SIGNAL COMPACTION BY MAXIMUM VERISIMILITUDE

Dan STEFANOIU[1], Janetta CULITA[2]

De regulă, datele eşantionate furnizate de un proces sunt afectate de perturbaţii. Deşi înainte de prelucrarea datelor este necesară deparazitarea lor, zgomotele perturbatoare pot îngloba o parte a informaţiei dorite. În acest caz, este adecvată o deparazitare parţială, ca etapă preliminară a procedurii de prelucrare. O altă cerinţă frecvent întâlnită în pre-procesarea datelor o constituie compresia informaţiei utile într-un număr mai mic de date, care împreună cu deparazitarea parţială descriu conceptul de "compactare de semnal". Articolul prezintă o metodă originală de obţinere a datelor preliminare compactate, bazată pe conceptul de "verosimilitate", nefiind necesar un semnal de sincronizare care însoţeşte datele achiziţionate.

Sampled data provided by a process are usually corrupted by various noises. Although data are required to be denoised before applying any further processing, the corrupting noises could encode a part of desired information. A partial denoising is suitable in this case, as part of pre-processing procedure. Another requirement frequently used in data pre-processing is to compress the useful information in a smaller number of data. Partial denoising and compression are gathered in the concept of "signal compaction". The paper introduces an original method for providing compacted preliminary data, relied on the concept of "verisimilitude" and requiring no synchronization signal accompanying the acquired data.

Keywords: time domain synchronous averaging, maximum verisimilitude

1. Introduction

An interesting problem of signal pre-processing (spp) is to extract a "useful" signal from noise-corrupted acquired data. The "usefulness" is regarded here from two points of view. Firstly, a part of noise should be attenuated such that the information carried by the initial signal is mostly encoded by the extracted signal as well. One refers to this operation as *denoising*. Secondly, some applications (like e.g. image interpretation or fault diagnosis and detection)

[1] Prof., Dept. of Automatic Control and Computer Science, University "Politehnica" of Bucharest, Romania. (E-mail: danny@indinf.pub.ro)
[2] Lecturer, Dept. of Automatic Control and Computer Science, University "Politehnica" of Bucharest, Romania. (E-mail: jculita@yahoo.com)

require redundant data. For other applications (such as speech or image coding), on the contrary, the less redundant the data, the better the performance. Redundancy is usually quantified by the *compression ratio*, coarsely computed as the ratio between the extracted signal and the original signal sizes. In general, pre-processed signals are not required to be maximally compressed, but to preserve some redundancy in a smaller number of samples. By definition, one refers to partial denoising together with partial compression as *(signal) compaction*.

The first difficulty in constructing a compacted signal arises from the fact that the rule of combination between the true (deterministic) data and the (stochastic) noise is unknown. (The noise is often considered as white Gaussian, with null statistical mean, as consequence of Central Limit Theorem). A second difficulty is concerned with the separation between the deterministic and the stochastic components of data. Even in case of signal-noise superposition, it is extremely difficult, if not impossible, to draw a fine line between the two components. Therefore, the compacted signal may contain a part of noise, whereas the removed signal may include some useful information that should not have been removed. The quality of the resulted separation strongly depends upon the denoising method and is assessed by means of *Signal-to-Noise Ratio* (SNR).

Depending on the SNR and on the subsequent purpose of pre-processed data, one can identify three main classes of signal extraction methods, based on models in *time*, in *frequency* or in *time-frequency*. In *time domain*, the oldest approach is perhaps based on interpolation techniques, mostly coming from early works of classical Mathematics. The interpolation model includes in general the most part of initial noise, because the model is maximally fitted to data. Another approach using parametric models is to find the waveform which matches the data the best, according to a given criterion, usually *Least Squares* (LS) based. The waveform results by means of a system identification recipe [1]. In general, the identification methods are however quite complex. Within the spp stage, simpler methods are usually preferred.

A very practical approach in time domain is concerned with *averaging* and non-parametric models. The average is however a very coarse estimation of data provider behavior. In spite of such a limitation, when cleverly used, the average could outperform other more sophisticated models in simplicity and effectiveness. In [2] has been devised a technique known as *Time Domain Synchronous Averaging* (TDSA). A combination between TDSA and Lagrange interpolation techniques is introduced in [3]. Nowadays, TDSA is employed in numerous applications where data are collected from rotating machinery, despite the fact it requires 2 data sets: the main one and the synchronization impulses.

In *frequency domain*, the basic methods of compaction during spp are based upon spectral estimation techniques. A good description of these techniques can be found in [4] or [5] and they aim to provide a smoother spectrum than the original one. Smoothing the spectrum implies noise attenuation and a more accurate estimation of compacted signal spectrum. The frequency approaches operate within the hypothesis that the measured data are *quasi-stationary*, i.e. their spectrum is quasi-constant in time. But, in general, the most systems provide data which are more or less non-stationary. Therefore, an approach in joint *time-frequency domain* is more suitable when operating with large sets of non-stationary data. A quasi complete description of time-frequency analysis methods can be found in [6]. In spp, the time-frequency methods are seldom employed because of their complexity.

