Available online at www.sciencedirect.com
ScienceDirect
Analytica Chimica Acta 581 (2007) 159–167
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

Sub-optimal wavelet denoising of coaveraged spectra employing statistics from individual scans

Roberto Kawakami Harrop Galvão [1], Heronides Adonias Dantas Filho [2], Marcelo Nascimento Martins [1], Mário César Ugulino Araújo *, Celio Pasquini [2]

Universidade Federal da Paraíba, CCEN, Departamento de Química-Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), Caixa Postal 5093, CEP 58051-970 João Pessoa, PB, Brazil

Received 6 June 2006; received in revised form 25 July 2006; accepted 28 July 2006
Available online 7 August 2006

Abstract

This paper proposes a novel wavelet denoising method, which exploits the statistics of individual scans acquired in the course of a coaveraging process. The proposed method consists of shrinking the wavelet coefficients of the noisy signal by a factor that minimizes the expected square error with respect to the true signal. Since the true signal is not known, a sub-optimal estimate of the shrinking factor is calculated by using the sample statistics of the acquired scans. It is shown that such an estimate can be generated as the limit value of a recursive formulation. In a simulated example, the performance of the proposed method is seen to be equivalent to the best choice between hard and soft thresholding for different signal-to-noise ratios. Such a conclusion is also supported by an experimental investigation involving near-infrared (NIR) scans of a diesel sample. It is worth emphasizing that this experimental example concerns the removal of actual instrumental noise, in contrast to other case studies in the denoising literature, which usually present simulations with artificial noise. The simulated and experimental cases indicate that, in classic denoising based on wavelet coefficient thresholding, choosing between the hard and soft options is not straightforward and may lead to considerably different outcomes. By resorting to the proposed method, the analyst is not required to make such a critical decision in order to achieve appropriate results.

Keywords: Wavelet transform; Denoising; Individual scans; Coaveraging process; Near-infrared spectroscopy; Diesel analysis

1. Introduction

A usual approach to improving the signal-to-noise ratio (SNR) in an instrumental signal consists of carrying out repeated measurements and coaveraging the individual scans [1,2]. However, the number of scans cannot be arbitrarily large because of time constraints, as well as drift in the intensity and wavelength axes. An alternative consists of applying signal processing techniques to further reduce the noise in the coaveraged signal. Such a procedure is termed "denoising" [3,4], "filtering" [5] or "noise suppression" [6,7].

In this context, the discrete wavelet transform (DWT) may be a useful tool [8]. For this purpose, the signal is converted to the wavelet domain, usually by a filter bank algorithm [9], and a transformation is applied to attenuate ("shrink") the wavelet coefficients with low signal-to-noise ratio. A simple procedure consists of maintaining the approximation coefficients (low-frequency components of the signal) of the DWT, while replacing all the detail coefficients (high-frequency components of the signal) with zeros [10]. Such an approach assumes that the most important signal features are concentrated in low frequencies, whereas the high-frequency components are dominated by noise. More elaborate methods usually employ an apodization operation, in which coefficients below a given threshold are replaced with zeros. The remaining coefficients can be kept unaltered ("hard thresholding") or can be decreased by the threshold value ("soft thresholding") [11].

Apodization methods require a judicious choice of a threshold, which may not be trivial if the noise statistics (or at least the noise intensity) is unknown. In a seminal paper, Donoho [11] proposed a criterion known as universal threshold, which has been used in several papers thereafter [4,6,12]. A number

* Corresponding author. Tel.: +55 83 3216 7438; fax: +55 83 3216 7437.
E-mail address: laqa@quimica.ufpb.br (M.C.U. Araújo).
[1] Divisão de Engenharia Eletrônica, Instituto Tecnológico de Aeronáutica, Brazil.
[2] Instituto de Química, Universidade Estadual de Campinas, Brazil.

0003-2670/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2006.07.078

of comparisons between Donoho's wavelet denoising with other filtering techniques have been reported [3,10,13,14].

Barclay et al. [4] compared wavelet denoising with Savitzky-Golay (SG) smoothing and discrete Fourier transform (DFT) filtering and concluded that the wavelet-based techniques provide better results. Similar findings were also presented in [3,14].

In this work, an alternative wavelet denoising method for signals obtained by a coaveraging procedure is proposed. Denoising is carried out by attenuating ("shrinking") the wavelet coefficients in order to minimize the expected square error between the reconstructed and the "true" signals. Since the true signal is not known in practice, a sub-optimal estimate of the shrinking factor is calculated by using sample statistics of the acquired scans. A mathematical development is presented to show that such an estimate can be generated as the limit value of a recursive formulation.

The proposed strategy is compared with hard and soft thresholding in a simulated example. Moreover, an experimental study is carried out by using near-infrared (NIR) spectral scans of a diesel sample. In this case, rather than using simulated noise, the investigation is concerned with actual instrumental noise.

2. Background and theory

2.1. Notation

A sequence of elements indexed by λ is denoted by $\{x(\lambda)\}$ or, when there is no need to specify the indexing element, simply by x. A single element of the sequence is indicated as $x(\lambda)$. Operator E denotes the expected value of a random variable.

2.2. Discrete wavelet transform

As described in the wavelet literature [8,9], the DWT of a signal $\{x(\lambda)\}$, $\lambda = 1, 2, \ldots, J$, can be calculated in a fast manner by using a bank of digital filters of the form depicted in Fig. 1. The basic structure of this filter bank consists of a pair of low-pass (H) and high-pass (G) filters followed by a downsampling operation, which consists of discarding every other point of the filter outputs. Filters H and G usually have finite length (that is, finite impulse response) and can be chosen such that the overall transformation is orthogonal, thus preserving the information content of the input signal. The first filtering/downsampling operation results in two sequences $\{c^1(k_1)\}$ (low-pass) and $\{d^1(k_1)\}$ (high-pass), with $k_1 = 1, \ldots, J/2$. Sequences c^1 and d^1 are termed first-level approximation and detail coefficients, respectively. Each approximation and detail coefficient is associated to a region of the original sequence x. Therefore, the position information is partially retained by DWT.

