Optimum Rate-Distortion Dictionary Selection for Compression of Atomic Decompositions of Electric Disturbance Signals

Michel P. Tcheou, *Student Member, IEEE*, Lisandro Lovisolo, Eduardo A. B. da Silva, *Senior Member, IEEE*, Marco A. M. Rodrigues, *Member, IEEE*, and Paulo S. R. Diniz, *Fellow, IEEE*

***Abstract*—In this letter, we address rate-distortion-optimum compression of signals from electric power system disturbances, using atomic decompositions. Usually, such optimization is obtained assuming a single dictionary and consists of finding the best compromise between the quantization of the coefficients in the atomic decomposition and its number of terms. Here, several parameterized dictionaries are used instead. This allows the selection of the dictionary leading to the best rate-distortion (R-D) compromise. Distinct dictionaries correspond to different quantizers for the parameters of the atoms. Side information must be transmitted in order to indicate the dictionary employed. The R-D performance in this case depends on a complex interplay between the quantizers of the parameters of the atoms and the coefficient quantizers. Using a training stage, we select a reduced set of parameter and coefficient quantizers that give near-optimum R-D performance. Simulation results show that the proposed scheme indeed achieves near-optimum R-D performance with low computational complexity in the coding stage.**

***Index Terms*—Atomic decompositions, dictionaries, electric disturbance signals, rate-distortion optimization.**

I. Introduction

Atomic decompositions represent signals using linear combinations of functions (atoms) drawn from a dictionary. They are important tools for the compression of several signal sources [1]–[4]. When based on a redundant dictionary, atomic decompositions can provide good adaptive signal approximations. The approximation is adaptive since the atoms are selected from the dictionary according to the signal being decomposed. The use of highly redundant dictionaries enables efficient decompositions of a wide range of signals. Several methods have been used to obtain these representations [5], such as the matching pursuit (MP) algorithm [6].

A signal $\mathbf{x}$ can be approximated by an atomic decomposition as

$$\hat{\mathbf{x}} = \sum_{n=0}^{M-1} \alpha_n \mathbf{g}_{\gamma(n)}. \quad (1)$$

The atoms $\mathbf{g}_{\gamma(n)}$ are selected from a redundant dictionary D, being indexed by the mapping $\gamma(n)$. If C_D is the dictionary cardinality, this mapping is defined as $\gamma : \mathbb{Z}^+ \rightarrow \{1, \ldots, C_D\}$. In compression applications, one encodes the number of terms M, the coefficients α_n, and atom indexes $\gamma(n)$. The optimum R-D tradeoff is achieved by finding a compromise between the number of expansion elements and the quantization of each coefficient [7]. Instead of using a single dictionary, consider a set of redundant dictionaries described as $\mathcal{D} = \{D_i\}_{i=1,\ldots,I}$, where I is the number of dictionaries included in $\mathcal{D}$. This framework is illustrated in Fig. 1. In this case, the dictionary used must be indicated to the decoder as side information. The optimum R-D performance corresponds to the tradeoff among the bits spent on side information, atom indexes, and coefficients leading to the minimum distortion. The solution to this tradeoff usually involves high computational demands, which increase with the number of distinct dictionaries in $\mathcal{D}$.

In this letter, we address atomic decompositions of electric power system disturbance signals using parameterized dictionaries [4]. The atoms consist of pieces of damped sinusoids that can be defined by a set of five parameters (frequency, damping factor, phase, start, and ending times). For coding, the parameters must be quantized. Referring to Fig. 1, each quantizer for the parameter space defines one dictionary D_i. One possible way to obtain the optimum R-D tradeoff would be to find the optimum coefficient quantizer and number of terms M for each dictionary D_i (parameter quantizers) and choosing the one with the best performance. However, some combinations of coefficient quantizers and dictionaries lead to poor performance, irrespective of the signal being compressed. We avoid such combinations by employing a reduced set of dictionaries belonging to the convex hull of operational R-D characteristics of signals from a training set. The results show that the proposed strategy allows to achieve near-optimal R-D performance with comparatively low computational complexity.