The *Frequency Averaging Method* (FAM) described hereafter belongs to the frequency domain approaches. To the best of our knowledge, the method is genuine, i.e. it has not been devised by other scientists. The article is organized as follows. Into the next section the TDSA method is succinctly overviewed. Section 3 is concerned with basic hypotheses regarding the noises and the presentation of Maximum Verisimilitude Method as theoretical basic tool. Within the Section 4, the FAM is described. Simulation examples using artificial and real measured data (bearing vibrations) are given in Section 5. Some interesting insights about method effectiveness are also revealed.

2. On TDSA

The principle of TDSA originates from the early works in Signal Processing. A clear evidence of time domain averaging is reported for example in [7] – the Welch method. But the TDSA technique has been devised by McFadden in [2]. The main idea is to perform the signal compaction in case the measured data are provided by a harmonic system with the output y described as follows:

$$y(t) = x(t) + v(t), \quad \forall t \in \mathbb{R}, \tag{1}$$

where x is the main harmonic signal with a *known* period T_r and v is an *unknown* noise. For example, such signals are returned by rotating machineries, where T_r is the rotation period, usually known or measurable. No assumptions are made regarding the noise v in [2], but, obviously, its average has been considered null. The problem is to extract x from y, i.e. to provide an estimation of signal x. The compacted signal is then constructed by restricting x to one period length.

The solution of this problem is devised in 2 steps. Firstly, x is estimated by the *averaged signal* below:

$$a_N(t) = \frac{1}{N}\sum_{n=0}^{N-1} y\left(t + nT_r\right), \quad \forall\, t \in \mathbb{R}, \tag{2}$$

where $N \geq 1$ is the number of periods on which the average should be computed. Obviously, a_N can be expressed as the output of a *comb filter* c_N [4], [5], inputted by y, with selectivity controlled by N. There are two main drawbacks of this approach: the period T_r must accurately be known and a_N is not necessarily periodic, as the sum in (2) is finite. Thus, the input signal should be measured over an infinite horizon of time, albeit the average sum in (2) is finite. In the second step, y is sampled with a period T_s. The measuring horizon of time is finite and included into $[0, NT_r]$ interval. Then the sampling operation can be simulated in continuous time by multiplying the data y with an infinite train of equally spaced Dirac impulses:

$$s(t) \overset{def}{=} \sum_{k\in\mathbb{Z}} \delta_0\left(t - kT_s\right), \quad \forall\, t \in \mathbb{R}. \tag{3}$$

If data are restricted to some measuring horizon by window w_N (usually rectangular on $[0, NT_r]$), the average signal approximating the main harmonic x can be defined as follows (by accounting equation (3)):

$$a_N(t) \overset{def}{=} \frac{s(t)w_N(t)}{N}\sum_{n=0}^{N-1} y\left(t + nT_r\right), \quad \forall\, t \in \mathbb{R}. \tag{4}$$

In definition (4), the comb filter has been used again. Unlike in definition (2), the time range of signal a_N in (4) has to be restricted to one period: $t \in [0, T_r)$, due to windowing. The signal could extend beyond this interval, but its values are less accurate. In definition (4), one can operate with finite data sets (because of windowing), which improves the definition (2). However, in this case, a_N is not necessarily periodic as well, because definitions (2) and (4) are based on the same averaging technique. Unfortunately, the control over the filter selectivity is lost: increasing N does not necessarily improve a_N accuracy. Also, the main period T_r still has to be known in advance. This is the major restriction of TDSA.

Nonetheless, TDSA is very appealing in applications due to its simplicity, albeit, in practice, not only T_r cannot accurately be known, but it usually varies in time. Therefore, a *synchronization signal* accompanying the data is necessary. This signal looks like a comb as well and slides over the data. For any position of comb over the data, the average is computed by extracting only the values pointed by the comb teeth. Obviously, this mechanism works identically whenever the teeth are equally spaced or not. Providing the synchronization signal for data is however not always an easy task.

