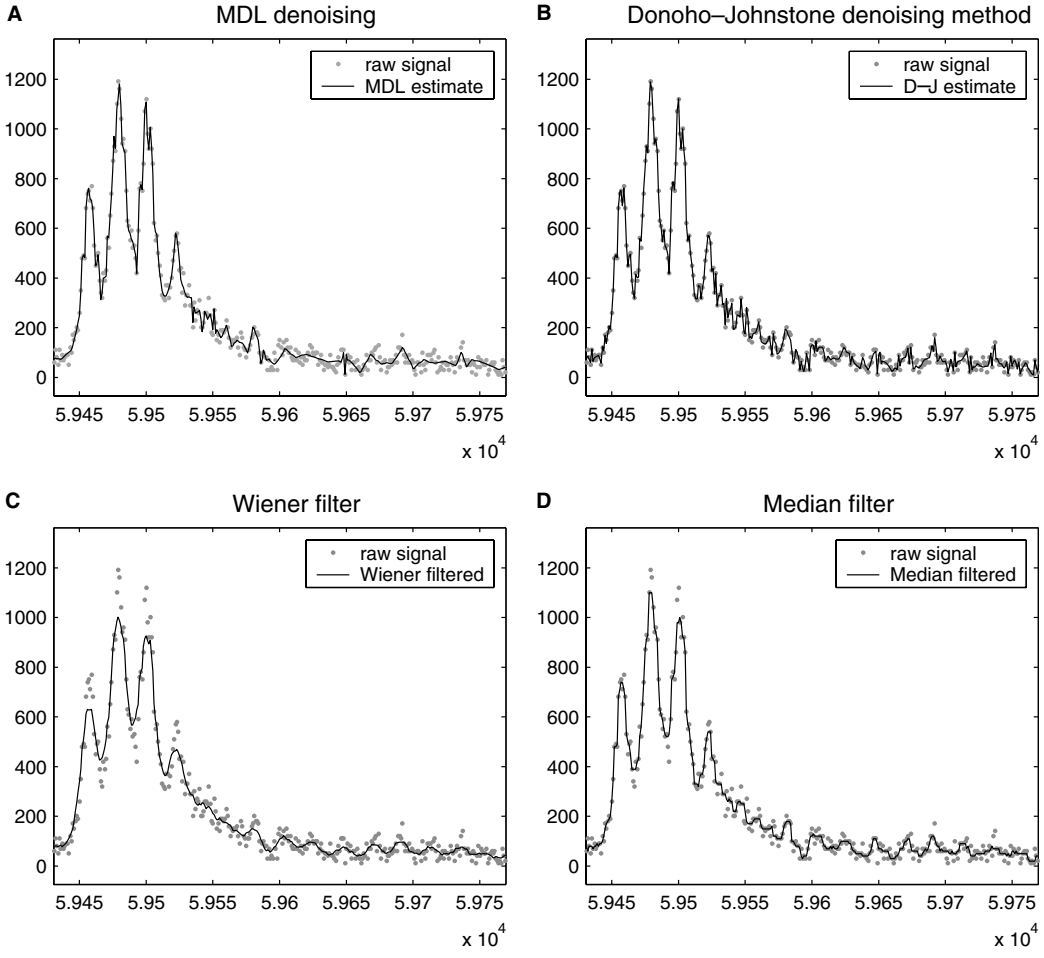

Fig. 1. A short segment of a mass spectrometer signal denoised with: (A) MDL, (B) DJ with MAD noise estimate, (C) Wiener filter, and (D) Median filter. The chosen segment is from a low amplitude part of the signal.

correspond to the largest coefficients in absolute value $c_{(1)}, \ldots, c_{(k)}$ for some k, say $\hat{k}$.

The first term in particular in the minimized expression gives the amount of removed noise as that part in the data that cannot be compressed with the class of Gaussian models considered.

3. Results and discussion

3.1. Results of denoising real signals

Since no 'true' noise-free signal exists in real life data different denoising methods are difficult to compare objectively. Instead, we make the comparison in terms of qualitative characteristics of the signals such as smoothness of the fit and the shape of the peaks. Each of the four methods mentioned was applied to the test signals and the results are shown in the following figures. In each case we have chosen a segment which shows the typical behavior of the characteristics of the signal.

The MS signal contains sparsely located peaks and a noisy baseline. A typical MS signal may contain both high and low intensity peaks; the high amplitude peaks correspond to the most abundant substances while the low amplitude peaks are caused by substances present in small quantities. Successful denoising methods in mass spectroscopy data should be able to distinguish the smallest peaks from noise while conserving relative peak intensities. Fig. 1 shows denoising results on mass spectrometry signals.

