Jrl Syst Sci & Complexity (2007) 20: 325–343

DENOISING AND HARMONIC DETECTION USING NONORTHOGONAL WAVELET PACKETS IN INDUSTRIAL APPLICATIONS

P. MERCORELLI

Received: 8 May 2006 / Revised: 20 December 2006

Abstract New industrial applications call for new methods and new ideas in signal analysis. Wavelet packets are new tools in industrial applications and they have just recently appeared in projects and patents. In training neural networks, for the sake of dimensionality and of ratio of time, compact information is needed. This paper deals with simultaneous noise suppression and signal compression of quasi-harmonic signals. A quasi-harmonic signal is a signal with one dominant harmonic and some more sub harmonics in superposition. Such signals often occur in rail vehicle systems, in which noisy signals are present. Typically, they are signals which come from rail overhead power lines and are generated by intermodulation phenomena and radio interferences. An important task is to monitor and recognize them. This paper proposes an algorithm to differentiate discrete signals from their noisy observations using a library of nonorthonormal bases. The algorithm combines the shrinkage technique and techniques in regression analysis using Shannon Entropy function and Cross Entropy function to select the best discernable bases. Cosine and sine wavelet bases in wavelet packets are used. The algorithm is totally general and can be used in many industrial applications. The effectiveness of the proposed method consists of using as few as possible samples of the measured signal and in the meantime highlighting the difference between the noise and the desired signal. The problem is a difficult one, but well posed. In fact, compression reduces the level of the measured noise and undesired signals but introduces the well known compression noise. The goal is to extract a coherent signal from the measured signal which will be "well represented" by suitable waveforms and a noisy signal or incoherent signal which cannot be "compressed well" by the waveforms. Recursive residual iterations with cosine and sine bases allow the extraction of elements of the required signal and the noise. The algorithm that has been developed is utilized as a filter to extract features for training neural networks. It is currently integrated in the inferential modelling platform of the unit for Advanced Control and Simulation Solutions within ABB's industry division. An application using real measured data from an electrical railway line is presented to illustrate and analyze the effectiveness of the proposed method. Another industrial application in fault detection, in which coherent and incoherent signals are univocally visible, is also shown.

Key words Data compression, denoising, rail vehicle control, trigonometric bases, wavelet packets.

P. MERCORELLI
Department of Vehicles, Production and Process Engineering, University of Applied Sciences Wolfsburg, Robert-Koch-Platz 8-a., 38440 *Wolfsburg, Germany.* Email: p.mercorelli@fh-wolfsburg.de, mercorelli@ing.unisi.it.

1 Motivations and Introduction to the Paper

1.1 Motivations

The construction of electrical rail vehicles has greatly changed due to advances in the fields of power electronics and computer control. Converter control vehicles are active devices and their characteristics and functionality depend on control algorithms implemented in distributed real-time computer systems. The harmonic interference problem in electrical railway systems has recently received particular attention. The widespread utilization of modern electronic devices, such as GTO-Thyristor (Gate Turn Off-Thyristor) or IGBT (Insulated Gate Bipolar Transistors), can cause interference in signal circuits and communication systems as well as lead to stability problems[1]. Harmonic detection techniques are also of great importance for vehicles. Real-time distortion current monitoring, in many practical situations, is not an easy task because the current magnitude and phase change over time. Several approaches can be found in the literature on rail vehicles as in [2], where an adaptive Kalman filter based on the correlation analysis is proposed. Other works in this direction[3−4] have indicated wavelets as a promising approach for off-line analysis, monitoring and classification of transients in electrical railway systems. In rail vehicles, inrush currents are quasi–harmonic signals characterized by very high rectified current levels. These currents are typically dangerous for electronic power systems. In on-line detection of harmonic features interesting contributions have been presented in [5–6] where solutions to the problem of detection of dominant frequency vibration in pantograph systems are proposed. It has been highlighted in [7–8], that one of the most important problems in rail vehicle control is to model the nonlinearity of the locomotive transformer as well as to classify the transformer inrush current. Progress in this direction is marked by [7] which proposes an efficient algorithm in order to model strong non-linear systems. The transformer inrush current is caused by the transformer nonlinearity and occurs with the discontinuity of the magnetic flow. A typical example of inrush current phenomenon is when the locomotive is passing through a neutral section of line*. Note that usually the number of connections (and disconnections) of the pantograph to the overhead line is very high and this causes a high number of inrush currents to the transformer, thus rapidly degrading the transformer performance. To conclude, the main motivation and the aim of this paper consists of presenting a developed industrial algorithm for extracting relevant features of quasi–harmonic signals from noisy measurements due to interference. In other words, the primary motivation is to separate the noise from the signal. Secondly, such features are used for training neural networks to recognize dangerous from non-dangerous inrush currents. A possible scheme is reported in Fig.1. In accordance with the primary motivation the algorithm presented can be used to separate two signals. An example in which inrush current is extracted from its noise measurements is presented at the end of the paper. Furthermore, the paper presents an application dedicated to fault detection which is obtained through a neural network in which the training signals are filtered through the algorithm introduced here. All these applications show the generality of the technique. From the main and the second motivation it emerges that, when the signal is affected by a high level of noise, as in rail vehicle, the problem becomes difficult but also really interesting. In fact, an over compression reduces the noise but introduces the well known compression noise. By using a small number of parameters the noise due to the compression may be too high. On the contrary by using a big number of parameters the level of the compression may not be enough. Moreover, in the case of data overfitting, the measured and overhead power line noise could actually corrupt the data. This paper presents an optimal algorithm which tries to find a compromise and it is conceived using wavelet bases structured in a tree structure.

*A neutral section of the overhead line is a section of the line with tension equal to zero.

1.2 Introduction to the Paper

This paper proposes an algorithm for signal denoising by using libraries of non-orthogonal bases (frames) such as local smooth trigonometric libraries. This method extracts, from the observed discrete signals, a coherent part which is well represented by the given waveforms and a noisy, or incoherent part which cannot be compressed well by the waveforms. This paper also describes an algorithm which can be used to differentiate a discrete signal from its noisy components using a library of nonorthonormal wavelet packets of smooth trigonometric bases. The proposed technique is essentially a nonparametric regression analysis. The developed algorithm consists of building a map for the values of the Shannon Entropy function on every time-frequency cell of the sine and cosine packets for the measured signal. The libraries are split into two classes: coherent (for instance with the even and odd part of the sine bases) and incoherent (with the even and odd part of the cosine bases) decomposition, by minimizing and maximizing the Shannon Entropy function respectively. Then the time-frequency cells, maximizing the Cross Entropy function between the two groups, are chosen. In such a way one selects the bases with the best compression level and the bases which highlight the difference between the noise (incoherent decomposition) and the signal (coherent decomposition). Recursive residual iterations with sine bases for the coherent decomposition and with cosine bases for the noise allow the reconstruction of the signal and the noise with the best discernable bases. It is known that the Shannon entropy function is a measure of the flatness of the energy distribution of the signal so that its minimization leads to an efficient representation, mainly for signal compression[9]. It is known that the Cross Entropy function is a measure of the discrepancy between two or more bases and can be used to illuminate the difference between the noise and signal, see [10]. It is necessary to define a language to describe the signals. The language must be as versatile as possible in order to describe various local features of the signal. The method must be computationally efficient to be practically applied. The wavelet frames provides a flexible coordinate system with their redundant adaptive time-frequency cells. The smooth trigonometric bases match the desired harmonic signal very well and can detect information in small amounts of coherent data. Furthermore, the non-orthogonal libraries allow more elasticity in order to approximate the measured signals. In fact by relaxing the orthogonality, much more freedom on the choice of the wavelet function is gained to guarantee good choices of the compressed parameters, even though the fast algorithms associated with the orthogonality are lost[11]. In order to consider and use the non-orthogonality of the frames which generate an interaction between the elements of the bases† the algorithm considers, at each step, all the elements of the bases previously selected, without any elimination, see [11]. As the decomposition on a non-orthogonal basis is not unique, it is necessary to stop the algorithm. In order to reduce the dimensionality of the problem the algorithm works on initial compression data (data shrinkage), and uses the best basis paradigm as in [9] or [10] which allows a rapid search among a large collection of bases. The computational complexity is $\mathcal{O}(n(\log(n))^p)$, where p is equal to 1 or 2 depending on the basis type, wavelet dictionaries or trigonometric wavelet dictionaries respectively, being n the length of data signals. The compression method by using wavelets has already been used in control applications as in signal processing, see for instance [12], [13], and [14]. Meanwhile, several works claimed that wavelets are also useful for reducing noise [15], [16], and [17]. This paper tries to take advantage of both. It proposes an algorithm for the simultaneous suppression of random noise in data and the compression of signals. In [18] the minimum description length principle is adopted to find the best decomposition in order to suppress noise and detect the desired signal in the orthogonal wavelet libraries. The proposed algorithm can be used as a filter for the raw signals before applying the classification technique

†In a frame the decomposition is not unique.

already described in [8]. The algorithm is totally general and can be used in any signal feature detection problem. Suffice to say that cosine and sine bases are already used in JPEG (Joint Photographic Experts Group) compress techniques. The processing of real measured data for a railway vehicle line is presented here to illustrate and discuss the effectiveness of the proposed method. The paper is organized as follows. In Section 2 the problem is formalized. In Section 3 the non-orthogonality, the smooth trigonometric wavelet packets and the choice of the best regressor family are discussed. Sections 4 and 5 are devoted to the presentation of the algorithm and the discussion of the results.