In [2] a Fourier analysis is also performed, but one operates with continuous time signals, which forced the extension of FT definition to an infinite train of Dirac impulses. An interesting effect of TDSA (not emphasized in McFadden articles) is concerned with the discrete frequency representation of signals. The difficulty in giving frequency interpretations in [2] results from the ad hoc transfer of entities that naturally lie into the framework of discrete time signals to continuous time. A simpler approach is to consider only discrete time signals and to work with the *Discrete Fourier Transform* (DFT) [4], [5]. Compute the number of samples acquired during one period, T_r:

$$K_s \overset{def}{=} \lfloor T_r / T_s \rfloor . \tag{5}$$

The integer K_s of (5) plays the role of estimated period, in terms of normalized time, for the harmonic signal x (with an error smaller than T_s). Assume the synchronization signal consists of N impulses located at instants: $K_0 = 0 < K_1 < \cdots < K_{N-1}$. Then the average signal a_N is simply expressed on one harmonic period by the following definition:

$$a_N[k] \overset{def}{=} \frac{1}{N} \sum_{n=0}^{N-1} y\left[k + K_n\right], \quad \forall\, k \in \overline{0, K_s - 1}. \tag{6}$$

Obviously, the definition (6) relies on the finite set of acquired data $y[p] = y(pT_s)$, $p \in \overline{0, NK_s - 1}$ and works exactly when the sampling period divides the harmonic period. Write in capitals the DFTs associated to the discrete signals above. Then the following remarkable result holds:

Theorem 1 *The DFT of average signal a_N is expressed as a weighted average of DFTs applied on initial data y, for one single harmonic period K_s. More specific:*

$$A_N[p] = \frac{1}{N}\sum_{n=0}^{N-1} Y_n[p]\mathrm{e}^{-2\pi jpK_n/K_s}\,, \quad \forall k \in \overline{0, K_s - 1}, \tag{7}$$

where:

$$Y_n[p] \stackrel{def}{=} \sum_{q=K_n}^{K_n+K_s-1} y[q]\mathrm{e}^{-2\pi jqp/K_s}\,, \quad \forall p \in \overline{0, K_s - 1}. \tag{8}$$

(The proof is straightforward and therefore omitted.)

Theorem 1 shows that the frequency contents of compacted signal can be estimated by the following scenario: segment the data into N successive frames, compute the DFT of order K_s for each resulted frame and average the results.

Consider the N computed DFTs are concatenated in a "frequency signal". Then, in case of uniform synchronization, equations (7) and (8) become:

$$A_N[p] = \frac{1}{N}\sum_{n=0}^{N-1} Y_n[p]\,, \quad \forall p \in \overline{0, K_s - 1} \tag{9}$$

(i.e. simply the average) and, respectively:

$$Y_n[p] \stackrel{def}{=} \sum_{q=nK_s}^{(n+1)K_s-1} y[q]\mathrm{e}^{-2\pi jqp/K_s}\,, \quad \forall p \in \overline{0, K_s - 1} \tag{10}$$

(i.e. the frames do not overlap – see the sum limits). A sliding comb with N equally spaced teeth at K_s instants can be used to extract the averaged values for each of the K_s positions in a period. This picture is quite intuitive and relies on the hypothesis that, if the main harmonic has a constant period K_s, then the N DFTs are quite similar and thus, by averaging them, a noise reduction is obtained. Moreover, the interpretation holds for non-uniform synchronization as well. But, in this case, the frames could overlap, depending of index values in (6).

3. Noise Hypotheses and MVM

Return to equation (1) and consider that the measured data y, the compacted signal x and the noise v are discrete time signals. Usually, the data set includes N acquired samples measured from a system. Thus, y is a finite length discrete time signal with support included in $\overline{0, N-1}$. One wants that the compacted signal x be a finite length discrete time signal as well, but with smaller support, for example inside $\overline{0, P-1}$, where $P \le N$. Finally, the noise v is also discrete time but not necessarily additive to x in the sense of equation (1). (Actually, in this context, the signals have different supports.) Two natural hypotheses regarding the noise are assumed, as it will be shown next.

One can split the discrete spectrum of y into M non overlapped sub-bands with the same width. Consider for simplicity, that M is a divisor of N, i.e. $N = K \cdot M$ for some $K \ge 1$. Obviously, this condition is not very restrictive. Also, denote by DFT_l the DFT operator applied to a signal of length l. Then the natural hypotheses below are assumed hereafter:

- H_1 The DFT of signal y is affected by a set of M complex valued and additive sub-band noises V_m, $m \in \overline{0, M-1}$ with finite supports included into corresponding sub-bands. (Thus, the noises V_m are orthogonal each other.)
- H_2 Noises V_m are white Gaussian with null mean and variances λ_m^2, $m \in \overline{0, M-1}$ (unknown).