In this signal the useful information is in the high fluctuations, which the Wiener filter Fig. 1C and the Median filter Fig. 1D, distort somewhat. The MDL and DJ techniques, Fig. 1A and B, respectively, preserve the shape of the high fluctuations very well while reducing the noise elsewhere. However, the MDL signal estimate is smoother than the DJ denoised signal in the low fluctuations part which over-fits to the noise and does not quite remove it. This is reflected in the estimates of the noise variance, which in the MDL method is the variance of the removed part in the data, but in the DJ method is defined by Eq. (6). The MAD estimate in DJ is $\hat{\sigma}^2_{\mathrm{MAD}} = 39.48$ and the corresponding MDL estimate is $\hat{\sigma}^2_{\mathrm{MDL}} = 453.02$.

The optimal model of the denoised data is defined by the retained coefficients, and hence the fewer coefficients are retained the simpler the resulting model is. The number $1 - C$, which we may call the *compressibility index*, is an approximate measure of the amount of information received, where $C = \hat{k}/n$ is the ratio of the number of retained coefficients to the number of all the coefficients and is an approximate indication of the data compression obtained. After all, since the non-informative noise cannot be compressed the larger the index $1 - C$ is the more of the data have been compressed leaving less as the incompressible noise. The compressibility index is a useful measure in illustrating and comparing how the criteria for selecting the wavelet coefficients behave. In the

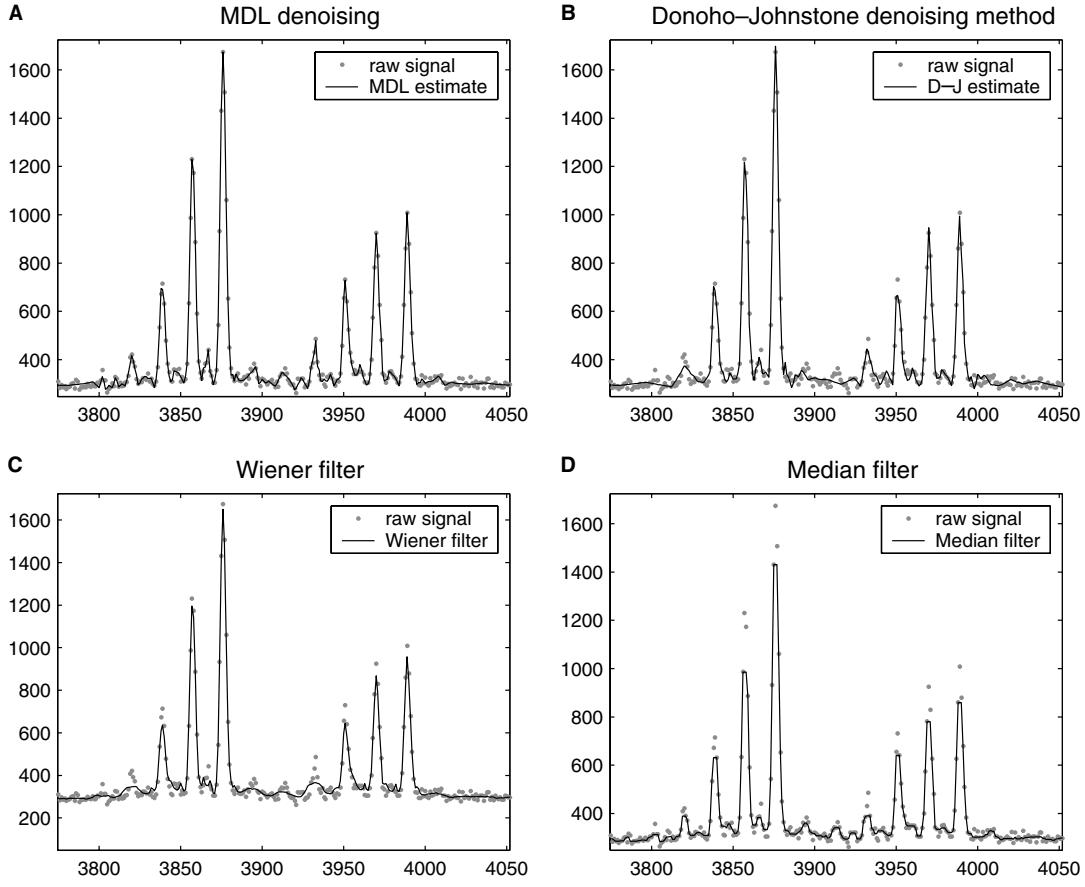

Fig. 2. A segment of a capillary electrophoresis microsatellite genotyping signal denoised with: (A) MDL, (B) DJ, (C) Wiener filter, and (D) Median filter.

mass spectrometer signal the MDL method gives $1 - C = 0.97$ and the DJ method $1 - C = 0.81$.