2 Problem Formulation

Useful features of inrush current, despite the presence of noisy signals, must be detected in order to recognize this phenomenon and shut down the transformer of the locomotive. The inrush current is a quasi–harmonic signal well described in [4]. Though a rigorous definition of a quasi-harmonic signal does not exist in any literature, it is commonly accepted that:

A quasi-harmonic signal is a signal with one dominant harmonic and "some" (two or three) relevant sub-harmonics in superposition.

In neural networks, in order to obtain short pattern recognition time and a low percentage of errors, a particular training is required and normally demanded

Good compressed features: small amount of data with high level of information.

When the signal is affected by noise, as in a rail vehicle, the problem becomes really difficult. In fact, an over compression reduces the noise but introduces the well known compression noise. To sum up, there are two different requirements in order:

For the sake of data compression the signal should be compressed with a small number of parameters.

For the sake of minimizing the distortion between the estimate and the true signal a great number of parameters are needed.

Now the conflict is clear. By using a small number of parameters the noise due to the compression may be too high, by using a data overfitting the level of the compression may not be enough. Moreover, in the case of data overfitting, the measured and overhead line noise could really corrupt the data. The problem could be stated as a nonparametric model identification problem with little a-priori knowledge in which the choice of the most suitable basis plays a crucial role. In general the main problem could be stated in the following way. Let us consider a discrete degradation model

$$\boldsymbol{d} = \boldsymbol{f} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{d}$, $\boldsymbol{f}$, $\boldsymbol{n} \in \mathcal{X} \subseteq \Re^{d_0}$ and $d_0 = 2^{n_0}$ ($n_0 \in N$). The subspace $\mathcal{X}$ is called signal space and d_0 is the number of samples of the signal. The vector $\boldsymbol{d}$ represents the noisy observed data and $\boldsymbol{f}$ is the unknown true signal to be estimated. The vector $\boldsymbol{n}$ is white Gaussian noise (WGN), whose distribution is assumed to be unknown. The distribution is assumed with an unknown average because of undesired signals from the overhead line, and unknown deviation because of measured errors. The problem consists of differentiating signal $\boldsymbol{f}$ from noisy observation $\boldsymbol{d}$. Assume that the signal $\boldsymbol{f}$ is a quasi-harmonic signal. First, the trigonometric bases are selected. Let $\mathcal{B}$ be the candidate library packet tree to describe the signal $\boldsymbol{f}$.

Problem Given a measured data $\boldsymbol{d} = \boldsymbol{f} + \boldsymbol{n}$ as in (1) where $\boldsymbol{f}$ is assumed to be a "quasi-harmonic" signal, $\boldsymbol{n}$ is Gaussian noise with unknown distribution. Given the library

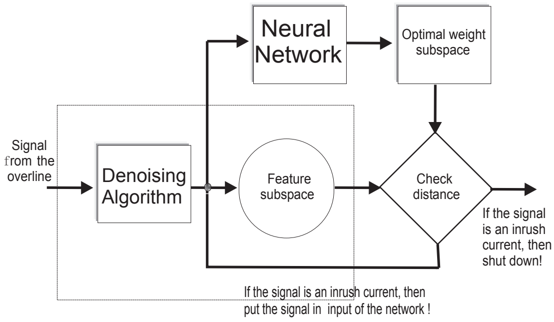

Figure 1 On-line structure to recognize inrush current.

of trigonometric bases $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_M\}$, let $\boldsymbol{f} = \boldsymbol{W}_m \boldsymbol{\alpha}_m^k$, where $\boldsymbol{W}_m \in \Re^{k \times k}$ is a non-orthogonal matrix whose column vectors are the basis elements of $\mathcal{B}_m$, $m = 0, 1, \cdots, M$, and the $\boldsymbol{\alpha}_m^k$ are the expansion coefficients of $\boldsymbol{f}$ with only k no-negligible coefficients. In order to extract relevant features, find a map $\mathcal{K}$, called feature extractor, $\mathcal{K} : \mathcal{X} \to \mathcal{F} \subset \Re^k$ (k is normally more than one) with $k \ll d_0$ such that

$$\min_{\{\boldsymbol{W}_m, \boldsymbol{\alpha}_m^k\}} \widehat{\sigma}^2, \tag{2}$$

where

$$\widehat{\sigma}^2 = \|\boldsymbol{d} - \boldsymbol{W}_m \boldsymbol{\alpha}_m^k\|^2,$$

and $\|\cdot\|$ is the Euclidean norm. Note that in order to reduce the computational complexity, a data shrinkage pre processing phase is required to reduce the number d_0 of samples.

Remark 1 It is useful to note that the function $\boldsymbol{f}$ belongs to the subspace of dimension equal to d_0. In other words, function $\boldsymbol{f}$ could be represented as a column with d_o elements. For the sake of data compression the signal should be compressed with a small number k of parameters. All the m bases of the wavelet library have the dimension $k \times k$. Moreover, it will be explained that the $\boldsymbol{W}_m$ can be composed from linear dependant elements, in other words non-orthogonal bases.

Remark 2 Given a k, an optimum for (1) consists of finding $\{\boldsymbol{W}_m, \boldsymbol{\alpha}_m^k\}$ with an iterative procedure. The non-orthogonality gives more elasticity but at the same time requires more calculation effort.

3 Giving up on Orthogonality

Wavelet and wavelet series are popular in signal processing and numerical analysis. Loosely speaking, a function $f(t)$ can be decomposed into

$$f(t) = \sum_j \sum_n w_{j,n} \boldsymbol{\psi}_{j,n}(t), \tag{3}$$

where $\boldsymbol{\psi}_{j,n}(t)$ are wavelet functions, normally obtained by dilating and translating a mother function $\boldsymbol{\psi}(t)$, index j and n denote dilation and translation respectively and $w_{j,n}$ is the weight coefficient for $\boldsymbol{\psi}_{j,n}(t)$. The most popular algorithms deal with the orthonormal wavelet bases, see [19]. Besides these, wavelet frames, see [19], consist of non-orthogonal wavelet families and are redundant bases. Wavelet frames are more powerful but computationally more involved.

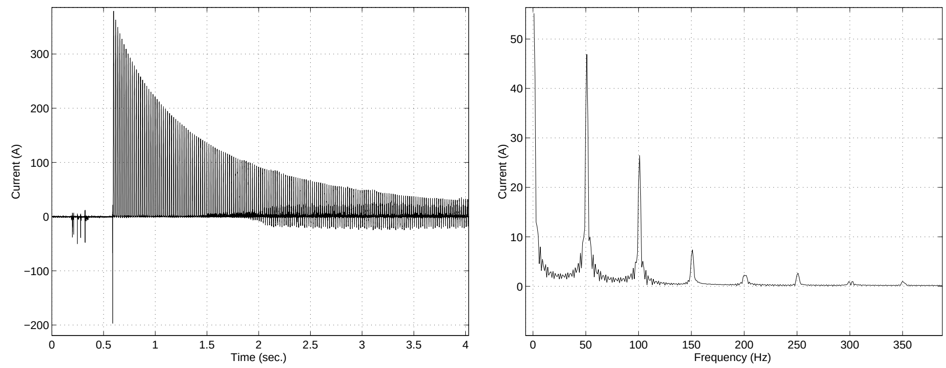

Figure 2 Real Signal. Left: Inrush current: time domain. Right: Inrush current: windowed spectral analysis

Definition 1 A family of functions $\{\boldsymbol{\psi}_{j,n}(t); (j,n) \in \boldsymbol{Z}, t \in \boldsymbol{R}\}$ in a Hilbert $\mathcal{H}$ space is called a frame of $\mathcal{H}$ if there are two positive constants $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\forall\ f(t) \in \mathcal{H}$ it holds:

$$\boldsymbol{A}\|f(t)\|^2 \leq \sum_{j,n} \|\langle f(t), \boldsymbol{\psi}_{j,n}(t)\rangle\|^2 \leq \boldsymbol{B}\|f(t)\|^2, \tag{4}$$

where $\langle \cdot, \cdot \rangle$ indicates the inner product and with $\|\cdot\|$ a norm.