II. Compression of Electric Disturbance Signals Using Parameterized Dictionaries

Large electric power systems consist of complex interconnected grids where several players perform distinct functions such as generation, transmission, and distribution of electrical power. For each player, it is crucial that any disturbance be monitored, not only to meet regulatory requirements but also to identify its causes. With the wide availability of dedicated monitoring devices known as digital fault recorders (DFRs), an increasing number of events can be acquired. Therefore, specific compression methods for reducing the file sizes of the DFR data are needed. Typical DFR data are formed by voltage and current

Manuscript received May 12, 2006; revised June 10, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiang-Gen Xia.

M. P. Tcheou and M. A. M. Rodrigues are the Centro de Pesquisas de Energia Elétrica-CEPEL, Rio de Janeiro, RJ, 21941-590, Brazil (e-mail: pompeu@cepel.br; mamr@cepel.br).

L. Lovisolo is with the Universidade Estadual do Rio de Janeiro, Rio de Janeiro, RJ, 20550-900, Brazil (e-mail: lisandro@uerj.br).

E. A. B. da Silva and P. S. R. Diniz are with the Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21945-970, Brazil (e-mail: eduardo@lps.ufrj.br; diniz@lps.ufrj.br).

Digital Object Identifier 10.1109/LSP.2006.882117

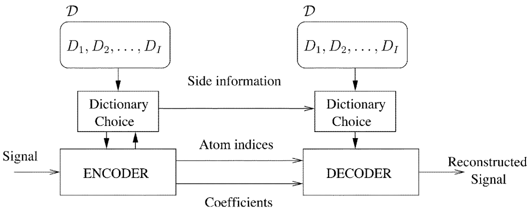

Fig. 1. Compression framework using a set $\mathcal{D}$ of redundant dictionaries.

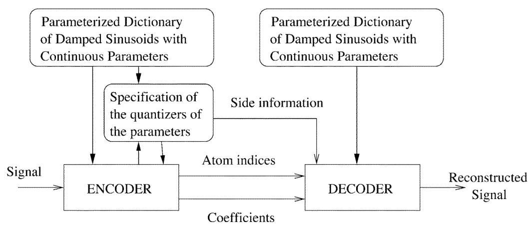

Fig. 2. Compression framework using a set of parameterized dictionaries.

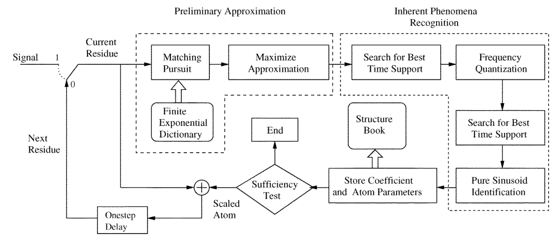

Fig. 3. Atomic decomposition algorithm. In the first iteration, the switch is in position 1 and in the remaining iterations stays in position 0.

waveforms from a transmission line. It is usually related with a fault and contains some cycles before and after the fault instant.

Roughly, we can consider that electric power systems are basically formed by sources, loads, and transmission lines, i.e., RLC circuits, whose transient behavior is modeled by damped sinusoids. There are also discontinuities in these signals due to switching. Following these premises, a disturbance signal $x(t)$ can be approximated via

$$x(t) = \sum_{i=0}^{M-1} \alpha_i e^{-\rho_i\left(t-t_{s_i}\right)} \cos(2\pi k_i F t + \phi_i) \times \left[u\left(t-t_{s_i}\right) - u\left(t-t_{e_i}\right)\right] \quad (2)$$

where M is the number of expansion elements, F is the fundamental frequency (50/60 Hz), and each element is represented by a 6-tuple $(\alpha_i, k_i, \rho_i, \phi_i, t_{s_i}, t_{e_i})$, where α_i is the amplitude, k_i is an integer, ρ_i is the decaying factor, ϕ_i is the phase, t_{s_i} and t_{e_i} are the starting and ending times, and $u(\cdot)$ corresponds to the unit step function. Based on the signal model given by (2), in [4], a compression method of disturbance signals that employs parameterized dictionaries of damped sinusoids, resembling Fig. 1, is proposed. First, one performs the atomic decomposition using a parameterized dictionary of damped sinusoids with continuous parameters through the MP algorithm [4], [6]. Then, the parameters of the atoms are quantized along with the coefficients. Each quantizer for the atom parameters corresponds to a different dictionary D_i (see Fig. 2). The dictionary choice (parameter quantizer used) must be informed to the decoder as side information.