The CE microsatellite genotyping signals contain few peak clusters and a noisy baseline. A special characteristic is the signal cross-talk phenomenon resulting from overlapping dye spectra. High intensity peaks may leak to other channels and create new peaks or enhance existing peaks. Because denoising has to be done before the spectral separation process, the intensity of the real peaks and bleed through peaks should be preserved. The denoising results for capillary electrophoresis microsatellite genotyping signals are shown in Fig. 2.

Both the Wiener and the Median filter, Fig. 2C and D, respectively, tend to dampen the peaks and lose information, especially in the smaller peaks. The MDL and the DJ methods produce good results, the former retaining slightly better the small peaks. The MAD estimate of the noise variance is $\hat{\sigma}^2_{\mathrm{MAD}} = 367.54$ and the MDL estimate $\hat{\sigma}^2_{\mathrm{MDL}} = 137.66$. The DJ compressibility index is 0.96 and MDL index is 0.93. For this signal the compressibility index is slightly misleading, because we see that DJ, Fig. 2B, cuts the two narrow peaks around index 3950, thereby regarding a part of the useful information as noise. The sequencing signal is similar to the genotyping signal except that there are more peaks and they are located closer to each other. Also, the baseline is more ambiguous due to the number of peaks and the bleed through peaks. Fig. 3 shows the denoising results on capillary electrophoresis sequencing signals.

The Median filter, Fig. 3D, reduces slightly peak heights, but retains the peak shapes well, and over all rivals the MDL technique. The Wiener filter Fig. 3C performs poorly. DJ method Fig. 3B, maybe perhaps ranked as the third best. It reduces again the heights of the smaller peaks, and between indexes 7300 and 7350 fails completely to locate two smaller peaks. The MAD estimate of the variance of the noise is $\hat{\sigma}^2_{\mathrm{MAD}} = 973.98$ and the MDL estimate is $\hat{\sigma}^2_{\mathrm{MDL}} = 1473.58$. The DJ compressibility index is 0.91, and the MDL index is virtually the same 0.90.

In the tested data both of the MDL and the DJ denoising methods produce reasonable results. The performance of the DJ method depends heavily on the quality of the variance estimate. When the noise variance estimate is too small the DJ method retains too many coefficients which amounts to overfitting to the noise as can be seen from Fig. 1B. On the other hand when the variance estimate is too large DJ method tends to suppress high frequency peaks as in Fig. 2B. There are also more subtle differences between the two techniques due to the fact that the threshold resulting from the minimization of the MDL criterion is based on a completely different principle and is a more complex function of the data than just a single estimate of the noise variance. An indication of the difficulty with the prejudiced estimation of the noise variance in the DJ technique is evident in Fig. 3B. In this sequencing signal case the MAD estimate of the noise variance is quite small, but