According to hypothesis H_1, the following model of $Y \equiv DFT_N(y)$ can be defined for any $m \in \overline{0, M-1}$ and $k \in \overline{0, K-1}$:

$$Y[mK + k] = A_m[k] + V_m[k], \tag{11}$$

where: $mK + k = n \in \overline{0, N-1}$ has been expressed by using the *Theorem of Division with Reminder* (TDR) and A_m is a deterministic model of DFT. Usually, A_m is an auto-regressive model or a polynomial. By concatenation of all A_m models, the DFT of compacted signal x is obtained. The overall model (11) is a description of DFT for every frequency sub-band. Hypothesis H_2 can be expressed as follows, for any $m \in \overline{0, M-1}$, $l \in \overline{0, K-1}$ and $k \in \overline{0, K-1}$:

$$\not{p}\left(V_m[k]\right) = \frac{1}{\sqrt{2\pi}\lambda_m} \exp\left(-\frac{\left|V_m[k]\right|^2}{2\lambda_m^2}\right); \quad E\left\{V_m[k]V_m^*[l]\right\} = \lambda_m^2 \delta_0[k-l]. \tag{12}$$

In (12), $\not{p}$ is the density of probability, E stands for the statistical average operator (the *expecting* operator), a^* is the complex conjugate of a and $\delta_0[\bullet]$

denotes the discrete unit impulse (the Kronecker symbol). As consequence of hypothesis H_2, every noise V_m has independent values.

The *problem* is to provide an estimation of deterministic models A_m ($m \in \overline{0, M-1}$), by using the DFT of measured data $Y[n]$, $n \in \overline{0, N-1}$. The estimations of variances λ_m^2, ($m \in \overline{0, M-1}$) can be used to assess the models accuracy. Intuitively, one wants that the spectrum of compacted signal *keeps the appearance* of original data spectrum. With other words, the two spectra should exhibit similar shapes, but the compacted spectrum must be less noisy. This requirement can be quantified by means of *verisimilitude* concept, which comes from System Identification [1]. Thus, in order to estimate the parameters of models A_m, the *Maximum Verisimilitude Method* (MVM) can be employed [1], [9]. Before describing the MVM, consider that every model A_m is a linear:

$$A_m[k] = \boldsymbol{\varphi}_m^T[k]\boldsymbol{\theta}_m , \quad \forall\, k \in \mathbb{N} , \tag{13}$$

where $\boldsymbol{\varphi}_m[k]$ (data) and $\boldsymbol{\theta}_m$ (parameters) are column vectors with the same length (the number of parameters), while T denotes the transposition operator. The linearity of model aims to keep the low complexity required by spp methods. For example, if A_m is a polynomial with degree p_m:

$$A_m[k] = \alpha_{m,0} + \alpha_{m,1}k + \cdots + \alpha_{m,p_m} k^{p_m} , \quad \forall\, k \in \mathbb{N} , \tag{14}$$

then, in (13): $\boldsymbol{\varphi}_m^T[k] = \begin{bmatrix} 1 \; k \; \cdots \; k^{p_m} \end{bmatrix}$ encompasses the data, real valued, while $\boldsymbol{\theta}_m^T = \begin{bmatrix} \alpha_{m,0} \; \alpha_{m,1} \; \cdots \; \alpha_{m,p_m} \end{bmatrix}$ includes all parameters, real or complex valued.

If $\boldsymbol{\Theta}_m$ is the vector of parameters $\boldsymbol{\theta}_m$ extended by the unknown variance λ_m^2, the MVM estimation is defined by:

$$\hat{\boldsymbol{\Theta}}_m \stackrel{def}{=} \arg\max_{\boldsymbol{\Theta}_m \in \mathcal{S}_m} \wp\left(Y_m \,|\, \boldsymbol{\Theta}_m\right) , \quad \forall\, m \in \overline{0, M-1} . \tag{15}$$

In (15), $\mathcal{S}_m$ is the stability domain of model A_m, the data segment in sub-band $m \in \overline{0, M-1}$ is $Y_m = \{Y[mK + k]\}_{k \in \overline{0, K-1}}$ and $\wp(Y_m \,|\, \boldsymbol{\Theta}_m)$ is the density of conditional probability between the data Y_m and the parameters $\boldsymbol{\Theta}_m$. Thus, the parameters should be selected such that the measured data occur with a maximum probability. This actually means that the parameters have *maximum verisimilitude* to the measured data. Without any knowledge about the noises affecting the data, the equation (16) has little impact in practice. Hypotheses H_1 and H_2 lead however the following result.

Theorem 2 *Under H_1 and H_2 hypotheses, the MVM estimation (15) is identical to LS estimation, i.e., for any* $m \in \overline{0, M-1}$*:*

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_m = \left[\frac{1}{K}\sum_{k=0}^{K-1} \boldsymbol{\varphi}_m[k]\boldsymbol{\varphi}_m^T[k]\right]^{-1}\left[\frac{1}{K}\sum_{k=0}^{K-1} \boldsymbol{\varphi}_m[k]Y[mK+k]\right] \\ \hat{\lambda}_m^2 = \frac{1}{K}\sum_{k=0}^{K-1}\left|Y[mK+k] - \boldsymbol{\varphi}_m^T[k]\hat{\boldsymbol{\theta}}_m\right|^2 \end{bmatrix}. \tag{16}$$

(The proof is well known and requires no special manipulations.)