The MP method is used in the encoder of Fig. 2. It performs successive approximations of signals iteratively employing the dictionary elements. At the first iteration, the MP algorithm chooses the atom with the highest correlation with the signal. The chosen atom is then scaled and subtracted from the signal obtaining a residue. The process is repeated (using the previously calculated residue) until the residue energy becomes sufficiently small or until another stopping criterion is met [5], [6].

The parameterized damped sinusoidal atom $\mathbf{g}_\gamma$ is given by

$$g_\gamma(k) = K_\gamma g(k) \cos(\xi k + \phi)\left[u(k - m^s) - u(k - m^e)\right]$$
$$k = 0, \ldots, N_s - 1$$

$$g(k) = \begin{cases} 1, & \text{if } \rho = 0 \quad \text{Pure sinusoid } (\xi \neq 0), \\ & \quad \text{dc or unit impulse } (\xi = 0) \\ e^{-\rho(k-m^s)}, & \text{if } \rho > 0 \\ & \quad \text{Decreasing exponential} \\ e^{\rho(m^e-k)}, & \text{if } \rho < 0 \\ & \quad \text{Increasing exponential} \end{cases} \quad (3)$$

where K_γ is set so that $\|\mathbf{g}_\gamma\| = 1$ and N_s is the signal length. Furthermore, the atom in (3) is defined by the 5-tuple $\gamma = (\rho, \xi, \phi, m^s, m^e)$ in which ρ is the decaying factor, ξ denotes the frequency, ϕ denotes the phase, and m^s and m^e are the starting and ending samples.

In [4], some heuristics are introduced inside the MP loop in order to achieve physically interpretable representations with respect to disturbance signal phenomena. Fig. 3 shows the block diagram of the resulting algorithm. First, it carries out a preliminary approximation of the current residue considering its entire time support, which means that $m^s = 0$ and $m^e = N_s - 1$. The algorithm searches the most correlated atom with respect to the residue in a finite exponential dictionary with presampled parameter space. Then, one looks for the γ (value of the atom parameters) that maximizes the match between the atom and the current residue using a Newton-like optimization.

After the preliminary approximation, the algorithm performs inherent phenomena recognition by first reducing the time support of the atom determined by the variables m^s and m^e. It reduces the region of support sample by sample and verifies whether the inner product between the atom with new time support and the current residue increases. It then quantizes the atom frequency to a multiple of the fundamental frequency of the power system and repeats the time support search. After that, an attempt is made to identify pure sinusoids within the support region by using a heuristic based on a similarity metric. In other words, the algorithm decides between damped sinusoidal atom and pure sinusoid, choosing the one that achieves the smallest error per sample with respect to the residue. The algorithm stops when the approximation achieved is good enough. Otherwise, it scales and subtracts the atom from the current residue and produces a new residue to be approximated in the following iteration. At the end of the decomposition, we obtain the signal approximation in (1) represented by the sequence of pairs $(\alpha_n, \gamma(n))$, $n = 0, \ldots, M-1$, where $\gamma(n) = (\rho_n, \xi_n, \phi_n, m_n^s, m_n^e)$ [see (2) and (3)]. More details can be found in [4].

For compression, the coefficients and atoms parameters are quantized after the decomposition. Here α_n and each parameter of $\gamma(n)$ are quantized using uniform scalar quantizers as [8]

$$x_q = I_x \times \Delta_{q(x)}, \quad \text{where } I_x = \left\lfloor \frac{x + \frac{\Delta_{q(x)}}{2}}{\Delta_{q(x)}} \right\rfloor \quad (4)$$

where x is the parameter, x_q denotes its quantized version, $\Delta_{q(x)}$ denotes the quantization step, and I_x is the index assigned to x.