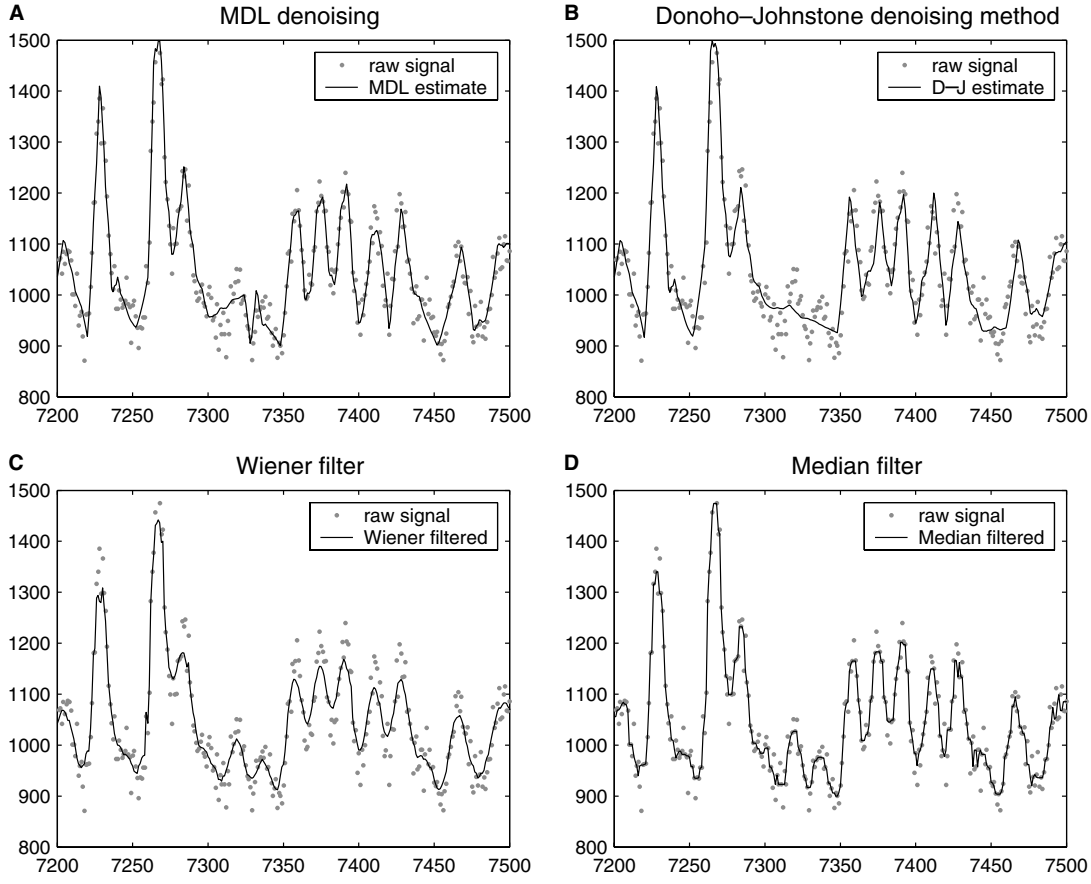

Fig. 3. Capillary electrophoresis sequencing signal. The unprocessed raw signal is shown with gray dots. The denoised signals are shown with black line. Results for: (A) MDL, (B) the DJ method with MAD noise variance estimate, (C) the Wiener filtered signal, and (D) the Median filtered signal.