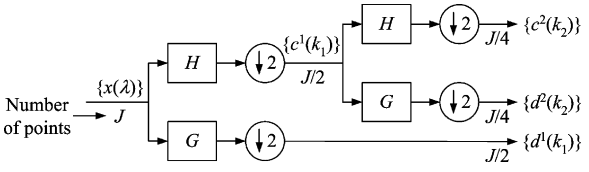

Fig. 1. Wavelet filter bank with two decomposition levels. H and G represent low-pass and high-pass filters, respectively, whereas $\downarrow 2$ denotes the downsampling operation.

Sequence $\{c^1(k_1)\}$ can be further decomposed to generate sequences $\{c^2(k_2)\}$ (second-level approximation coefficients) and $\{d^2(k_2)\}$ (second-level detail coefficients), with $k_2 = 1, \ldots, J/4$. If this procedure is repeated up to the Nth decomposition level, the final result is a sequence $\{t(k)\}$, $k = 1, \ldots, J$, formed by concatenating the level-N approximation coefficients (c^N), and all detail coefficients ($d^N, \ldots, d^1$). In this case, the approximation coefficients c^N comprise the low-frequency components of x, the detail coefficients d^1 are associated to the high-frequency components, and the remaining detail coefficients account for intermediate frequencies. As can be seen, DWT performs a frequency decomposition of the input signal, which is similar to the application of a DFT. However, in contrast to DFT, the position information is partially retained by DWT, as mentioned above. Therefore, localized features in the original domain (such as sharp peaks), which are spread over the entire frequency domain in DFT, may be concentrated in a small number of wavelet coefficients. Such an attribute is of value in denoising applications, because it allows signal features to be more easily distinguished from background noise.

Choosing the best number of decomposition levels in DWT is still an open problem. As reported in [12], it is usual practice to employ the maximum possible number of decomposition levels.

2.3. Hard and soft thresholding

In the hard thresholding technique, the wavelet coefficients $\{t(k)\}$ of the noisy signal are subjected to the following transformation:

$$t_{\mathrm{f}}^{\mathrm{hard}}(k) = \begin{cases} 0, & \text{if } |t(k)| \le h \\ t(k), & \text{if } |t(k)| > h \end{cases} \quad (1)$$

for $k = 1, \ldots, J$, where h is a threshold value and subscript f stands for "filtered". The transformed coefficients $\{t_{\mathrm{f}}^{\mathrm{hard}}(k)\}$ are then used to reconstruct the signal by inverse DWT.

Soft thresholding is more "aggressive" in that all coefficients are modified, as shown in Eq. (2). In this case, the coefficients that are not replaced with zeros are shrunk towards zero by the threshold value h.

$$t_{\mathrm{f}}^{\mathrm{soft}}(k) = \begin{cases} 0, & \text{if } |t(k)| \le h \\ \mathrm{sign}[t(k)][|t(k)| - h], & \text{if } |t(k)| > h \end{cases} \quad (2)$$

The main difficulty associated to these methods consists of determining an appropriate threshold value. If h is too small, the denoising will not be effective, but if h is too large, important features of the signal may be distorted ("over-smoothing"). The so-called universal threshold, which has been extensively used in the literature [10,12,15], is defined by

$$h = \sigma\sqrt{2\ \ln(J)} \quad (3)$$

where σ is the standard deviation of the noise, which is considered to be white, and J is the length of the signal under consideration. If σ is unknown, an estimate can be obtained

by [6,10,11,15]

$$\hat{\sigma}^{\text{median}} = \frac{\text{median}(|d^1|)}{0.6745} \tag{4}$$

Such an approximation assumes that the highest frequency components, which are contained in the first-level detail coefficients d^1, are mostly associated to noise.

It is worth noting that most implementations of hard and soft thresholding leave the approximation coefficients unaltered because such coefficients usually correspond to relevant features of the signal [10]. This formulation will be adopted throughout the present work.

2.4. Proposed method

In the proposed method, each individual scan $\{x_I(\lambda)\}$ is assumed to be the sum of the deterministic term ("true signal") and a zero-mean, uncorrelated noise term (subscript I stands for individual). If an orthogonal DWT is employed, the noise in the wavelet coefficients is also zero-mean and uncorrelated [16]. Therefore, the kth wavelet coefficient can be written as

$$t_I(k) = \mu(k) + \eta_I(k) \tag{5}$$

where $\mu(k)$ corresponds to the true signal and $\eta_I(k)$ is a noise term such that

$$E\{\eta_I(k)\} = 0 \tag{6}$$

$$E\{\eta_I(k_1)\eta_I(k_2)\} = \begin{cases} \sigma_I^2(k), & k_1 = k_2 = k \\ 0, & k_1 \neq k_2 \end{cases} \tag{7}$$

Consequently, the kth wavelet coefficient for the average of M scans is given by

$$t(k) = \mu(k) + \eta(k) \tag{8}$$

with

$$E\{\eta(k)\} = 0 \tag{9}$$

$$E\{\eta(k_1)\eta(k_2)\} = \begin{cases} \sigma^2(k), & k_1 = k_2 = k \\ 0, & k_1 \neq k_2 \end{cases} \tag{10}$$

where $\sigma^2(k) = \sigma_I^2(k)/M$ [1,16], that is, the noise variance is inversely proportional to the number of coaveraged scans.