The quantizers of each parameter are defined by a dynamic range and a number of levels, constrained here to powers of two, such that they are indexed using an integer number of bits. The starting and ending samples, m_n^s and m_n^e, are quantized with $\log_2(N_s)$ bits. In addition, the frequency ξ_n is quantized with $\log_2((F_s/2)/F)$ bits, where F_s is the sampling frequency, and F denotes the fundamental frequency in which the power system operates. For the coefficients α_n, the decaying factors ρ_n, and the phases ϕ_n, the dynamic range is defined by their respective maximum and minimum values, that is, $\Delta_{q(x)} = (x_{\max} - x_{\min})/(2^{b_x} - 1)$, where b_x is the number of bits allocated to x. For the coefficient quantizer, $\alpha_{min} = 0$. The dynamic ranges and the number of bits allocated to α_n, ρ_n, and ϕ_n, the signal length, and the fundamental and sampling frequencies are encoded as side information.

The quantization of the parameters and the coefficients give rise to the following signal approximation:

$$\tilde{\mathbf{x}} = \sum_{n=0}^{M-1} Q_\alpha\{\alpha_n\}\mathbf{g}_{Q_i\{\gamma(n)\}} \tag{5}$$

where $Q_\alpha\{\cdot\}$ is the quantizer of the coefficients, and $Q_i\{\cdot\}$ denotes the quantizer of the parameters. Note that each $Q_i\{\cdot\}$ corresponds to a dictionary D_i. Thus, for each bit allocation among the parameters, a distinct dictionary is used, that is, different atoms will compose the signal approximation. The dictionary D_i used is defined by the mapping $Q_i\{\cdot\}$, and $\tilde{\mathbf{x}}$ corresponds to a weighted sum of the elements selected from D_i. The weights of the atoms in $\tilde{\mathbf{x}}$ depend on $Q_\alpha\{\cdot\}$, since an atom can appear more than once in a certain decomposition, (5) implies that some atoms have more impact on the resulting signal reconstruction than others. The optimum R-D solution is provided by the quantizers $Q_\alpha\{\cdot\}$ and $Q_i\{\cdot\}$ that lead to the minimum distortion for a given rate.

Usually, compression systems based on MP in the related literature achieve signal compression by retaining a certain number of terms M and quantizing the coefficients [7]. The compression framework presented substantially differs from these. Here, since we use a continuous parameter dictionary, compression is achieved by also quantizing the parameters of atoms. This compression scheme is equivalent to the use of multiple dictionaries followed by the selection of one of them for coding a given signal (see Fig. 1).

III. R-D Optimum Dictionary Selection

The purpose of R-D optimization is to achieve the best signal reproduction for a desired compression target [9]. In the framework at hand, one has to find a compromise between the number of atoms in the signal representation, the quantization of the coefficients, and the choice of the dictionary $D_i \in \mathcal{D}$ that is defined by the quantizers of the atom parameters.

First, define the number of bits for a given atom as

$$r = r_\alpha + r_\xi + r_\rho + r_\phi + r_{m^s} + r_{m^e} \tag{6}$$

where $r_\alpha, r_\xi, r_\rho, r_\phi, r_{m^s}$, and r_{m^e} denote the amount of bits allocated to $\alpha, \xi, \rho, \phi, m^s$, and m^e, respectively. The total number of bits spent in the signal compression, except for side information, will be $r \times M$, where M is the number of terms in the signal reconstruction [see (5)]. The quantities r_{m^s} and r_{m^e} are defined by the signal length, while r_ξ is defined by the sampling and the fundamental frequencies, i.e., they do not change in $\mathcal{D}$. Therefore, the total distortion is expressed just as a function of the number of bits spent on the coefficients, the decaying factors and the phases, resulting in the mean squared value (MSE)

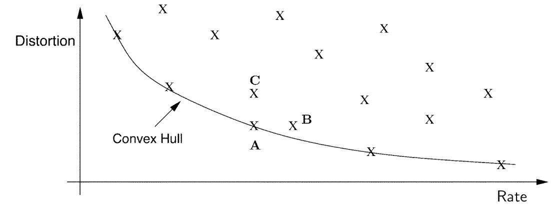

Fig. 4. Convex hull comprising the R-D optimal points.

$$d = \frac{1}{N_s}\|\mathbf{x} - \tilde{\mathbf{x}}\|^2 = f(r_\alpha, r_\rho, r_\phi) \tag{7}$$

where N_s is the signal length.