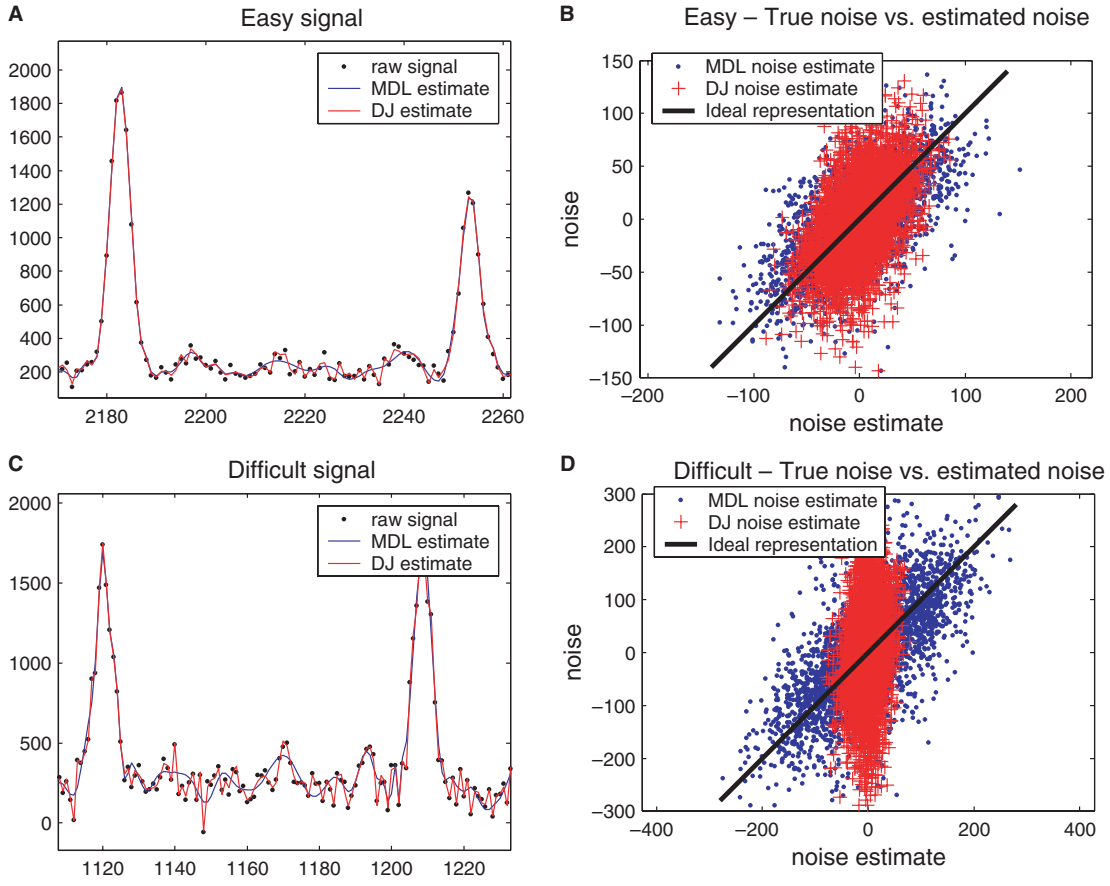

Fig. 4. A synthetic sequencing signal segment. The observed signal values are shown with black dots and the denoised signal estimates are with red (DJ) and blue (MDL) lines. (A) Low noise variance $\sigma_1^2 = 1647.6$ signal. (B) A scatter plot of the added noise values against their estimates $\hat{\varepsilon}_{\mathrm{MDL}}$ and $\hat{\varepsilon}_{\mathrm{MAD}}$. (C) High noise variance $\sigma_2^2 = 9675.1$ signal. (D) A scatter plot of the added noise values against their estimates $\hat{\varepsilon}_{\mathrm{MDL}}$ and $\hat{\varepsilon}_{\mathrm{MAD}}$.

peak cutting takes place anyway. This is due to the variations in peak width and frequencies.

3.2. Results on denoising synthetic signals

Because of the generally superior performance of the MDL and DJ methods in our previous data sets we do not include the Wiener and the Median filters in the further comparison of the techniques when applied to synthetic data.

We generated an artificial sequencing signal and added to it Gaussian noise of mean zero and variance, first $\sigma_1^2 = 1647.6$ and then $\sigma_2^2 = 9675.1$. Fig. 4A shows the observed signal y and the corresponding MDL and DJ estimates $\hat{y}$ of the signal with low noise variance σ_1^2. Both methods recover the large peaks well but only the MDL technique is capable of following the smooth fluctuations. The high variance σ_2^2 signal y and its denoised version $\hat{y}$ are shown in Fig. 4C. Both techniques are still capable to recover the large peaks well. In order to get an idea of the recovery of the small fluctuations we study how well the two techniques recover the added noise component.

With synthetic signals the residual noise $e = y - \hat{y}$ can be compared with the added real noise component. If we plot the values of the added noise against themselves we get a straight line as shown in Fig. 4B and D. Similarly when these values are plotted against the MDL and the DJ residuals we get two clouds of points around the ideal line. We see in Fig. 4B in case of the small variance signal that the MDL estimates of the noise signal follow the line a bit better, especially when the true noise values deviate from the mean zero by 20 more than 50.

Fig. 4D shows the same scatter plot for the high variance signal. We see that the MDL estimates of the noise are much better than the DJ estimates, which fail completely to follow the ideal straight line; they scatter about zero in a random manner in a range from −100 to 100 regardless of the range $(-300, 300)$ of the true noise. The DJ method incorrectly identifies the added noise as part of the informative signal. This can be clearly seen in Fig. 4C.

The MDL variance estimate $\hat{\sigma}^2_{\mathrm{MDL}}$ for the low noise variance signal is $\hat{\sigma}^2_{\mathrm{MDL}} = 2039.9$ and the compressibility index is 0.79. The DJ MAD estimate is $\hat{\sigma}^2_{\mathrm{MAD}} = 253.89$ and the compressibility index is 0.63. For the high noise variance signal the variance estimates are $\hat{\sigma}^2_{\mathrm{MDL}} = 8686.7$ while the MAD estimate in the DJ technique $\hat{\sigma}^2_{\mathrm{MAD}} = 258.64$ and the compressibility indices 0.40 for DJ MAD and 0.83 for the MDL technique. When the noise level is known to be as high as in the latter case the resulting low compressibility index of DJ method may also further indicate choosing a too complex signal model resulting in over-fitting to the noise.