In the proposed strategy, each wavelet coefficient $t(k)$ is modified by a transformation of the form

$$t_f(k) = f(k)t(k) \tag{11}$$

and the inverse DWT is then applied to the resulting sequence $\{t_f(k)\}$ in order to obtain the denoised signal. The sequence of weights $\{f(k)\}$ is calculated to minimize the cost function C defined as

$$C = \sum_{k=1}^{J} E\{[t_f(k) - \mu(k)]^2\} \tag{12}$$

which is the expected sum of square errors between the modified coefficients and their ideal ("true") values. By inserting Eqs. (8) and (11) in Eq. (12), it follows that

$$\begin{aligned} C &= \sum_{k=1}^{J} E\{[f(k)\mu(k) + f(k)\eta(k) - \mu(k)]^2\} \\ &= \sum_{k=1}^{J} E\{[\mu(k)[f(k) - 1] + f(k)\eta(k)]^2\} \\ &= \sum_{k=1}^{J} E\{\mu^2(k)[f(k) - 1]^2 + 2f(k)\eta(k)\mu(k)[f(k) - 1] \\ &\quad + f^2(k)\eta^2(k)\} \\ &= \sum_{k=1}^{J} \mu^2(k)[f(k) - 1]^2 + 2f(k)E\{\eta(k)\}\mu(k)[f(k) - 1] \\ &\quad + f^2(k)E\{\eta^2(k)\} = \sum_{k=1}^{J} \mu^2(k)[f(k) - 1]^2 + f^2(k)\sigma^2(k) \end{aligned} \tag{13}$$

where the last equality follows from Eqs. (9) and (10). The derivative of C with respect to $f(k)$ is given by

$$\frac{\partial C}{\partial f(k)} = 2\mu^2(k)[f(k) - 1] + 2f(k)\sigma^2(k) \tag{14}$$

The minimum of the cost is achieved by imposing $\partial C/\partial f(k) = 0$ for each k, which leads to

$$\begin{aligned} &\mu^2(k)[f^*(k) - 1] + f^*(k)\sigma^2(k) = 0 \\ &\therefore f^*(k)[\mu^2(k) + \sigma^2(k)] = \mu^2(k) \\ &\therefore f^*(k) = \frac{\mu^2(k)}{\mu^2(k) + \sigma^2(k)} \end{aligned} \tag{15}$$

where symbol * is used to denote the optimal value. Moreover, it follows from Eq. (14) that the second derivatives of the cost are given by

$$\frac{\partial^2 C}{\partial f(k_1)\partial f(k_2)} = \begin{cases} 2\mu^2(k) + 2\sigma^2(k), & k_1 = k_2 = k \\ 0, & k_1 \neq k_2 \end{cases} \tag{16}$$

which shows that the Hessian matrix is diagonal with positive diagonal elements. Therefore, it can be concluded that the Hessian matrix is positive-definite, which means that the sequence of weights $\{f^*(k)\}$ does correspond to a minimum of the cost (rather than a maximum or an inflection point).

Eq. (15) shows that smaller weights are assigned to coefficients in which the signal-to-noise ratio is smaller. In fact, the expression for $f^*(k)$ can be re-written as

$$f^*(k) = \left(1 + \frac{\sigma^2(k)}{\mu^2(k)}\right)^{-1} = (1 + \text{SNR}(k)^{-1})^{-1} \tag{17}$$

where SNR(k) is the signal-to-noise ratio (relation between the signal power and the noise power) in the kth coefficient. Therefore, coefficients with a large SNR remain approximately the same ($f^*(k) \approx 1$), whereas coefficients with a small SNR are shrunk towards zero ($f^*(k) \approx 0$).

Since $\mu(k)$ and $\sigma(k)$ are unknown, appropriate estimates must be employed in their place. For this purpose, the sample mean $m(k)$ calculated over M scans can be used as an estimate for $\mu(k)$. If the noise is homoscedastic, an estimate for $\sigma(k)$ is given by $\hat{\sigma}^{\text{median}}$, as defined in Eq. (4). Therefore, Eq. (15) becomes

$$\hat{f}(k) = \frac{m^2(k)}{m^2(k) + (\hat{\sigma}^{\text{median}})^2} \tag{18}$$

with

$$m(k) = \frac{1}{M}\sum_{i=1}^{M} t^i(k) \tag{19}$$

where $t^i(k)$ denotes the kth wavelet coefficient of the ith scan that was actually acquired. The hat symbol ($\wedge$) is used in Eq. (18) to indicate that $\hat{f}(k)$ is sub-optimal, since estimates for $\mu(k)$ and $\sigma(k)$ are used instead of their true values. By using $\hat{f}(k)$ instead of $f(k)$ in Eq. (11), a sub-optimal expression for the denoised coefficient $\hat{t}_{\text{f}}(k)$ can be written as

$$\hat{t}_{\text{f}}(k) = \frac{m^2(k)}{m^2(k) + (\hat{\sigma}^{\text{median}})^2} t(k) \tag{20}$$

The calculation expressed in Eq. (20) can be re-iterated by employing $\hat{t}_{\text{f}}(k)$, instead of $m(k)$, as a better estimate of $\mu(k)$. By doing so, one arrives at the following recursion:

$$\hat{t}_{\text{f}}^{(n)}(k) = \frac{[\hat{t}_{\text{f}}^{(n-1)}(k)]^2}{[\hat{t}_{\text{f}}^{(n-1)}(k)]^2 + (\hat{\sigma}^{\text{median}})^2} t(k) \tag{21}$$

where $\hat{t}_{\text{f}}^{(n)}(k)$ is the result of the nth iteration, with $\hat{t}_{\text{f}}^{(0)}(k) = m(k)$.