Theorem 2 simplifies the approach and keeps the same interpretation regarding the maximum of verisimilitude. Moreover, due to LS properties (see [1] or [9]), the estimates (16) are convergent to the true values as K increases.

The complexity of model (11) and the computational effort can be controlled through the selected number of parameters (e.g. $p_m + 1$ in model (14)). One of the simplest models is obtained by selecting $p_m = 0$ in (14). In this case, due to (16), the model becomes:

$$\hat{A}_m = \frac{1}{K}\sum_{q=0}^{K-1} Y[mK+q], \quad \forall\, m \in \overline{0, M-1}, \tag{17}$$

i.e. it is expressed as simple averages of frequency data in corresponding sub-bands. Obviously, there is a big difference between equations (9) and (17). Within TDSA, the averages are computed following the comb rule, whereas by means of MVM frequency averaging, consecutive values inside the same sub-band are employed. Also, note that, in (17), not the spectral values are used, but the frequency data obtained by computing the DFT of original data.

4. The FAM

The MVM has theoretically shown how the DFT of compacted signal can be estimated, by using frequency data. Theorem 2 involves that the accuracy of estimation improves with the number of spectral lines allocated to each sub-band (K). Usually, the number of sub-bands (M) is constant and thus, the accuracy increases if the number of acquired data (N) increases. But the most interesting consequence of Theorem 2 is concerned with the reconstruction of compacted signal from its DFT, which constitutes the core of FAM.

After the models A_m being estimated, the DFT of compacted signal results by concatenation:

$$\hat{X}[mK+k] = \hat{A}_m[k], \quad \forall\, m \in \overline{0, M-1},\ \forall\, k \in \overline{0, M-1}. \tag{18}$$

In (18), the TDR has been invoked again. The models A_m being estimated by MVM, the DFT of compacted signal is the nearest deterministic waveform to the DFT of initial data, in the LS sense. So, the spectrum of compacted signal keeps the best the appearance of original data spectrum and, moreover, it is smoother.

In equation (18), one can see that the compacted signal has the same support as the original measured one, i.e. $P = N$. Thus, the original data have (partially) been denoised, but not compressed. Some compression is achieved if each of the models A_m is interpolated in a smaller number of spectral lines than K. Let $L < K$ be the number of interpolation spectral lines to be considered. Then $P = ML < MK = N$ and the DFT of compacted signal is similarly expressed like in (18), but L replaces K.

Finally, the estimated compacted signal is computed with the help of an inverse DFT_P. The FAM is then described by the following procedure:

Step 1. Compute the frequency data $Y \equiv DFT_N(y)$.

Step 2. Use MVM to estimate the deterministic models $\{A_m\}_{m\in\overline{0,M-1}}$.

Step 3. Perform the interpolation of each model $\{A_m\}_{m\in\overline{0,M-1}}$ in L equally spaced spectral lines, with $L < K$.

Step 4. Construct the DFT_P of compacted signal by concatenation, like in equation (18) (with L instead of K), where $P = ML$.

Step 5. Apply the inverse DFT_P to estimate the time values of compacted signal $\hat{x}$ on $\overline{0, P-1}$.

The procedure above is quite general and only requires that M be a divisor of N. No synchronization signal is necessary and even the main (rotation) harmonic period T_r could miss or be unknown. Thus, even asynchronous signals can be compacted by using FAM, which is not possible with TDSA. However, if T_r exists and can be estimated, then the number of interpolation spectral lines L has to be set accordingly. Usually, it is suitable that the compacted signal be represented on a small number of main rotations (up to 10), which involves that the maximum number of samples and the corresponding number of spectral lines in each sub-band can easily be derived. However, it is not necessary that T_r be known with high accuracy. If inaccurately estimated, the compacted signal will only lie in a support with length non divisible by T_r. For example, if one wants x to be represented on 5 full rotations, but $0.9 \cdot T_r$ is used instead of T_r, then the resulted support length is $4.5 \times T_r$ instead of $5 \times T_r$. The information about the main harmonic period is basically not affected, even in case this period is variable. Also, by interpolation, the aliasing is avoided, since the operation is applied on DFT and not on data.

The main drawback of FAM in its general form is the computational effort. If the effect of interpolation is ignored, then the procedure requires about $(N \log_2 N + P \log_2 P + MK^3)$ operations. The interpolation is more or less increasing this number.

A difficulty when using FAM is the selection of parameters N, M and K as result of a trade-off. On one hand, the MVM estimates are accurate for a big number of spectral lines per sub-band K, which involves either N is large or M is small. On the other hand, the original spectrum is better “imitated” by the compacted one if the number of sub-bands M is large enough (i.e. the bandwidth is small enough), which involves either N is large or K is small. Since both M and K must be set with sufficiently large values, this involves the number of acquired data N has to be large. This is the price paid for the absence of synchronization signal.