In this framework, the drawback is that the optimal decomposition on a non-orthogonal basis is a NP-complete problem and one needs to stop the algorithm, for instance, with a threshold criterion at the stage i for the l_2 norm of the differential error. Now, the first question is which function to use as the basis function. Thus the problem consists of picking one that 'suits the application', in the sense that only a few terms will be needed. A suitable criterion already known in literature is to select the basis which, once a threshold level has been fixed, has the minimum number of elements in the selected frame. Now, having chosen the best family, how does one go about choosing the size of the frame subset? Finally, how is it possible to select the terms of the subset?

3.1 Choosing the Best Regressor Family: Smooth Trigonometric Wavelet Packets

A regressor family is a family of functions, orthogonal or non-orthogonal, with which it is possible to approximate a given signal.

The case presented in this paper has quasi-harmonic signals that change amplitude and phase over time. The latter aspect suggests the wavelet as the basis function. In Fig. 2, a measured signal in the time domain and its windowed Fourier transform are depicted. The data is very well concentrated around several frequencies, in this case it is multiples of the fundamental (50 Hz). The picture in Fig. 2 seems to suggest a function with a frequency window and time support. As shown in [7] a suitable family for this case is the smooth trigonometric wavelet packet. By choosing adjacent functions in particular, orthonormal bases are obtained. But if bases on different levels of the tree are considered as in Fig. 3, these do not form an orthonormal basis. The functions

$$\boldsymbol{S}_{i,k}(t) = \frac{2}{\sqrt{2T_i}}\mathcal{W}_i(t)\sin\left[(2k+1)\frac{\pi}{2T_i}(t-\alpha_i)\right] \tag{5}$$

form an orthonormal basis of $\mathcal{L}^2(\Re)$ subordinate to the partition $\mathcal{W}_i$. The collection of such bases forms a library of orthonormal bases[20]. One can form a library of orthonormal local

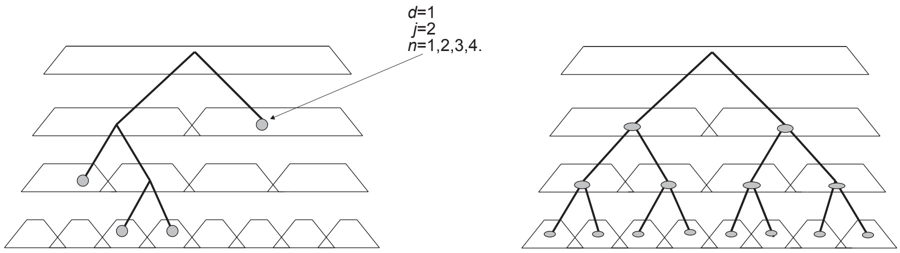

Figure 3 Organization of local intervals into a binary tree for smooth local trigonometric wavelets. Orthogonal basis (left) and frame (right)

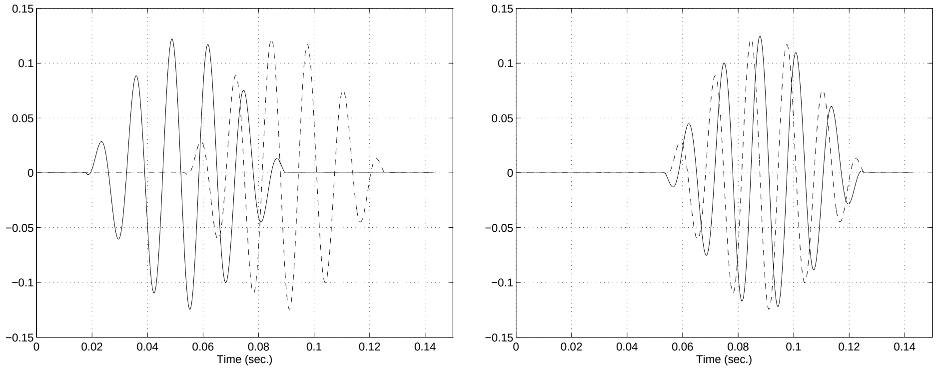

Figure 4 Left: Adjacent (orthogonal) cosine waveforms with smooth window $\boldsymbol{C}_{2,1}(t)$ and $\boldsymbol{C}_{2,2}(t)$. Right: Biorthogonal smooth local sine and cosine function $\boldsymbol{S}_{2,2}(t)$ and $\boldsymbol{C}_{2,2}(t)$

cosine bases:

$$\boldsymbol{C}_{i,k}(t) = \frac{2}{\sqrt{2\mathcal{T}_i}} \mathcal{W}_i(t) \cos\left[(2k+1)\frac{\pi}{2\mathcal{T}_i}(t-\alpha_i)\right], \tag{6}$$

where $\Re = \bigcup_{-\infty}^{\infty} \mathcal{I}_i$ is considered, where $\mathcal{I}_i = [\alpha_i, \alpha_{i+1})$ and $\alpha_i < \alpha_{i+1}$. Write $\mathcal{T}_i = \alpha_{i+1} - \alpha_i = |\mathcal{I}_i|$ and let $\mathcal{W}_i(t)$ be a window function supported in $[\alpha_i - \frac{\mathcal{T}_{i-1}}{2}, \alpha_{i+1} + \frac{\mathcal{T}_{i+1}}{2}]$ such that

$$\sum_{-\infty}^{\infty} \mathcal{W}_i^2(t) = 1 \tag{7}$$

and

$$\mathcal{W}_i^2(t) = 1 - \mathcal{W}_i^2(2\alpha_{i+1} - t), \qquad \text{for } t \text{ near } \alpha_{i+1}. \tag{8}$$

Definition 2 Let a library of wavelet packets be a collection of functions of the form

$$\boldsymbol{\psi}_{d,j,n}(t) = \boldsymbol{\psi}_j(2^d t - n), \tag{9}$$

where $(d, n) \in \boldsymbol{Z}$ and $j \in N$.

Here, the pyramidal packet is represented by the indices d, j, n, where d is the level of the tree (scaling parameter), j is a frequency cell (oscillation parameter) and n is a time cell

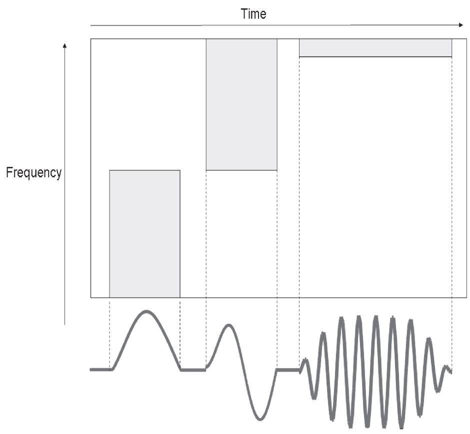

Figure 5 Time frequency cells in phase plane for trigonometric functions

(localization parameter), see Fig. 3. Another representation of the pyramidal tree is given in Fig. 6. The function $\boldsymbol{\psi}_{d,j,n}(t) = \boldsymbol{\psi}_j(2^d t - n)$ is roughly centered at $2^{-d}n$, and has support of size $\approx 2^{-d}$ and oscillates with frequency $\approx j$.

If we look at it in more detail we can see that if the signal consists of $d_0 = 2^{n_0}$ dyadic and equally spaced samples, the basis functions will be indexed by the triplet d, j, n: if d_0 is the total number of the samples then the corresponding samples related to the d level with relative desampling are $d_{0_d} = 2^d$ and

$$0 \leq d \leq n_0, \quad 0 \leq j < 2^{n_0 - d}, \quad 0 \leq n < 2^d. \tag{10}$$

The scale parameter d divides the number of decompositions of the original signal window into subwindows and the position index n numbers the adjacent windows. Thus the information cell is drawn over the horizontal (time) interval $I_n = [2^{n_0-d}n, 2^{n_0-d}(n+1)[$. In general, the local trigonometric bases, for instance the cosine basis, for the subspace over the time subinterval I_n consists of the function with the associated information cell alongside the frequency interval $I_j = [2^d j, 2^d(j+1)[$ on the vertical axis (frequency), see Fig. 5. The basis functions have nominal frequencies in $2^d(j + \frac{1}{2})$. Each subdivision halves the nominal window width and thus the resolution level. In particular the resolution level on the tree could be represented as a collection of rectangles:

$$[2^{n_0-d}n, 2^{n_0-d}(n+1)[\quad \times \quad [2^d j, 2^d(j+1)[. \tag{11}$$

Taking a basis with cells on different levels of the tree, a non-orthogonal basis (frames) is obtained. Taking basis elements on different levels of the tree, which cover the real axis $\Re$ one is considering superpositions of bases with different resolution frequency cells. In other words the orthogonality is lost.