It is worth noting that $t(k)$ in Eq. (21) is a random variable, as defined in Eq. (8). However, in order to obtain $\hat{t}_{\text{f}}^{(n)}(k)$ from Eq. (21), $t(k)$ must be replaced by the value actually obtained from measured data. Such a value corresponds to the sample mean $m(k)$ defined in Eq. (19). Therefore, Eq. (21) becomes

$$\hat{t}_{\text{f}}^{(n)}(k) = \frac{[\hat{t}_{\text{f}}^{(n-1)}(k)]^2}{[\hat{t}_{\text{f}}^{(n-1)}(k)]^2 + (\hat{\sigma}^{\text{median}})^2} m(k) \tag{22}$$

The limit value $\hat{t}_{\text{f}}^{\text{lim}}(k)$ for the recursion expressed in Eq. (22) can be determined by imposing the steady-state condition $\hat{t}_{\text{f}}^{(n)}(k) = \hat{t}_{\text{f}}^{(n-1)}(k) = \hat{t}_{\text{f}}^{\text{lim}}(k)$. As a result, the following equation arises:

$$[\hat{t}_{\text{f}}^{\text{lim}}(k)]^3 - [\hat{t}_{\text{f}}^{\text{lim}}(k)]^2 m(k) + \hat{t}_{\text{f}}^{\text{lim}}(k)(\hat{\sigma}^{\text{median}})^2 = 0 \tag{23}$$

which has three possible real-valued solutions:

$$\hat{t}_{\text{f}}^{\text{lim}}(k) = \left\{ 0,\ \frac{m(k) - \sqrt{m^2(k) - 4(\hat{\sigma}^{\text{median}})^2}}{2},\ \frac{m(k) + \sqrt{m^2(k) - 4(\hat{\sigma}^{\text{median}})^2}}{2} \right\} \tag{24}$$

provided that $|m(k)| \geq 2\hat{\sigma}^{\text{median}}$. However, by using the initial condition $\hat{t}_{\text{f}}^{(0)}(k) = m(k)$, the recursion in Eq. (22) actually converges to the third solution if $m(k) > 0$ and to the second solution if $m(k) < 0$, that is

$$\hat{t}_{\text{f}}^{\text{lim}}(k) = \begin{cases} \dfrac{m(k) + \sqrt{m^2(k) - 4(\hat{\sigma}^{\text{median}})^2}}{2}, & m(k) > 0 \\ \dfrac{m(k) - \sqrt{m^2(k) - 4(\hat{\sigma}^{\text{median}})^2}}{2}, & m(k) < 0 \end{cases} \tag{25}$$

as shown in Fig. 2a and b for $m(k) > 0$. A similar graphic analysis can also be accomplished if $m(k) < 0$.

If $|m(k)| < 2\hat{\sigma}^{\text{median}}$, the only real-valued solution for Eq. (23) is $\hat{t}_{\text{f}}^{\text{lim}}(k) = 0$. Such a conclusion is illustrated in Fig. 2c.

Finally, the proposed denoising rule can be summarized as follows:

$$\hat{t}_{\text{f}}^{\text{lim}}(k) = \begin{cases} \dfrac{m(k) + \text{sign}[m(k)]\sqrt{m^2(k) - 4(\hat{\sigma}^{\text{median}})^2}}{2}, & |m(k)| \geq 2\hat{\sigma}^{\text{median}} \\ 0, & |m(k)| < 2\hat{\sigma}^{\text{median}} \end{cases} \tag{26}$$

where sign$[m(k)]$ equals +1 if $m(k) > 0$ and −1 otherwise.

It is worth noting that using a single estimate $\hat{\sigma}^{\text{median}}$ for the standard deviation of the noise in all wavelet coefficients may not be appropriate if the noise is heteroscedastic. In this case, $\sigma(k)$ could be estimated as

$$\hat{\sigma}^{\text{sample}}(k) = \sqrt{\frac{1}{M(M-1)}\sum_{i=1}^{M}[t^i(k) - m(k)]^2} \tag{27}$$

for $k = 1, \ldots, J$. In this equation, the additional factor M is included in the denominator to account for the coaveraging effect, as discussed at the beginning of this section. In order to use this estimate, suffice it to replace $\hat{\sigma}^{\text{median}}$ with $\hat{\sigma}^{\text{sample}}(k)$ in Eq. (26).

3. Experimental

3.1. Simulated example

In the simulated example, a synthetic signal $\{x(\lambda)\}$, $\lambda = 1, 2, \ldots, 1024$, was generated as the superposition of six Gaussian functions with means equally distributed in the interval (1–1024). The peak heights and standard deviations of the Gaussians were randomly generated in the range (2–4) and (0–30), respectively. The resulting signal is presented in Fig. 3a. Each simulated scan was generated by adding a zero-mean, white Gaussian noise to the true signal. The standard deviation of the noise was set to 0.1. A noisy scan simulated in this manner can be seen in Fig. 3b.

A Symlet 8 wavelet filter bank was employed in the denoising procedure. In this case, filters H (low-pass) and G (high-pass) have 16 weights each. The filtering operations were accomplished by using circular convolution. The number of decomposition levels was set to six, which is the maximum possible value given the length of the input signal. In fact, the number of approximation coefficients at level 6 is 1024/(2^6) = 16, which is equal to the length of the H and G filters. Therefore, if an