An interesting particular case results when considering constant polynomials as deterministic models, like in equation (17). In this case, the constants are simply the averages computed over frequency data in every sub-band. If the sub-bands are narrow enough, one can substitute the local DFT variation by the frequency data averages, as instantaneous frequency contents. This assumption involves a reduction of computational effort, which makes the FAM appealing in applications. The compacted signal lies inside the support $\overline{0, M-1}$ and can directly be constructed, as outlined by the following result.

<u>Theorem 3</u> *With the deterministic models (17), the compacted signal* $\hat{x}$ *can be estimated as follows:*

$$\begin{cases} \hat{x}[0] = y[0] & , \ m = 0 \\ \hat{x}[m] = \dfrac{1 - w_M^m}{K} \displaystyle\sum_{k=0}^{K-1} \dfrac{y[kM+m]}{1 - w_K^k w_N^m} & , \ m \in \overline{1, M-1} \end{cases} . \tag{19}$$

(The proof only relies on algebraic manipulations and is therefore omitted.)

If M is equal to period K_s (see definition (5)) and the time average is restricted to K periods, then the compacted signal resulted by applying TDSA is the following (according to equation (6)):

$$a_K[m] \overset{def}{=} \frac{1}{K}\sum_{k=0}^{K-1} y[kM+m], \quad \forall m \in \overline{0, M-1}. \tag{20}$$

Obviously, although equations (19) and (20) are different, they still have a common kernel represented by "$y[kM+m]$", i.e. the comb rule in computing the averages is the same. Theorem 3 shows that, when using FAM, one operates with weighted averages of initial data, such that the spectrum of compacted signal would keep the appearance of initial spectrum.

The amount of computations necessary to evaluate $\hat{x}$ from equations (19) is about $(8K+5)(M-1)$ operations. (No interpolation is necessary.) This amount is sensibly lower than the number involved by the general procedure of FAM.

5. Simulation results

Implementation of FAM in particular case of model (17) arises no special problems. The only restriction that should be verified is $N = MK$. This is however only a soft requirement.

The FAM has been implemented in MATLAB environment with model (17). The following data have been employed in simulation experiments:

1. A sine wave of period K_s, corrupted by a Gaussian white noise.
2. A raw vibration signal of length N.
3. The high pass filtered vibration signal of 2.

The compacted signal is represented on $M < N$ samples in all cases.

Fig. 1 illustrates a sine wave of period $K_s = 500$ (a) that has been compacted to signals with lengths $M = 333$ (b) and $M = 71$ (c).

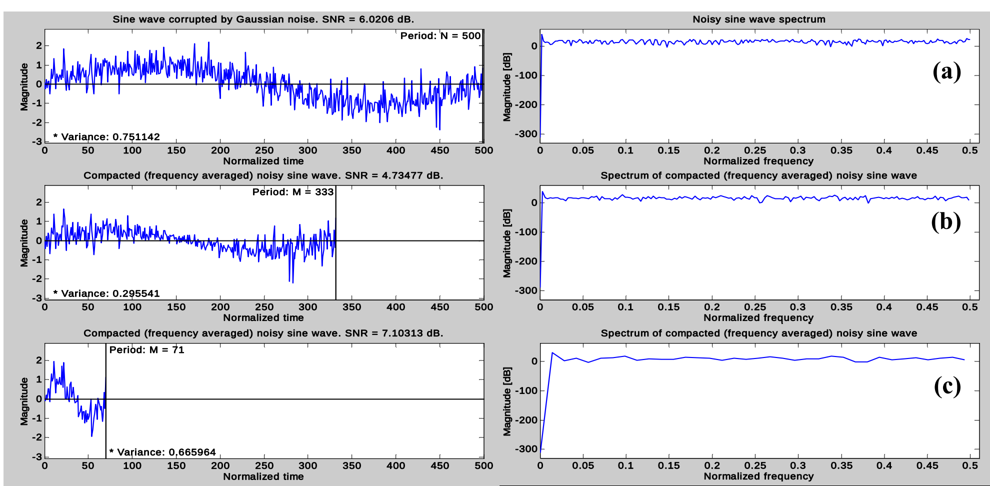

Fig. 1. A noisy sine wave and two compacted signals with their spectra.

None of these lengths are divisors of K_s. However, both compacted signals recovered the shape of the original one in time as well as in frequency. On the right column of figure, the corresponding spectra are depicted. The sine wave was corrupted by a quite strong Gaussian white noise ($\text{SNR} \cong 6$ dB). The spectra of compacted signals in Figs. 1(b) and (c) are smoother than the original spectrum (see the right column), but the resulted SNR is not necessarily increasing. Thus, for $M = 333$ (b) the SNR is smaller (4.73 dB), while for $M = 71$ (c) the SNR is bigger (7.1 dB). This implies that increasing the SNR has to be realized by selecting N, M and K appropriately. Unfortunately, the variation of resulted SNR with M (for a given N) is extremely non linear, as Fig. 2 displays.