4 Denoising and Harmonic Detection Algorithm

Our algorithm will work transversally on the wavelet packet tree without any restriction in order to use all the possible combinations of the bases, all the possible frames. A family regressor is selected, for instance the sine/cosine wavelets, the d, j, n parameterized family:

$$\left\{\mathcal{R}_c, \mathcal{R}_s\right\} = \left\{\boldsymbol{\psi}^c_{d,j,n}(t), \boldsymbol{\psi}^s_{d,j,n}(t); (d,n) \in \boldsymbol{Z}, j \in N, t \in \Re\right\}$$

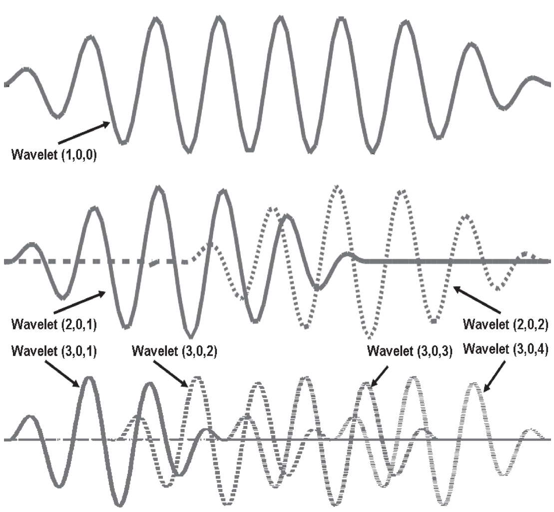

Figure 6 A structure of the wavelet tree

should contain a finite number of wavelets, as few as possible, so that the regressor selection procedure can be efficiently applied. In an approximating wavelet library not all the wavelet functions are useful. Normally only a small number of the coefficients are important and the other ones can be neglected. In general one can select a candidate library as follows:

$$\big\{\mathcal{R}_c,\ \mathcal{R}_s\big\} = \big\{\boldsymbol{\psi}^c_{d,j,n}(t), \boldsymbol{\psi}^s_{d,j,n}(t) : d,j,n \in \mathcal{I}_1 \cup \mathcal{I}_2 \cup \cdots \cup \mathcal{I}_K\big\} \tag{12}$$

with $K = 1, 2, \cdots, L$ and

$$\mathcal{I}_k = \big\{d, j, n : \|w^c_{d,j,n}(k)\|_p > \epsilon, \|w^s_{d,j,n}(k)\|_p > \epsilon\big\}. \tag{13}$$

where

$$w^c_{d,j,n}(k) = \frac{\langle \boldsymbol{Y}_k(t), \boldsymbol{\psi}^c_{d,j,n}(t)\rangle}{\|\boldsymbol{\psi}^c_{d,j,n}(t)\|^2_p} \quad \text{and} \quad w^s_{d,j,n}(k) = \frac{\langle \boldsymbol{Y}_k(t), \boldsymbol{\psi}^s_{d,j,n}(t)\rangle}{\|\boldsymbol{\psi}^s_{d,j,n}(t)\|^2_p}$$

with $\boldsymbol{Y}_k(t)$ is a set of testing signals, ϵ is a chosen small positive number, $\langle\cdot\,,\cdot\rangle$ and $\|\cdot\|_p$ represents its p induced norm. In this way 'empty' wavelets are eliminated from the wavelet frame. In other words, starting with a regular tree packet (library), select only those whose support fit our training data. This method is called wavelet shrinkage by some authors[21]. The algorithm seeks a set of the wavelet bases characterized by the indices d, j, n, which minimize and maximize the Entropy functions (16) and (17), respectively. Among these sets, the indices which maximize the Cross Entropy (18) are selected. Before describing the proposed algorithm in a mathematical way a summary description is given.

Once the cosine and sine frames have been built.

Step 0 The measured signal is charged.

Step 1 The measured signal is decomposed on all the wavelet functions of the tree.

Step 2 Calculate for the measured signal in sine frame the best basis representation ("coherent" part of the signal) and in cosine frame the worst basis representation ("incoherent" part of the signal). In this step at least two bases for each representation are calculated.

Step 3 Among the selected bases (at least 4 bases) select the couple which maximizes "the cross Entropy function".

Step 4 Through these two bases build "the coherent" and the "incoherent" part of the signal. Subtract these two parts from the analyzed signal. Iterate again till a stop criterion is not satisfied.

Let $\mathcal{R}_c = \left\{\boldsymbol{\psi}^c_{d,j,n}(t); (d,n) \in \boldsymbol{Z}, j \in N, t \in \Re\right\}$ and $\mathcal{R}_s = \left\{\boldsymbol{\psi}^s_{d,j,n}(t); (d,n) \in \boldsymbol{Z}, j \in N, t \in \Re\right\}$ be the truncated cosine and sine packet frames respectively as defined in [7]. The algorithm is developed as follows. Let $\boldsymbol{f}_{c(i)}(t)$ and $\boldsymbol{f}_{s(i)}(t)$ denote the estimate of the quasi–harmonic signal and noise respectively. Let us use function $Y_i(t)$, which represents the measured data. $\gamma_{c(i)}(t)$ and $\gamma_{s(i)}(t)$ represent the residual functions at the step i.

Algorithm

Step 0 (initialization): Initialize the iterative algorithm as follows

$$\begin{aligned}&i = 1,\\ &\boldsymbol{f}_{c(0)}(t) = \boldsymbol{f}_{s(0)}(t) = 0,\\ &\boldsymbol{\gamma}_{c(0)}(t) = \boldsymbol{\gamma}_{s(0)}(t) = \mathbf{Y}_0(t) = \boldsymbol{d}.\end{aligned}$$

Step 1 (computing weights): For each triple d, j, n within bounds (10), compute the weights $\boldsymbol{c}_{d,j,n}$, $\boldsymbol{s}_{d,j,n}$ of the signal's decomposition on the trigonometric wavelet library:

$$\begin{aligned}&\min_{(\boldsymbol{c}(.),\boldsymbol{s}(.))} \mathcal{J}_0(\boldsymbol{c}_{d,j,n}, \boldsymbol{s}_{d,j,n})\\ &= \min_{(\boldsymbol{c}(.),\boldsymbol{s}(.))} \left(\boldsymbol{Y}_{(i-1)}(t) - \sum_{d,j,n\in\mathcal{R}_c} \boldsymbol{c}_{d,j,n}\boldsymbol{\psi}^c_{d,j,n}(t) - \sum_{d,j,n\in\mathcal{R}_s} \boldsymbol{s}_{d,j,n}\boldsymbol{\psi}^s_{d,j,n}(t)\right)^2.\end{aligned}$$

This yields to

$$\boldsymbol{c}_{d,j,n} = \frac{\sum\limits_{d,j,n\in\mathcal{R}_c} \langle\gamma_{c(i-1)}(t), \boldsymbol{\psi}^c_{d,j,n}(t)\rangle}{\sum\limits_{d,j,n\in\mathcal{R}_c} \left(\boldsymbol{\psi}^c_{d,j,n}(t)\right)^2}, \tag{14}$$

$$\boldsymbol{s}_{d,j,n} = \frac{\sum\limits_{d,j,n\in\mathcal{R}_s} \langle\gamma_{s(i-1)}(t), \boldsymbol{\psi}^s_{d,j,n}(t)\rangle}{\sum\limits_{d,j,n\in\mathcal{R}_s} \left(\boldsymbol{\psi}^s_{d,j,n}(t)\right)^2}. \tag{15}$$