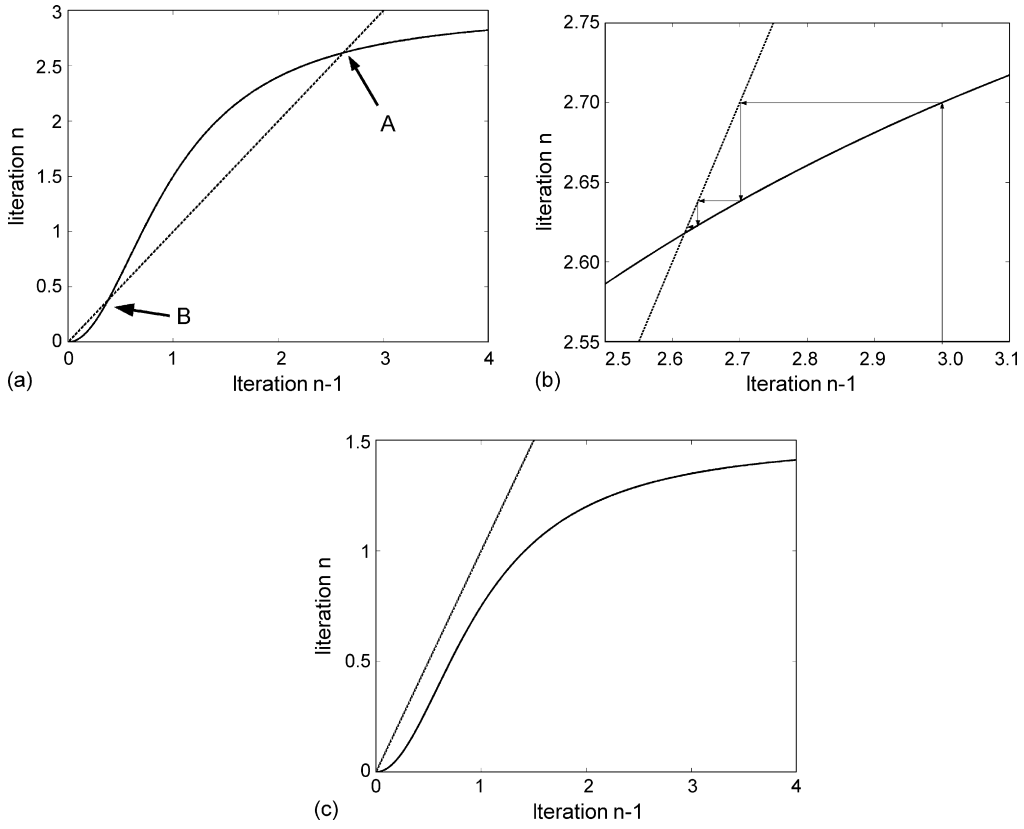

Fig. 2. (a) Graphical representation of the recursion expressed in Eq. (22) for $m(k) = 3$ and $\hat{\sigma}^{\text{median}} = 1$. The values of $\hat{t}_{\text{f}}^{(n-1)}(k)$ and $\hat{t}_{\text{f}}^{(n)}(k)$ are associated to the horizontal and vertical axes, respectively. The dashed straight line is the bisectrix of the first quadrant ($\hat{t}_{\text{f}}^{(n-1)}(k) = \hat{t}_{\text{f}}^{(n)}(k)$). The intersection points A and B correspond to the third and second solutions in Eq. (24), respectively. (b) Enlargement of the graph around point A. The arrows indicate the sequence of recursive calculations that start from $\hat{t}_{\text{f}}^{(0)}(k) = m(k)$ and converge to point A. (c) Graphical representation for $m(k) = 1.5$ and $\hat{\sigma}^{\text{median}} = 1$.

additional decomposition was carried out, the resulting approximation coefficients would no longer retain position information, which is a key advantage of the DWT.

In addition to the results obtained with the simulation conditions described above, an investigation of the signal-to-noise ratio on the denoising outcome was also carried out. For this purpose, three different values for the standard deviation of the noise were employed (0.01, 0.1 and 1.0).

The denoising results were evaluated in terms of the root-mean-square error (RMSE) between the reconstructed (x_{f}) and the true (x) signals, which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{J}\sum_{\lambda=1}^{J}[x_{\text{f}}(\lambda) - x(\lambda)]^2} \tag{28}$$

where J is the signal length.

3.2. Experimental case study

Spectra of a single diesel sample were acquired by using a Bomem FT-NIR spectrophotometer with a 2 cm^{-1} resolution in the 850–1300 nm range at room temperature (24 °C) and relative air humidity of 48%. The resulting spectra had 2560 points. With these settings, the measurements have excellent wavelength repeatability, which ensures that the individual scans are properly aligned. Initially, 50 individual scans were acquired (a single scan is depicted in Fig. 4a). Thereafter, a coaveraged spectrum (used as reference for RMSE calculations) was acquired for the same sample by using 100 scans (Fig. 4b).

Denoising was initially carried out by using a Symlet 8 wavelet filter bank with seven decomposition levels. Moreover, in order to assess the sensitivity of the denoising methods with respect to the choice of the wavelet filters, an investigation was also conducted with 21 wavelets from the Daubechies (db), Symlet (sym) and Coiflet (coif) families, as in previous studies [10,12].

All calculations in the simulated example and experimental case study were accomplished by using the Matlab 6.5 software and its Wavelet Toolbox.

4. Results and discussion

4.1. Simulated example

Fig. 5 presents the denoising results for the simulated data in terms of RMSE values as a function of the number of coaveraged scans. In this case, the standard deviation of the noise in each individual scan was set to 0.1. As can be seen, all denoising methods provided a reduction in RMSE with respect to the coaveraged signal ("no denoising"). The performance of the proposed method was similar to hard thresholding and considerably better than soft thresholding.

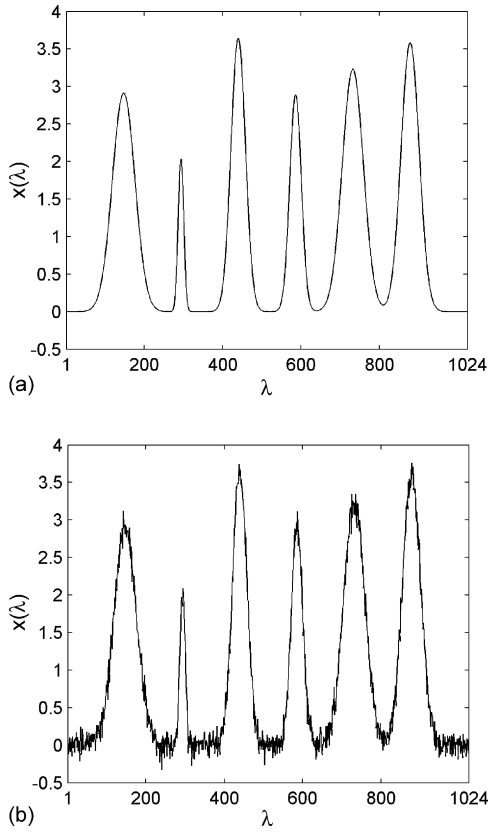

Fig. 3. (a) True signal used in the simulated example. (b) Single scan corrupted by noise with standard deviation of 0.1.