Consider the quantizer defined by (4) and the individual wordlengths contained in the 6-tuple

$$\mathbf{b}_k = (r_\alpha, r_\xi, r_\rho, r_\phi, r_{m^s}, r_{m^e}) \in \mathcal{B} \tag{8}$$

where $\mathcal{B}$ is defined by the ranges of r_α, r_ρ, and r_ϕ (r_ξ, r_{m^s}, and r_{m^e} are fixed, depending only on the original signal). Note that each $\mathbf{b}_k$ defines a coefficient quantizer and a quantizer for the parameters of the atoms. The parameter quantizer corresponds to a choice of dictionary D_i [see (5)]. In order to accomplish the best R-D tradeoff, one should search for the $\mathbf{b}_k$ that minimizes the total distortion inserted by the encoding process given a bit-budget r_{budget}. The solution is obtained by solving the following optimization problem [9]:

$$\min_{\mathbf{b}_k \in \mathcal{B}} \{d\}, \quad \text{subject to} \quad r_{\text{budget}} = Mr. \tag{9}$$

The classical solution for this problem is based on the Lagrangian optimization [9], which corresponds to minimizing a cost function of the form $J = d + \lambda r_{\text{budget}}$, where $\lambda \geq 0$ denotes the Lagrangian multiplier. As there is no closed form for d with respect to $(r_\alpha, r_\xi, r_\rho, r_\phi, r_{m^s}, r_{m^e})$, we used the operational R-D curves [9] described in the sequel.

For each $\mathbf{b}_k = (r_\alpha, r_\xi, r_\rho, r_\phi, r_{m^s}, r_{m^e})$, and for a given signal, we compute the R-D pair (r_k, d_k), yielding an R-D plot as illustrated in Fig. 4. The rate r_k is measured in bits/sample, and the distortion d_k corresponds to the MSE [see (7)]. The operational curve is obtained by connecting the points belonging to the convex hull of the region defined by the R-D pairs generated for each $\mathbf{b}_k \in \mathcal{B}$. Note in Fig. 4 that, for instance, the point B is certainly worse than the point A, because it presents identical distortion but has larger rate. Likewise, the point C is also worse than the point A, because it presents a larger distortion for the same rate. Thus, we select the point belonging to the convex hull that for a desired compression rate gives the dictionary and coefficient quantizer that yields minimum distortion.

The set of optimum R-D points differs from signal to signal, that is, one $\mathbf{b}_k$ found to be optimum for one signal may not be optimum for another. Thereby, we propose a strategy to diminish the size of $\mathcal{B}$ by using only the selected $\mathbf{b}_k$ providing an R-D optimum dictionary and coefficient quantizer for at least one signal. This reduced set of $\mathbf{b}_k$ is capable of providing near optimum R-D performance with low computational complexity.

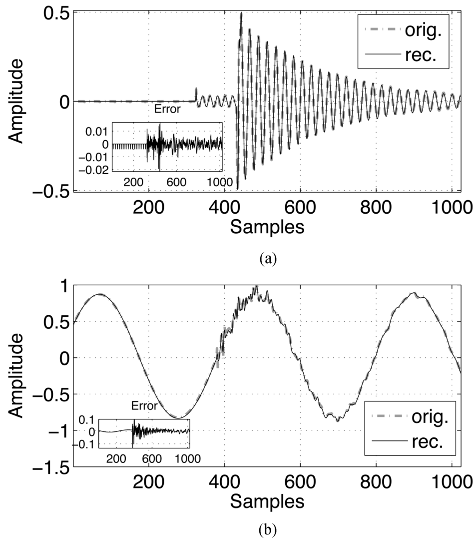

Fig. 5. Disturbance signals, their reconstructed versions, and errors at 0.95 and 0.578 bits/sample, respectively. (a) Signal R_1. (b) Signal R_2.