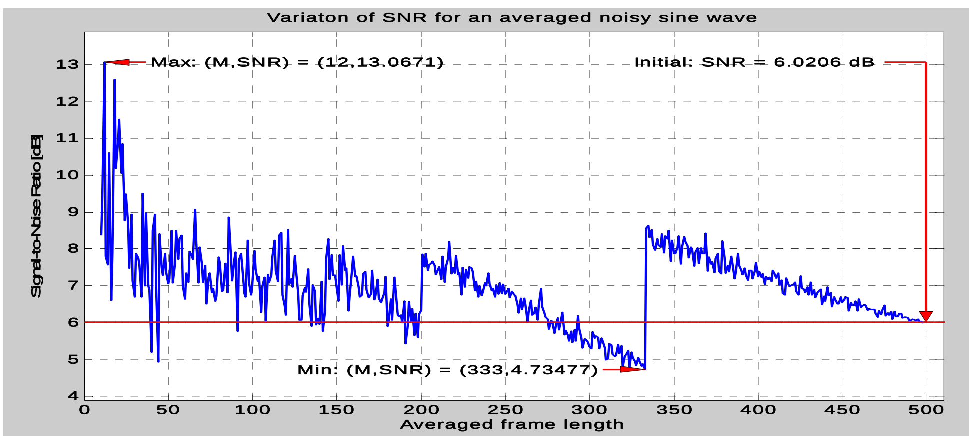

Fig. 2. SNR versus the compacted support length.

This phenomenon has been confirmed by different simulations, with different noises and initial SNR values. Thus, sometimes SNR is below the original one, though these cases are by far less numerous than the cases where SNR increases. The selected $M = 333$ happened to be the worst case in this example. In general, there are many possible selections of M such that the resulted SNR sensibly increases. The figure also shows that the best chances to increase SNR are obtained when $M << N$, i.e. when the number of acquired data is large enough.

A harmonic vibration has been acquired from a bearing with rolling balls, in order to perform fault detection and diagnosis [8]. The signal is depicted in Fig. 3(a), on top. The sampling rate was 20 kHz and the signal length is 809.35 ms ($N = 16\,187$ samples). During the measurements, one has noticed that the rotation speed varies around the nominal value of 44.37 Hz (about 2662 rpm) with a variance of more than $\pm 10\%$. The speed variation is mainly due to a non uniform load, but probably the defect that started to develop on the inner race of bearing also plays a role in this matter. This variation cannot easily be distinguished within the figure. So, without accurately measuring the rotation speed, the period of one rotation has been set to $T_r = 22.53$ ms (about 451 samples). Consequently, the support length of compacted signal has been established to 4 full rotations, i.e. $M = 1804$ samples. Note that M is not a divisor of N and all acquired data have been considered, without truncation or zero-padding. The resulted compacted signal is depicted on bottom of Fig. 3(a). A bit more than 4 full rotations can clearly be seen, because of the mentioned speed variation. As expected, the poor estimation of T_r has practically no influence on the compacted signal.

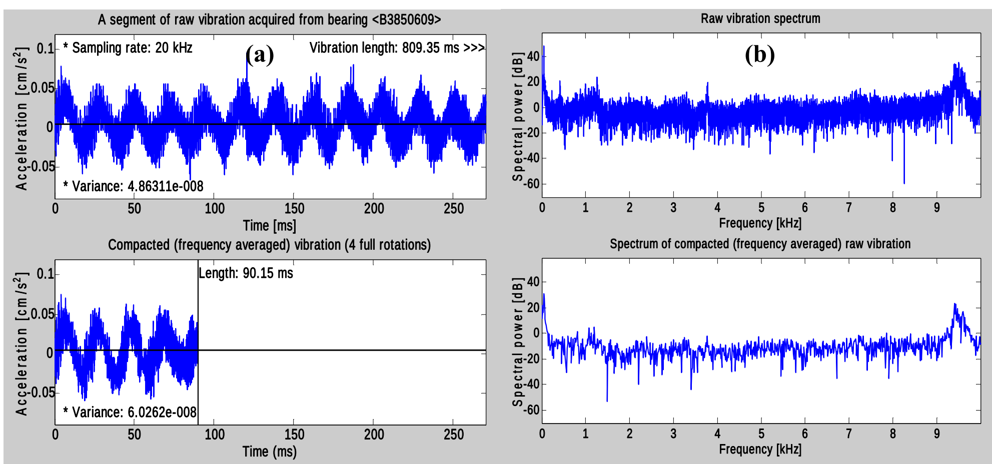

Fig. 3. A raw vibration (a, top), the compacted signal (a, bottom) and spectra (b).