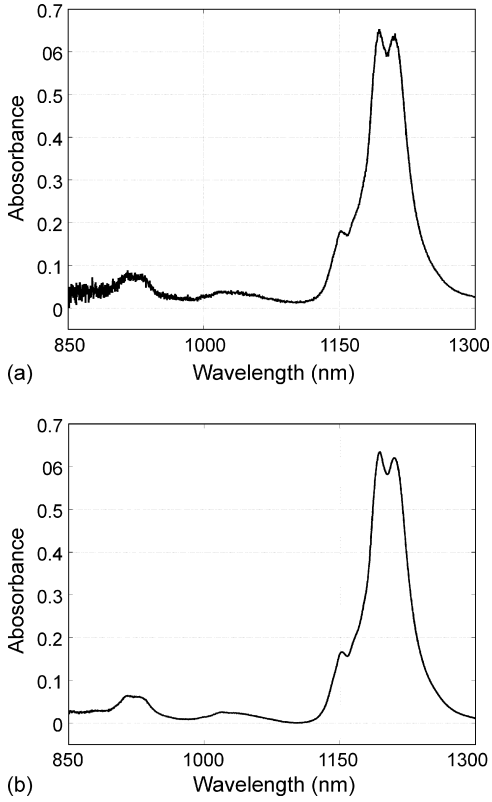

Fig. 4. (a) Individual NIR scan and (b) coaveraged spectrum used as reference for RMSE calculations (100 scans).

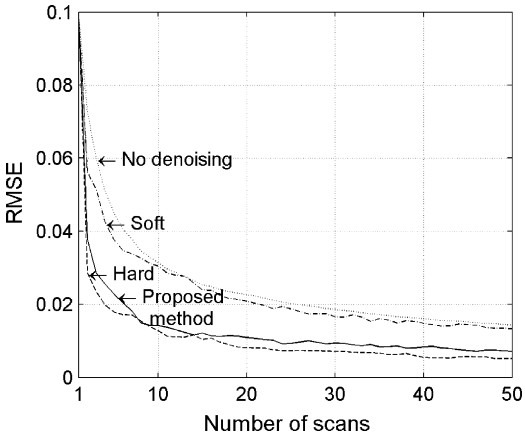

Fig. 5. RMSE with respect to the true signal as a function of the number of coaveraged scans in the simulated example. The standard deviation of the noise in each scan was set to 0.1. A Symlet 8 filter bank with six decomposition levels was employed.

The denoising outcome for eight scans can be visualized in Fig. 6a, in which the coaveraged signal ("no denoising") is compared with the reconstructed signals. The differences between the proposed method and hard thresholding appear mainly in regions of small signal intensity, where the former left some residual noise and the latter introduced oscillatory artifacts. Soft thresholding, on the other hand, caused distortions in the signal,

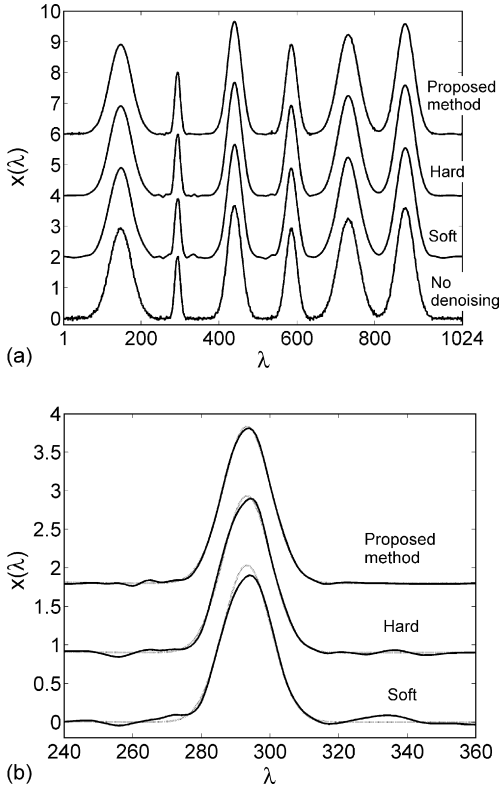

Fig. 6. (a) Comparison between the coaveraged signal resulting from eight scans ("no denoising") and the denoised signals. (b) Comparison between the denoised signals (solid lines) and the true one (dotted lines) in the range λ = 240–360. Different vertical offsets were added to the signals in (a and b) for better visualization.

Table 1
RMSE values (×1000) for eight scans and different noise levels (σ is the standard deviation of the noise in an individual scan)

Denoising method	σ		
	0.01	0.1	1.0
No denoising	3.5	34.6	346
Soft thresholding	4.0	32.8	219
Hard thresholding	1.7	15.5	124
Proposed method	1.6	15.0	142

especially around the peaks, which were over-smoothed. Such an effect is apparent in Fig. 6b, in which the denoised signals are compared with the true one. In this particular example, it can be argued that denoising by soft thresholding was too "agressive".

The effect of noise intensity on the performance of the denoising methods is presented in Table 1. As can be seen, the RMSE has a general tendency to increase with the noise level, as expected. The proposed method is seen to be slightly better than hard thresholding for small noise levels, whereas the situation is inverted if there is too much noise. In all cases, soft thresholding provided the worst results (for $\sigma = 0.01$, it even increased the RMSE value when compared to the coaveraged signal without denoising).

This example shows that choosing between hard and soft thresholding may be critical in order to obtain good denoising. The proposed method has the advantage of not requiring such a decision from the analyst.