Step 2 (optimizing Entropy functions): Select $N_c \geq 2$ bases (also redundant) around the minimum of the Entropy function of the cosine library

$$\mathcal{V}_c = -\sum_{d,j,n} \left(\frac{\widehat{\mathcal{P}}\big(\gamma_{c(i-1)}(t)\big)}{\mathcal{P}\big(\gamma_{c(i-1)}(t)\big)}\right) \ln\left(\frac{\widehat{\mathcal{P}}\big(\gamma_{c(i-1)}(t)\big)}{\mathcal{P}\big(\gamma_{c(i-1)}(t)\big)}\right), \tag{16}$$

and select $N_s \geq 2$ bases (also redundant) around the maximum of the Entropy function of the sine library

$$\mathcal{V}_s = -\sum_{d,j,n} \left(\frac{\widehat{\mathcal{P}}\big(\gamma_{s(i-1)}(t)\big)}{\mathcal{P}\big(\gamma_{s(i-1)}(t)\big)}\right) \ln\left(\frac{\widehat{\mathcal{P}}\big(\gamma_{s(i-1)}(t)\big)}{\mathcal{P}\big(\gamma_{s(i-1)}(t)\big)}\right), \tag{17}$$

where

$$\mathcal{P}\big(\gamma_{c(i-1)}(t)\big) = \|\gamma_{c(i-1)}(t)\|^2, \quad \widehat{\mathcal{P}}\big(\gamma_{c(i-1)}(t)\big) = \Bigg\| \sum_{d,j,n\in\mathcal{R}_c} \boldsymbol{c}_{d,j,n} \boldsymbol{\psi}^c_{d,j,n}(t) \Bigg\|^2.$$

and $\mathcal{P}\big(\gamma_{s(i-1)}(t)\big)$ and $\widehat{\mathcal{P}}\big(\gamma_{s(i-1)}(t)\big)$ are defined analogously. Refer to these bases as $\{l_{c_{d,j,n}}\}$ with $d,j,n \in \{\mathcal{R}_s\}$ and $\{l_{s_{d,j,n}}\}$ with $d,j,n \in \{\mathcal{R}_c\}$, i.e.,

$$\{l_{c_{d,j,n}}\} = \arg\Big(\min_{\{\mathcal{R}_c\}}(\|\mathcal{V}_c\|)\Big) \quad \text{and} \quad \{l_{s_{d,j,n}}\} = \arg\Big(\max_{\{\mathcal{R}_s\}}(\|\mathcal{V}_s\|)\Big).$$

Step 3 (maximizing Cross Entropy): Among the $N_c \times N_s$ selected bases, choose the basis pair which maximizes the Cross Entropy function

$$\mathcal{V}_{\text{cross}} = \sum_{d,j,n\in\mathcal{R}_c} \sum_{d,j,n\in\mathcal{R}_s} \left(\frac{\widehat{\mathcal{P}}(\gamma_{c(i-1)}(t))}{\mathcal{P}(\gamma_{c(i-1)}(t))}\right) \ln \left(\frac{\frac{\widehat{\mathcal{P}}(\gamma_{c(i-1)}(t))}{\mathcal{P}(\gamma_{c(i-1)}(t))}}{\frac{\widehat{\mathcal{P}}(\gamma_{s(i-1)}(t))}{\mathcal{P}(\gamma_{s(i-1)}(t))}}\right). \tag{18}$$

Analytically

$$\{l_{c^*_{d,j,n}}, l_{s^*_{d,j,n}}\} = \arg\Big(\max_{\{l_{c_{d,j,n}}\}\times\{l_{s_{d,j,n}}\}} \big(\mathcal{V}_{(\text{cross})}\big)\Big).$$

Step 4 Reconstruct, by using these two bases, the coherent and the incoherent part of the signal. Update $\boldsymbol{f}_c(t)$, $\gamma_c(t)$, $\boldsymbol{f}_s(t)$ and $\gamma_s(t)$:

$$\boldsymbol{f}_{c_i}(t) = \boldsymbol{f}_{c_{(i-1)}}(t) + \sum_{l_c\in\mathcal{R}_c} \boldsymbol{c}_{l_{c_{d,j,n}}} \boldsymbol{\psi}^c_{l_{c_{d,j,n}}}(t); \tag{19}$$

$$\boldsymbol{f}_{s_i(t)} = \boldsymbol{f}_{s_{(i-1)}}(t) + \sum_{l_{s_{d,j,n}}\in\mathcal{R}_s} \boldsymbol{s}_{l_{s_{d,j,n}}} \boldsymbol{\psi}^s_{l_{s_{d,j,n}}}(t); \tag{20}$$

$$\boldsymbol{\gamma}_{ci}(t) = \boldsymbol{\gamma}_{c(i-1)}(t) - \sum_{l_{c_{d,j,n}}\in\mathcal{R}_c} \boldsymbol{c}_{l_{c_{d,j,n}}} \boldsymbol{\psi}^c_{l_{d,j,n}}(t); \tag{21}$$

$$\boldsymbol{\gamma}_{si}(t) = \boldsymbol{\gamma}_{s(i-1)}(t) - \sum_{l_{s_{d,j,n}}\in\mathcal{R}_s} \boldsymbol{s}_{l_{s_{d,j,n}}} \boldsymbol{\psi}^s_{ls_{d,j,n}}(t), \tag{22}$$

$$\boldsymbol{Y}_i(t) = \boldsymbol{Y}_{(i-1)}(t) - \boldsymbol{f}_{c_i}(t) - \boldsymbol{f}_{s_i}(t). \tag{23}$$

Step 5 Stop the algorithm if index (1), i.e. the l_2 norm of $\boldsymbol{d} - \boldsymbol{Y}_i$, is lower than a given threshold, otherwise increment index i and go to Step 1.

The obtained feature of the signal ($\boldsymbol{c}_{l_{c_{d,j,n}}}$ and $\boldsymbol{s}_{l_{s_{d,j,n}}}$) could be used for training neural networks.

Remark 3 The basic idea of the algorithm is to split the signal into two parts which correspond to the coherent and incoherent part of the signal relative to the selected bases. This is found by seeking subspaces characterized by the indexes d,j,n where the Entropy functions of sine and cosine waveforms have a minimum and maximum, respectively. The Cross Entropy function is used successively, in order to select the most discriminating subspaces.

Remark 4 Concerning the minimum of the index (1) it will be shown that the problem is convex, thus implying the convergence of the proposed algorithm. The basic problem consists of looking for a linear vector combination $\mathbf{c}_{d,j,n}$ and $\mathbf{s}_{d,j,n}$ which minimizes and maximizes Shannon Entropy functions of sine and cosine waveforms respectively. The search is done among the following truncated frames:

$$\mathcal{R}_c = \left\{\boldsymbol{\psi}^c_{d,j,n}(t); (d,n) \in \boldsymbol{Z}, j \in N, t \in \Re\right\}$$

and

$$\mathcal{R}_s = \left\{\boldsymbol{\psi}^s_{d,j,n}(t); (d,n) \in \boldsymbol{Z}, j \in N, t \in \Re\right\}.$$

In order to show the convergence of the algorithm the first step ($i = 1$) is considered. Let $\gamma_0 = \boldsymbol{Y}_0(t)$ and

$$\min_{(\boldsymbol{c}(.),\boldsymbol{s}(.))} \mathcal{J}_0 = \min_{(\boldsymbol{c}(.),\boldsymbol{s}(.))} \Big(\gamma_0 - \boldsymbol{c}_{d,j,n}\psi^c_{d,j,n}(t) - \boldsymbol{s}_{d,j,n}\psi^s_{d,j,n}(t)\Big)^{\mathrm{T}} \cdot \Big(\gamma_0 - \boldsymbol{c}_{d,j,n}\psi^c_{d,j,n}(t) - \boldsymbol{s}_{d,j,n}\psi^s_{d,j,n}(t)\Big) \tag{24}$$

is solved by

$$\boldsymbol{c}_{d,j,n} = \Big(\big(\psi^c_{d,j,n}(t)\big)^{\mathrm{T}}\psi^c_{d,j,n}(t)\Big)^{-1}\big(\psi^c_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_0 = \big(\psi^c_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_0$$

and

$$\boldsymbol{s}_{d,j,n} = \Big(\big(\psi^s_{d,j,n}(t)\big)^{\mathrm{T}}\psi^s_{d,j,n}(t)\Big)^{-1}\big(\psi^s_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_0 = \big(\psi^s_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_0,$$

being $\psi^c_{d,j,n}(t)$ and $\psi^s_{d,j,n}(t)$ such that

$$\big(\psi^c_{d,j,n}(t)\big)^{\mathrm{T}}\psi^c_{d,j,n}(t)^{-1} = I, \quad \big(\psi^s_{d,j,n}(t)\big)^{\mathrm{T}}\psi^s_{d,j,n}(t)^{-1} = I.$$