The original signal looks very noisy. The estimated SNR is about 3.27 dB. However, FAM led to a noise reduction. Visually, the compacted signal looks less noisy. In order to quantify this observation, the SNR of compacted signal has been estimated by means of the best sine wave passing through the signal in the LS sense. It has been derived that the new SNR increased to 10.53 dB. The denoising effect is also emphasized by the spectra depicted in Fig. 3(b). The compacted signal spectrum is smoother than the original spectrum, keeping the same shape.

The vibration has been filtered by a high pass filter with cut-off frequency of 500 Hz (more than 10 times the main rotation frequency), in order to remove/attenuate the main harmonic and the natural harmonics of bearing. The resulted signal is quite asynchronous, in the sense that no predominant harmonic can easily be detected. Fig. 4(a) shows on top the filtered vibration. The apparent low frequency harmonic is due to the modulation between 2 noises: one encoding the information of bearing defect and another one issued from the environment and interference with different sources of vibration. The parameters of compacted signal depicted on bottom of Fig. 4(a) are the same (i.e. 4 full rotations long), but, this time, the main rotation cannot be seen. This example clearly shows how asynchronous (not necessarily harmonic signals) can be compacted by using the FAM. The previous remarks regarding the noise reduction hold in this example as well (see also the corresponding spectra in Fig. 4 (b)).

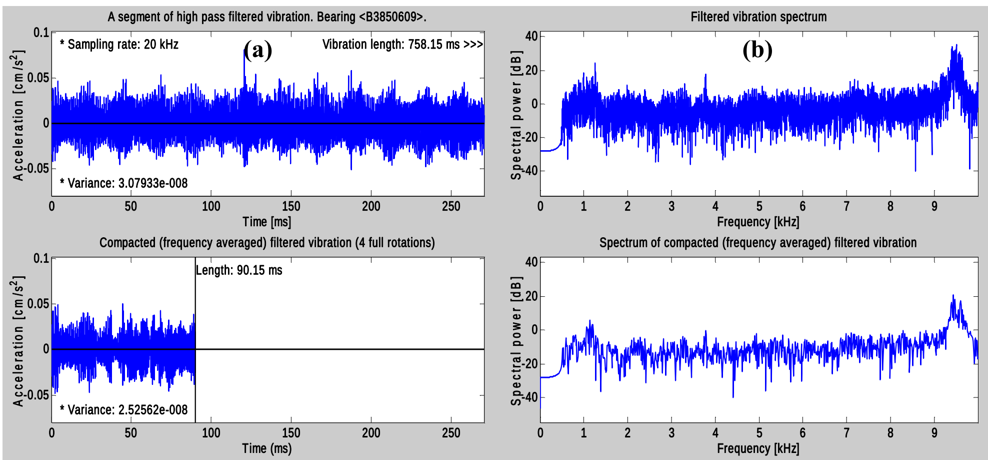

Fig. 4. A filtered vibration (a, top), the compacted signal (a, bottom) and spectra (b).

6. Conclusion

This paper dealt with the problem of signal compaction (partially denoising and compression). An alternative to TDSA technique has been introduced. The novel approach relies on frequency averaging with maximum verisimilitude between the original and the compacted FT. The simulations revealed many insights regarding the effectiveness of FAM, but its limitations too. Nonetheless, the FAM might be interesting in practice since no synchronization signals are required and asynchronous signals can also be compacted.

R E F E R E N C E S

[1] *T.Söderström, P. Stoica*, System Identification, Prentice Hall, London, UK, 1989.

[2] *P.D. McFadden*, “A Revised Model for the Extraction of Periodic Waveforms by Time Domain Averaging”, in Mech. Systems and Signal Processing, Vol. **1**, No. 1, 1987, pp. 83–95.

[3] *P.D. McFadden*, “Interpolation Techniques for Time Domain Averaging of Gear Vibration”, in Mechanical Systems and Signal Processing, Vol. **3**, No. 1, 1989, pp. 87–97.
[4] *A.V. Oppenheim, R. Schafer,* Digital Signal Processing, Prentice Hall, New York, USA, 1985.
[5] *J.G. Proakis, D.G. Manolakis*, Digital Signal Processing. Principles, Algorithms and Applications., Prentice Hall, New Jersey, USA, 1996.
[6] *L. Cohen,* Time-Frequency Analysis, Prentice Hall, New Jersey, USA, 1995.
[7] *P.D. Welch*, “The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short Modified Periodograms”, in IEEE Transactions on Audio and Electroacoustics, vol. **AU-15**, June 1967, pp. 70–73.
[8] *F. Ionescu , D. Arotăriței,* “Fault Diagnosis of Bearings by Neuro-Fuzzy Classification”, ISMA’23 Conference, KU-Leuven, Belgium, Sep. 16-18, 1998.
[9] *D. Stefanoiu, J. Culita, P. Stoica*, A Foundation of System Modeling and Identification, Printech Press, Bucharest, Romania, 2005 (in Romanian).