4.2. Experimental case study

An inspection of the individual scan in Fig. 4a indicates that noise intensity is not uniform along the entire spectral range. In fact, noise is stronger in shorter wavelengths because of limitations of the NIR detector. Therefore, the proposed method was implemented by using $\hat{\sigma}^{\text{sample}}$ instead of $\hat{\sigma}^{\text{median}}$ in Eq. (26).

According to the results in Fig. 7, soft thresholding was now superior to hard thresholding, in contrast with the findings of the simulated example. This difference between simulated and experimental results points once more to the difficulty of deciding whether to use hard or soft thresholding. On the other hand, the proposed method is again comparable to the best result between both thresholding methods.

In order to compare the denoising results in visual terms, a criterion was devised to choose an appropriate number of scans prior to the application of the denoising procedures. For this purpose, let $x_{a,n}$ denote the signal resulting from coaveraging n scans. As n is increased, the coaveraged signal changes as a result of the reduction in noise level. When such changes become insignificant, the coaveraging process could be stopped to save time. In order to quantify the significance of such changes, let $\Delta(M, N)$ be defined as

$$\Delta(M, N) = \sum_{\lambda=1}^{J} [x_{a,M}(\lambda) - x_{a,2}(\lambda)]^2 \times \left(\sum_{\lambda=1}^{J} [x_{a,N}(\lambda) - x_{a,2}(\lambda)]^2 \right)^{-1} \quad (29)$$

where $J = 2560$ is the number of points in each scan. Ratio $\Delta(M, N)$ compares the results of coaveraging M scans and N scans ($x_{a,M}$ and $x_{a,N}$, respectively) by using the initial average of two

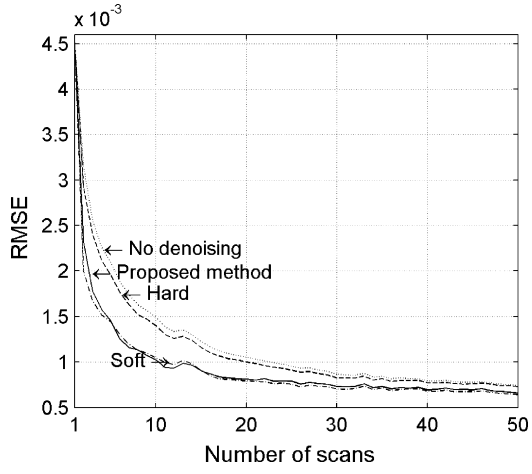

Fig. 7. RMSE with respect to the reference NIR spectrum as a function of the number of coaveraged scans. A Symlet 8 filter bank with seven decomposition levels was employed.

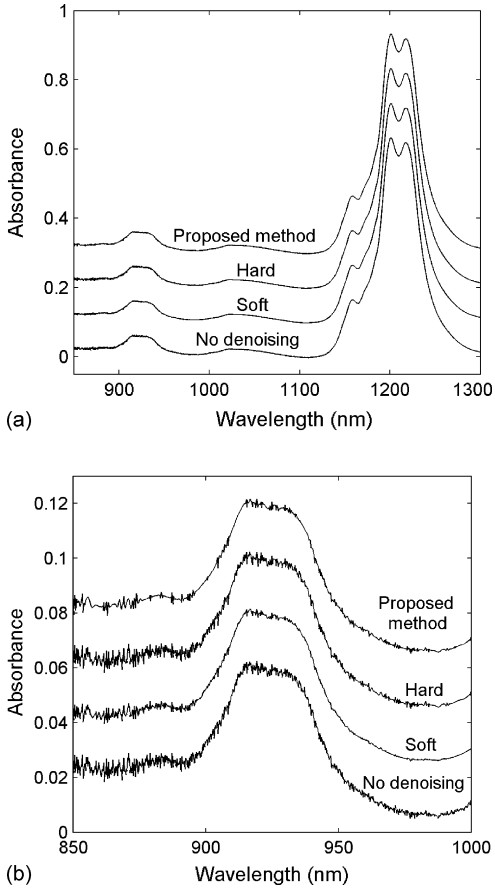

Fig. 8. (a) Comparison of the denoised NIR spectra obtained after 16 scans. (b) Enlargement of the denoised spectra in the range 850–1000 nm. Different vertical offsets were added to the spectra in (a and b) for better visualization.

Table 2
RMSE values (×1000) for 16 scans and different wavelet filters

Wavelet filters	Denoising method		
	Soft thresholding	Hard thresholding	Proposed method
db2	0.90	1.11	0.89
db3	0.88	1.10	0.89
db4	0.86	1.10	0.88
db5	0.85	1.10	0.86
db6	0.86	1.10	0.86
db7	0.85	1.10	0.87
db8	0.85	1.12	0.86
db9	0.85	1.11	0.86
db10	0.86	1.11	0.85
sym4	0.86	1.11	0.87
sym5	0.87	1.11	0.86
sym6	0.86	1.11	0.88
sym7	0.86	1.11	0.89
sym8	0.85	1.11	0.87
sym9	0.85	1.11	0.82
sym10	0.84	1.10	0.83
coif1	0.89	1.10	0.87
coif2	0.86	1.11	0.88
coif3	0.85	1.10	0.87
coif4	0.84	1.10	0.86
coif5	0.84	1.10	0.82

scans ($x_{a,2}$) as a basis for comparison. If $N > M$ and $\Delta(M, N)$ is not significantly larger than one, the additional time required to acquire N instead of M scans is not justified. In the present case, by successively doubling the number of scans and then applying Eq. (29), it follows that $\Delta(4, 8) = 1.38$, $\Delta(8, 16) = 1.22$, and $\Delta(16, 32) = 1.06$. By using an F-test with significance $\alpha = 0.05$, it can be concluded that $\alpha(16, 32)$ is not significantly larger than one. Therefore, 16 scans were adopted in the discussion below.