Index $\mathcal{J}_0$ can be written as:

$$\begin{aligned}\mathcal{J}_0 &= \Big(\gamma_0 - \psi^c_{\mathcal{B}}(t)^{\mathrm{T}}\gamma_0\psi^c_{\mathcal{B}}(t) - \psi^s_{\mathcal{B}}(t)^{\mathrm{T}}\gamma_0\psi^s_{\mathcal{B}}(t)\Big)^{\mathrm{T}} \\ &\quad \cdot \Big(\gamma_0 - \psi^c_{\mathcal{B}}(t)^{\mathrm{T}}\gamma_0\psi^c_{\mathcal{B}}(t) - \psi^s_{\mathcal{B}}(t)^{\mathrm{T}}\gamma_0\psi^s_{\mathcal{B}}(t)\Big) \\ &= \gamma_0^{\mathrm{T}}\gamma_0 + \Big(\psi^c_{\mathcal{B}}(t)^{\mathrm{T}}\gamma_0\Big)^2\psi^c_{\mathcal{B}}(t)^{\mathrm{T}}\psi^c_{\mathcal{B}}(t) - 2\Big(\psi^c_{\mathcal{B}}(t)^{\mathrm{T}}\gamma_0\Big)^2 \\ &\quad + \Big(\big(\psi^s_{\mathcal{B}}(t)\big)^{\mathrm{T}}\gamma_0\Big)^2\big(\psi^s_{\mathcal{B}}(t)\big)^{\mathrm{T}}\psi^s_{\mathcal{B}}(t) - 2\Big(\big(\psi^s_{\mathcal{B}}(t)\big)^{\mathrm{T}}\gamma_0\Big)^2.\end{aligned}$$

According to the transposition and biorthogonality properties, $\langle\psi^c_{d,j,n}(t), \psi^s_{d,j,n}(t)\rangle = 0$, $\forall\, d, j, n$, see [20], the minimization problem can be written as

$$\min_{(\boldsymbol{c}(.),\boldsymbol{s}(.))} \mathcal{J}_0 = \min_{(\boldsymbol{c}(.),\boldsymbol{s}(.))} \Big(\gamma_0^{\mathrm{T}}\gamma_0 - \Big(\big(\psi^c_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_0\Big)^2 - \Big(\big(\psi^s_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_0\Big)^2\Big). \tag{25}$$

Because of the biorthogonality, one can separately find the subspaces parameterized by d, j, n which optimize the Shannon entropy function. For the generic index i, because of the monotonicity of the Entropy function, the minimum in Step 2 corresponds to the minimum in

$$\mathcal{J}_i\big(\boldsymbol{c}_{d,j,n}\big) = \gamma_i^{\mathrm{T}}\gamma_i - \Big(\big(\psi^c_{d,j,n}(t)\big)^{\mathrm{T}}\gamma_i\Big)^2 \tag{26}$$

which corresponds to the minimum for the index (1). In [22] it was proved that $\gamma_i^{\mathrm{T}}\gamma_i$ is monotonically decreasing as i increases and this guarantees that the loop can be efficiently stopped by checking index (1).

Regarding the numerical aspects, the computational complexity is $\mathcal{O}(n(\log(n))^2)$ where n is the length of the considered signals. The algorithm is robust since it does not require any matrix inversion. The drawback of the proposed procedure is the low speed of convergence. Basically, this is due to the non-orthogonality of the bases.

5 Applications and Results Integrating Denoising Filter and Neural Networks

5.1 Validation

In order to validate the proposed algorithm, just an example is shown.

Fig. 7 shows the original signal and the reconstructed signal.

The following signal is considered:

$$Y_f = (400 + 70n(t))\sin(2\pi 70t) + 20,$$

where $n(t)$ is the white noise with variance equal to 1.

From Fig. 7 it is possible to see how constant 20 is included in the coherent part of the signal.

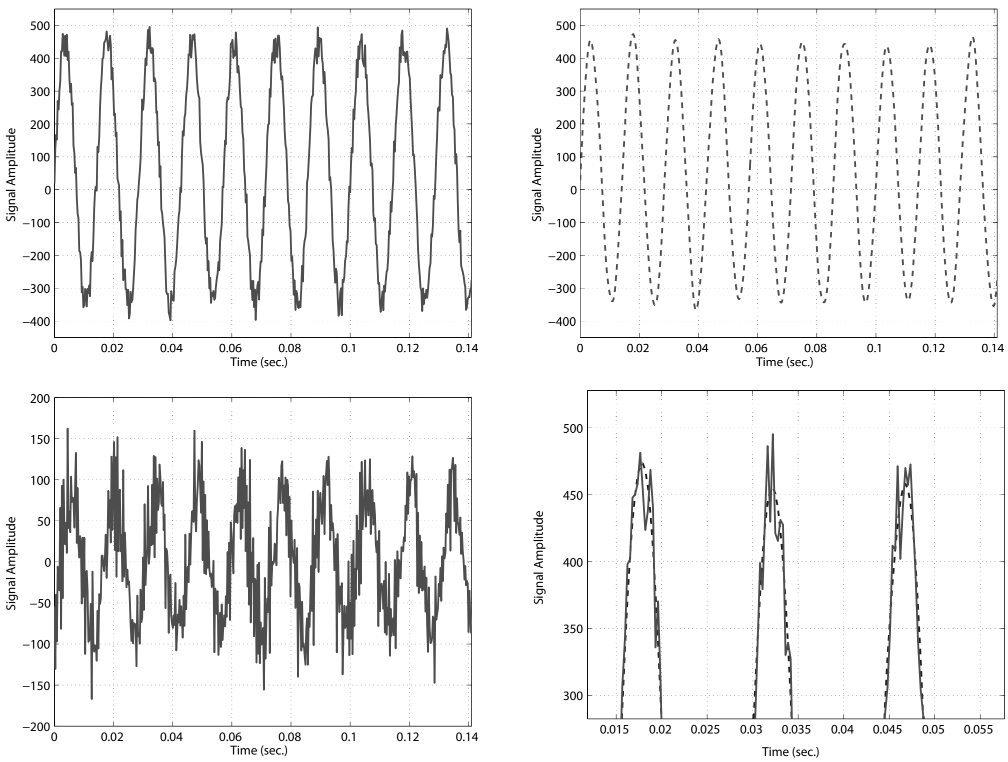

Figure 7 Left on the top: Generated signal. Right on the top: compressed and reconstructed signal. Left on the bottom: compressed and reconstructed noise. Right on the bottom: Details of the generated signal and its compression and reconstruction.

To validate the algorithm other kind of signal have been used but for sake of brevity just this two is shown. Real measured data from a locomotive transformer of a railway vehicle are processed through the proposed algorithm in order to detect the good features of the inrush current. The algorithm is well suited to such applications since the inrush currents are quasi-harmonic signals. The signal slowly changes amplitude and phase, making them comparatively narrow-band. This feature justifies the choice of the trigonometric bases in order to perform a coherent and incoherent expansion. The measured dyadic signal belongs to $\Re^{512}$, after the thresholding phase, a dyadic vector belonging to the space $\Re^{8}$ is obtained. This allows the consideration of only 8 frequencies $[0, 50, 100, 150, 200, 250, 300, 350]$ Hz for every time cell. The selected wavelet packet tree has three levels and considering the Nyquist frequency equal to 3.5 MHz and the length of the basis equal to 512 samples, one obtains a resolution around 3.5 Hz, 7 Hz and 14 Hz respectively.‡ The accuracy of the estimate depends on the setting of the threshold parameter and the nature of the disturbance. The usefulness of the technique consists of getting relevant compressed features even though the signal is corrupted by relevant noise. These features could be used in training wavelet networks in order to recognize dangerous inrush currents as they occur in a rail vehicles. In fact the problem connected with the training of the wavelet networks consists of obtaining the good features of the signal which are normally corrupted by electrical interference. Fig. 8 shows the results regarding the denoising problem. In the meantime it is easy to see the effect of the compression which introduces a new noise, the so called compression noise.

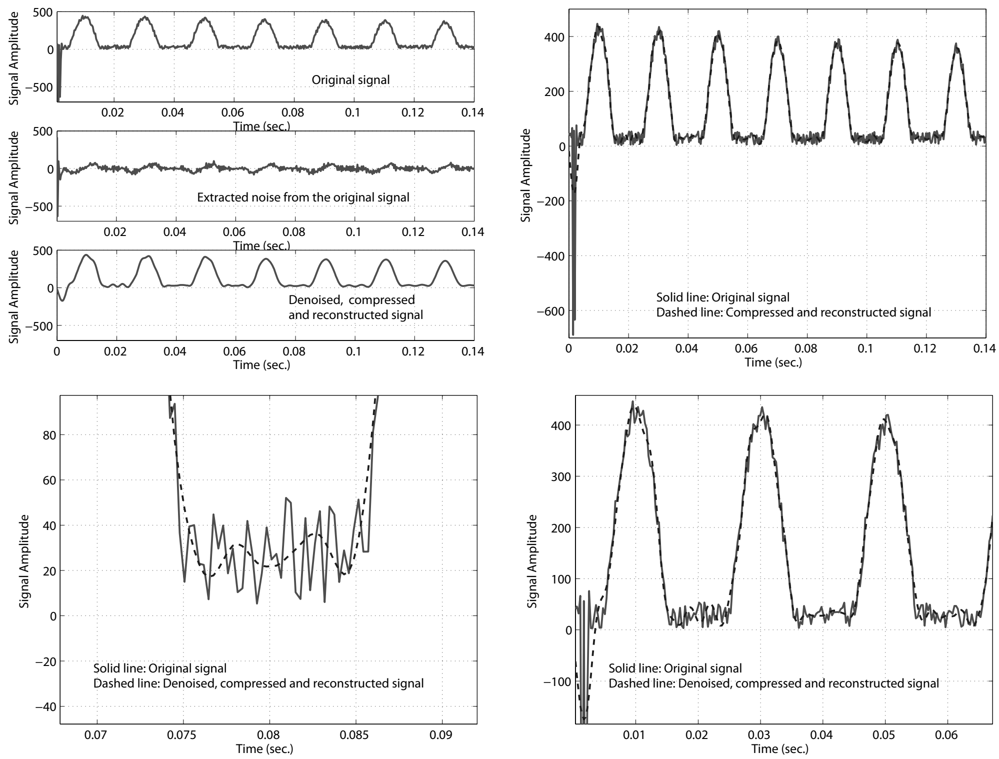

Figure 8 Original signal, denoised signal, compressed and reconstructed signal and details

‡Recall that the resolution $R = \frac{N_f 2^d}{N}$, where N is the length of the basis, d the level of the tree and N_f the Nyquist frequency.