3.3. Conclusions

Especially in high-throughput systems with automated data analysis the robustness of each processing step is essential in achieving reliable results. Our analysis shows that a poorly

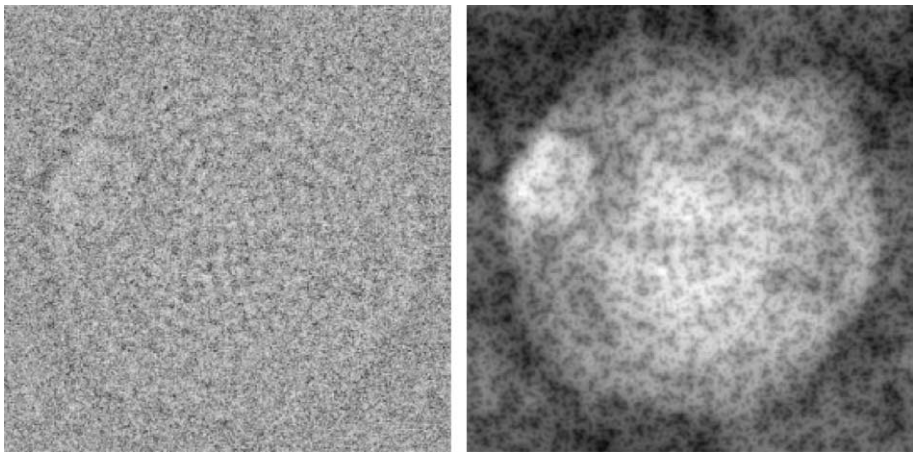

Fig. 5. (Left) Original cryo electron microscopy image of a virus. (Right) The same virus image after MDL denoising.

performing denoising method may alter the signal properties so much that further signal modelling becomes difficult or even impossible.

The denoising process should remove the random noise from the signal without changing the informative signal characteristics. In the signals studied here both the Wiener and the Median filters distort the amplitude of the informative peaks too much with one exception for the Median filter. Although these filters can be fine-tuned to perform properly, the treatment is always case sensitive and time consuming. In some cases they alter the relative intensities of the peaks, too, which are needed in mass spectroscopy signals to obtain the relative compound concentrations correctly.

Both the MDL and the DJ methods work well under these conditions, the former slightly but noticeably better. The former has a fundamental advantage over the latter in that it requires no separate estimate of the noise variance. In making such an estimate it is necessary to form some preselected notion of noise characteristics, which is not only somewhat arbitrary but involves circular reasoning in determining noise in terms of a characteristic of noise. By contrast, in the MDL criterion a reliable estimate of the noise variance is retrieved in addition to the denoised signal.

Although the treatment of the data is of a demonstrative nature, our results suggest that the MDL denoising method performs very well. While extensive further studies are required to validate our conclusions, similar results have been reported in other studies. As an example, a result from an MDL based image denoising method presented in [11] is shown in Fig. 5.

Acknowledgements: We thank Prof. Marc Baumann in University of Helsinki for providing the mass spectrometry data. We also thank M.Sc. Vibhor Kumar in Helsinki University of Technology for providing the cryo-EM images.

References

[1] Barron, A., Rissanen, J. and Yu, B. (1998) The minimum description length principle in coding and modeling. IEEE Transactions on Information Theory 44 (6), 2743–2760.
[2] Efron, B. and Tibshirani, R.J. (1993) An Introduction to the Bootstrap. Number 57 in Monographs on Statistics. Chapman & Hall, London.
[3] Bishop, C.M. (1995) Neural Networks for Pattern Recognition. Oxford University Press, Oxford.
[4] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003) Bayesian Data Analysis. Chapman & Hall/CRC Press, London.
[5] Mallat, S. (1998) A Wavelet Tour of Signal Processing. Academic Press, London.
[6] Daubechies, I. (1992) Ten Lectures on Wavelets, Vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics. Capital City Press.
[7] Hayes, M.H. (1996) Statistical Digital Signal Processing and Modeling. Wiley, New York.
[8] Blackledge, J.M. (1989) Quantitative Coherent Imaging. Academic Press, London.
[9] Jain, A.K. (1989) Fundamentals of Digital Image Processing. Prentice Hall Information and System Sciences Series. Prentice-Hall, Englewood cliffs, NJ.
[10] Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaption by wavelet shrinkage. Biometrika 81 (3), 425–455.
[11] Kumar, V., Heikkonen, J., Engelhardt, P. and Kaski, K. (2004) Robust filtering and particle picking in micrograph images towards 3D reconstruction of purified proteins with cryo-electron microscopy. Journal of Structural Biology 145 (1–2), 41–51.