Fig. 8a presents the coaveraged spectrum before denoising, as well as the spectra reconstructed by the three methods under comparison. The main differences can be seen in the short-wavelength region, where the signal-to-noise ratio is relatively poor. An enlargement of the spectra in the interval 850–1000 nm is presented in Fig. 8b. As can be seen, the proposed method provided better noise suppression in the range 850–900 nm, whereas soft thresholding was superior in the range 950–1000 nm.

Table 2 presents RMSE values resulting from the use of 16 coaveraged scans and different wavelet filters. In this table, the notations db$N$, sym$N$, and coif$N$ refer to filters of length $2N$, $2N$, and $6N$, respectively. According to an F-test with $\alpha = 0.05$, the proposed method always outperformed hard thresholding. In addition, it was equivalent to soft thresholding in all cases with the exception of sym7 and sym9. For sym9 the proposed method is significantly superior to soft thresholding, whereas the opposite situation is observed for sym7.

5. Conclusions

The wavelet denoising method proposed in this paper was shown to be a viable alternative to the standard techniques of hard and soft thresholding. In the simulated example, the proposed method was approximately equivalent to hard thresholding, by taking into account RMSE performance for different SNR scenarios. On the other hand, soft thresholding caused signal distortion by over-smoothing of peaks. This example showed that choosing between hard and soft thresholding may be critical in order to obtain a good denoising result. By resorting to the proposed method, the analyst is not required to make this decision.

In the second example, which employed real experimental data, soft thresholding was superior to hard thresholding, in contrast with the findings of the simulated example. Once more, the proposed method was as good as the best choice between both thresholding methods. Such a conclusion was supported by a comparison involving 21 different wavelet filters. It should be emphasized that this experimental example concerned the removal of actual instrumental noise, in contrast to most case studies in the denoising literature, which usually involve simulations with artificial noise.

It is worth noting that, despite the mathematical details underlying the proposed method, the final formulation consists of a simple equation (Eq. (26)), which can be implemented in a straightforward manner.

An interesting possibility for further research would be the application of the proposed method together with the stationary wavelet transform [17–20], originally called "á trous" algorithm [9,21], which makes the wavelet decomposition space-invariant. The wavelet packet transform [22–24], which offers more flexibility in the signal decomposition process, could also be employed.

In addition, future works may extend the scope of the investigations presented in this paper by considering case studies involving other instrumental techniques. The proposed method should be particularly useful in nuclear resonance or far-infrared spectroscopy, where a considerable number of scans may be required to achieve an adequate signal-to-noise ratio. In such cases, the use of denoising techniques would be of value to reduce the acquisition time needed to attain a given SNR or, conversely, to increase SNR for a given number of scans.

Acknowledgments

This work was supported by FAPESP (doctorate scholarship 04/06807-4), CNPq (Universal 475204/2004-2, PRONEX 015/98, scholarships and research fellowships) and CAPES (PROCAD 0081/05-1).

References

[1] D.A. Skoog, D.M. West, F.J. Holler, Fundamentals of Analytical Chemistry, Saunders College Publishing, Fort Worth, 1992.
[2] R.K.H. Galvão, S. Hadjiloucas, J.W. Bowen, Opt. Lett. 27 (2002) 643.
[3] B.K. Alsberg, A.M. Woodward, M.K. Winson, J. Rowland, D.B. Kell, Analyst 122 (1997) 645.
[4] V.J. Barclay, R.F. Bonner, I.P. Hamilton, Anal. Chem. 69 (1997) 78.
[5] A.V. Oppenheim, R.W. Schafer, Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, 1989.
[6] B. Walczak, D.L. Massart, Chemom. Intell. Lab. Syst. 36 (1997) 81.
[7] N. Saito, Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length crite-

rion, in: E. Foufoula-Georgiou, P. Kumar (Eds.), Wavelets in Geophysics, Academic Press, New York, 1994.
[8] B. Walczak, Wavelets in Chemistry, Elsevier Science, New York, 2000.
[9] M. Vetterli, J. Kovacevic, Wavelets and Subband Coding, Prentice-Hall, New Jersey, 1995.
[10] C.S. Cai, P.D. Harrington, J. Chem. Inf. Comput. Sci. 38 (1998) 1161.
[11] D.L. Donoho, IEEE Trans. Inf. Theory 41 (1995) 613.
[12] L. Pasti, B. Walczak, D.L. Massart, P. Reschiglian, Chemom. Intell. Lab. Syst. 48 (1999) 21.
[13] U. Depczynski, K. Jetter, K. Molt, A. Niemoller, Chemom. Intell. Lab. Syst. 49 (1999) 151.
[14] C.R. Mittermayr, S.G. Nikolov, H. Hutter, M. Grasserbauer, Chemom. Intell. Lab. Syst. 34 (1996) 187.
[15] B.R. Bakshi, J. Chemom. 13 (1999) 415.
[16] A. Papoulis, Probability, Random Variables and Stochastic Processes, third ed., McGraw-Hill, New York, 1991.
[17] E.M. Barj, M. Afifi, A.A. Idrissi, K. Nassim, S. Rachafi, Opt. Laser Tech. 38 (2006) 506.
[18] X.H. Wang, R.S.H. Istepanian, Y.H. Song, IEEE Trans. Nanobiosci. 2 (2003) 184.
[19] S. Foucher, G.B. Bénié, J.M. Boucher, IEEE Trans. Image Proc. 10 (2001) 49.
[20] M. Dai, C. Peng, A.K. Chan, D. Loguinov, IEEE Trans. Geosci. Remote Sens. 42 (2004) 1642.
[21] M.J. Shensha, IEEE Trans. Signal Proc. 40 (1992) 2464.
[22] B. Walzac, D.L. Massart, Chemom. Intell. Lab. Syst. 36 (1997) 81.
[23] X.G. Shao, W.S. Cai, Anal. Lett. 32 (1999) 743.
[24] L. Gao, S. Ren, Spectrochim. Acta A 65 (2005) 3013.