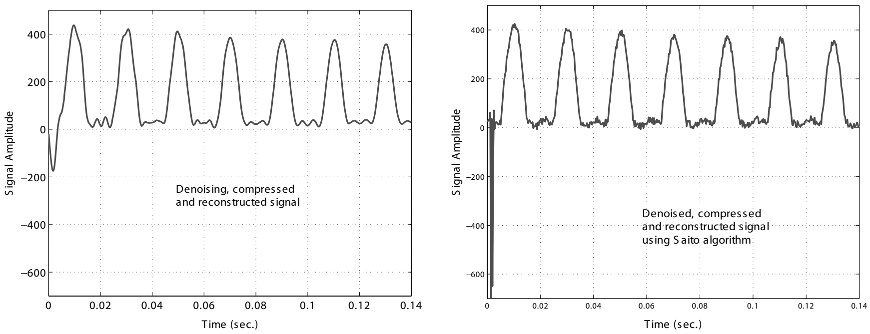

Figure 9 Left: Inrush current obtained through the proposed denoising and compression algorithm. Right: Inrush current obtained through the Saito denoising and compression algorithm.

5.2 Inrush Current

In Fig. 11 the structure of the network is represented. The sine pattern is devoted to the detection of the dangerous inrush currents. The cosine pattern is devoted to the detection to the non dangerous currents and the noise. In Table the results of the test are reported. In the network, sine and cosine layers are presented to recognize and differentiate dangerous transients from non-dangerous ones. By using the proposed algorithm as a filter for the neural network already presented in [8] the Inrush current is recognized without any error. In [8] a network structure is adaptively performed through subspaces of sine and cosine wavelet packet trees organized in three levels. The adaptive subspaces were characterized by sine and cosine vectors described as following. Once the prototype function is built by the training, this is compressed with a cosine/sine basis in vectors with 8 coefficients for every time cell, corresponding to the $[0; 50; 100; 150; 200; 250; 300; 350]$ Hz. These vectors represent our compressed prototype (coordinate system). The preliminary testing simulations, consisting of evaluating the Euclidean distance between the projected fresh signal and the coordinate basis vectors, show correct identification percentages of 95 % in 60 ms. The classical methods used in rail vehicle control, combine FFT (Fast Fourier Transform), DFT (Digital Fourier Transform) and a band pass plus threshold criterion and normally need more than 80 ms to recognize the inrush. Their equipment is very intricate and cannot distinguish dangerous transients from non-dangerous ones, see Fig. 10. In Fig. 11 the new scheme for identification of dangerous transients is schematically reported, where it is possible to see its elementary structure. The coordinated systems reported in Fig. 11 could be built through a neural network as reported in Fig. 1, see [8]. By using the Saito algorithm, as proposed in [18], the results show a low percentage of error (around 90 %) but with time significantly shorter (around 40 ms). This is due to the velocity connected with the orthogonality of the filter technique. In table I some results are summarized. The orthogonal algorithm is not able to make good distinction between noise and signal because of its fixed structure. The advantage consists of the introduction of a low disturbance compression. Fig. 9 shows, in the case of inrush current, the results of the two algorithms for the denoising problem.

5.3 Other Possible Applications

In Fig. 12 and 15 another industrial application is presented. It is a temperature sensor in a paper machine application. In [23] a similar problem is presented, where the authors proposed an algorithm to detect noise and outliers. Nevertheless the approach presented in

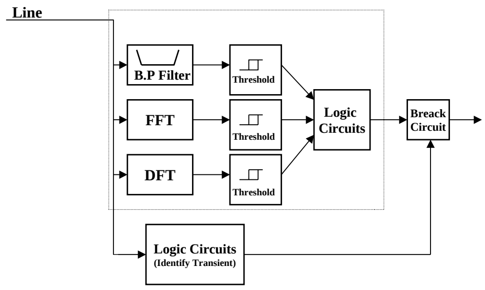

Figure 10 Actual Apparatus for Identification of Dangerous Transients

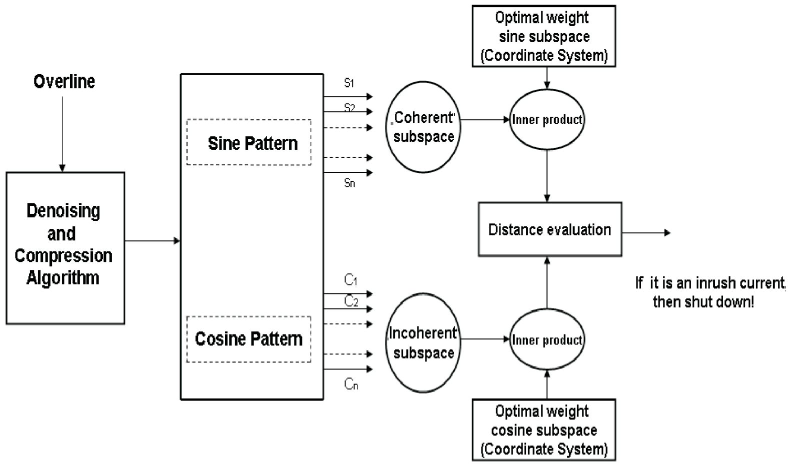

Figure 11 New Apparatus for Identification of Dangerous Transients

this paper offers the possibility to classify and increase the percentage of correct detection and increase the speed of detection time. With the technique presented in this paper, the outliers are performed as the incoherent part of the signal. The structure of the pattern recognition is identical to the case presented previously. This problem, according to the definitions, is performed through just one class of signals and its complement (outliers). The sine pattern is devoted to the detection of the desired signal. The cosine pattern is devoted to the detection of the outliers. The signal under consideration has a large bandwidth. The presented simulations are performed by vectors with the length of no more than 32 components, 16 components for every time cell (8 for sine and 8 for cosine decomposition). To be more exact, for every time cell, $[0; 50; 100; 150; 200; 250; 300; 350; 400; 450; 500; 550; 600; 650; 700; 750]$ Hz. In the preliminary simulation 50 signals are considered as training signals and other 50 signals as fresh testing ones. These signals were split into two complementary classes: 50 signals with outliers and the remaining 50 without them. The identified percentage of outliers is, for the kind of signal considered, around 98%. In Fig. 12, 13, and 14 some cases are shown. It is interesting to notice that there is a hyper selectivity due to "on-line operation" of the learning machine, see Fig. 15. The hyper selectivity decrees if a shorter window is adopted. In this case the percentage of the correctly identified outliers is lower and it does not increase through increased training. The inner product between the coordinate systems and the fresh data vectors resulting from the compression through the algorithm needs mostly only one time-frequency cell (on-line detection) to recognize the outliers. In these cases, (outliers' detection) Saito's approach is

Table 1 Inrush Current Test

Training wavelet network by using non-orthogonal denoising filter			
Class Tested	Correctly identified	Time of Identification	Incorrect
Training Data	100%	less than 20 ms	0%
Fresh Data	around 95%	less than 60 ms	around 5%

Training wavelet network without denoising filter			
Class Tested	Correctly identified	Time of Identification	Incorrect
Training Data	100%	less than 20 ms	0%
Fresh Data	around 80%	less than 40 ms	around 20%

Training wavelet network using Saito denoising filter			
Class Tested	Correctly identified	Time of Identification	Incorrect
Training Data	100%	less than 20 ms	0%
Fresh Data	around 90%	less than 50 ms	around 10%

Results by using actual apparatus for identification inrush current			
Class Tested	Correctly identified	Time of Identification	Incorrect
Fresh Data	around 90%	less than 200 ms	around 10%

preferred because it presents similar results but the structure is much simpler.

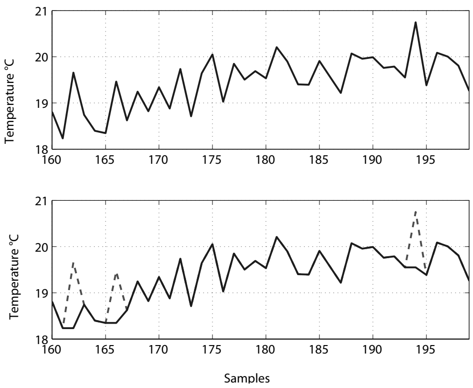

Figure 12 On the top: non-filtered signal. On the bottom the dashed line is the signal filtered through the learning machine

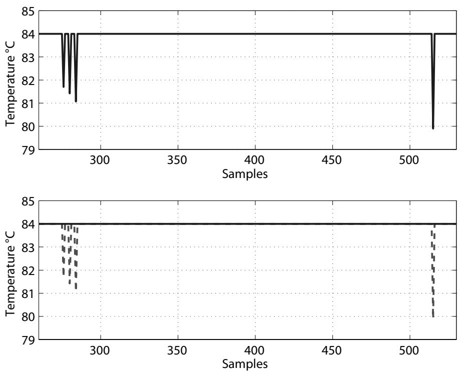

Figure 13 On the top: non- filtered signal. On the bottom the dashed line is the signal filtered through the learning machine

6 Conclusions

An algorithm for simultaneously suppressing any additional white Gaussian noise component and compressing the signal component in a data set is described. The aim of the paper consists of extracting relevant features of quasi–harmonic signals from noisy measurements with interferences. In railway applications such features can be used for training neural networks to recognize dangerous inrush currents from non-dangerous ones. In chemical processes the neural network can be used to detect a fault and an outlier from sensor measurements. The

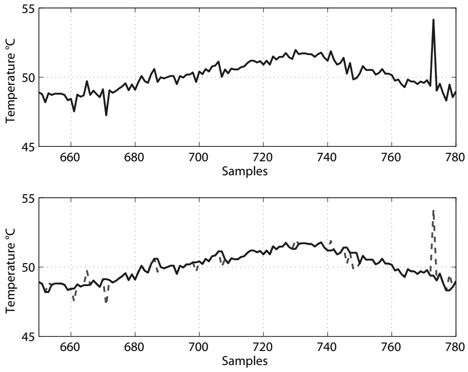

Figure 14 On the top: non-filtered signal. On the bottom the dashed line is the filtered signal through the learning machine

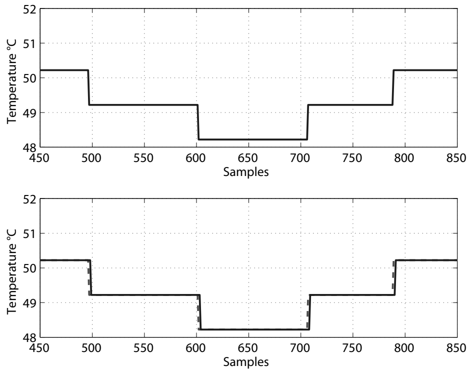

Figure 15 On the top: non-filtered signal. On the bottom the dashed line is the signal filtered through the learning machine

selected bases consist of wavelets, more precisely they consist of wavelet packets where the basis functions are local trigonometric bases. Cosine and sine bases with their biorthogonality allow an efficient coordinate system to perform properly. The bases are selected during every step by maximizing the cross entropy function which highlights the difference between the noise and the desired signal. Two industrial applications are reported to show the usefulness of the algorithm, the first one is related to inrush current detection and the second one is concerns an outliers problem.

References

[1] P. Terwiesch, Industrial modeling and control challanges related to electrical rail vehicles, in *Proc. European Control Conference ECC'*99, 1999.

[2] S. Liu, An Adaptive Kalman filter for dynamic estimation of harmonic signals, in *Proceedings-IEEE* 8*th International Conference on Harmonics and Quality of Power*, Athens 1998.

[3] B. Perunicić, M. Mallini, Z. Wang, and Y. Liu, Power quality disturbance detection and classification using wavelets and artificial neural networks, in *Proc.* 8*th IEEE International Conference on Harmonics and Quality of Power*, Athens 1998.

[4] P. Terwiesch, S. Menth, and S. Schmidt, Analysis of Transients in Electrical Railway Networks Using Wavelets, *IEEE Trans. on Industrial Electronics*, 1998, **45**(6): 955–959.

[5] P. Mercorelli, M. Rode, and P. Terwiesch, A Wavelet Packet Algorithm for Online Detection of Pantograph Vibrations, in *Proc.* 9*th IFAC Int. Symp. on Control in System Transportation 2000*, Braunschweig, 2000.

[6] P. Mercorelli, M. Rode, and P. Terwiesch, System and Methodology for Dominant Frequency Detection by Using a Set of Trigonometric Wavelet Functions, *Patent N* 7759 *in Patentamt ABB Corporate Research Mannheim. Code in the German Paten Office* 10 25 89 21. 6, 2001.

[7] P. Mercorelli and P. Terwiesch, A Black Box Identification in Harmonic Domain, *VDE European Transactions on Electrical Power*, 2003, **13**(1): 29–40.

[8] P. Terwiesch and P. Mercorelli, A Local Feature Extraction Using Biorthogonal Bases for Classification of Embedded Classes of Signals, in *Proc.* 14*th Int. Symp. on Mathematical theory of networks and systems* (MTNS2000), Perpignan, France, 2000.

[9] R. R. Coifman and M. V. Wickerhauser, Entropy based algorithm for best basis selection, *IEEE Trans. Inform. Theory*, 1992, **32**: 712–718.

[10] N. Saito and Ronald R. Coifman, Local discriminant bases, in *Mathematical Imaging: Wavelet Applications in Signal and Image Processing II, Proc. SPIE*, 1994, **2303**.

[11] Q. Zhang, Using Wavelet Network in Nonparametric Estimation, *IEEE Transaction on Neural Networks*, 1997, **8**(2): 227–236.

[12] J. M. Shapiro, Image coding using the embedded zerotree wavelet algorithm, in *Proc. SPIE Conf. on Mathematical Imaging: Wavelet Applications in signal and Image Processing* (ed. by A. F. Laine), 1992, **2034**: 180–193.

[13] R. A. De Vore, B. Jawerth, and B. J. Lucier, Image compression through wavelet transform coding, *IEEE Trans. Inform. Theory*, 1992, **38**(2): 719–746.

[14] M. V. Wickerhauser, High-resolution still picture compression, *Digital Signal Processing: A Review Journal*, 1992, **2**(4): 204–226.

[15] D. L. Donoho, Wavelet shrinkage and W. V. D.: a 10-minute tour, in *Progress in Wavelet Analisis and Applications*, Editions Frontieres, B. P. 33, 91192 Gif-sur-Yvette Cedex France, 1993, 109–128.

[16] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Preprint, Dept. of Statistics, Stanford University, Stanford, CA, Jun., 1993.

[17] S. Mallat and Z. Zhang, Matching Pursuit with Time-Frequency Dictionaries, *IEEE Trans. on Signal Processing*, 1993.

[18] N. Saito, Simultaneous Noise Suppression and Signal Compression using a Library of Orthonormal Bases and the Minimum Description Length Criterion, *Wavelets in Geophysics* (ed. by E. Foufoula-Georgiou and P. Kumar), Academic Press, 1994.

[19] I. Daubechies, *Ten Lectures on Wavelets*, Publisher Society for Industrial and Applied Mathematics, Philadelphia (Pennsylvania), 1995.

[20] P. Auscher, G. Weiss, and M. V. Wickerhauser, Wavelet-A Tutorial in Theory and Applications, in *Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets* (ed. by C-K Chui), Academic Press, Boston, 1992, 237–256.

[21] A. Benveniste, A. Juditscky, B. Delyon, Q. Zhang, and P. G. Glorenee, Wavelets in Identification, in *Proc. SYSID'*94 10*th IFAC Symp. Syst. Identification*, Copenhagen, 1994.

[22] S. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, Technical Report 619, New York University, Computer Science Departiment, 1993.

[23] P. Mercorelli and A. Frick, Noise Level Estimation Using Haar Wavelet Packet Trees for Sensor Robust Outlier Detection, Selected paper in *IEE The* 2006 *International Conference on Computational Science and its Applications (ICCSA* 2006 *Glasgow)*, Series: Lecture Note in Computer Sciences, Springer-Verlag publishing, 2006.