IV. Experimental Results

In this section, we verify the performance of the proposed R-D optimization scheme. The reduced set of $\mathbf{b}_k$ used in this optimization scheme was obtained from a set $\mathcal{B}$ defined by varying r_α from 3 to 16 bits and r_ρ and r_ϕ from 1 to 12 bits. This way, $\mathcal{B}$ has 2016 elements. Using $\mathcal{B}$, we obtained the operational R-D curves for a training set consisting of 29 disturbance signals acquired from the Brazilian power system. This training set is characterized by common phenomena in electric power system's disturbance signals [10]. The amount of signals in this set has been verified to be appropriate to generate a reduced set of $\mathbf{b}_k$ that is capable of providing near-optimum R-D performance as verified next. It was observed that, from the complete set of 2016 possible $\mathbf{b}_k$ used to obtain the training set operational R-D curves, only 292 $\mathbf{b}_k$ were actually optimal for at least one signal, yielding a seven-fold reduction in coding complexity.

The values of r_α, r_ρ, and r_ϕ, the fundamental and sampling frequencies, the signal length, as well as the dynamic ranges of the parameters and coefficients are encoded in a 149-bit header.

Fig. 5 illustrates two disturbance signals R_1 and R_2 [4]. They are originally represented using 16 bits per sample and do not belong to the training set employed in the R-D optimization. In [4], the quantizers were chosen disregarding the best R-D compromise. For example, for signal R_1 in [4] using a rate of 1.035 bits/sample, the signal was reconstructed with a signal-to-noise ratio ($\text{SNR} = 10\log_{10}(\|\mathbf{x}\|^2/\|\mathbf{x}-\tilde{\mathbf{x}}\|^2)$ dB) of 28.08 dB, whereas the R-D optimization provided a quantizer that codes signal R_1 at 0.95 bits/sample with 31.65 dB in SNR. For R_2, a quantizer defined by $r_\alpha = 6$, $r_\rho = 6$, and $r_\phi = 6$ was employed in [4], yielding an SNR of 31.12 dB at 0.584 bits/sample. The proposed R-D optimization found a better quantizer for R_2 at a similar bit-rate; this quantizer is given by $r_\alpha = 6$, $r_\rho = 3$, and $r_\phi = 4$, resulting in a rate equal to 0.578 bits/sample allied to an SNR equal to 31.39 dB (see plots in Fig. 6). Fig. 5 shows also the reconstructed versions of the compressed R_1 and R_2 at 0.95 and 0.578 bits/sample, respectively, and their reconstruction errors.

One can verify the effectiveness of the reduced set used in R-D optimization by observing Fig. 6. This figure presents the operational R-D curves of R_1 and R_2 using both the reduced and the complete set of $\mathbf{b}_k$. Note that, for both signals, the operational curves obtained by using the reduced set present good approximation to the curves obtained by using the complete set. This shows the effectiveness of the training stage that enabled the finding of a set of dictionaries (specified by $\mathbf{b}_k$) with good performance when applied to electric disturbance signals. This allows us to implement a compression scheme with low computational complexity that is also capable of reaching near-optimal R-D performance.

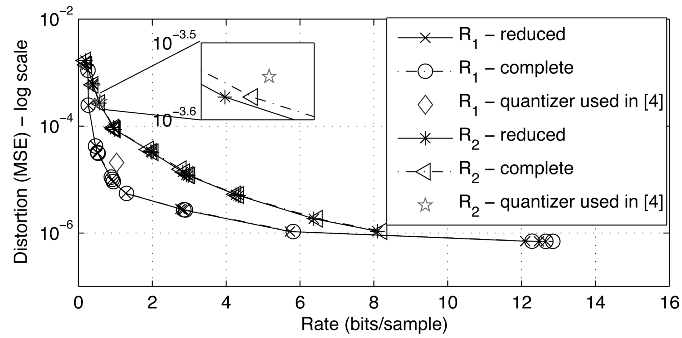

Fig. 6. Operational curves achieved by using the reduced and complete sets of quantizers for signals R_1 and R_2.

V. Conclusion

This letter proposes a framework for compression of electric disturbance signals using atomic decompositions with parameterized dictionaries. The dictionaries are defined by the quantizers applied to the parameters of the atoms. In this case, the R-D optimization is solved by finding the coefficient quantizer and the quantizers of the atom parameters that lead to the best R-D compromise. We obtain the R-D optimum dictionary and coefficient quantizer through the construction of R-D operational curves. In order to decrease the computational complexity of the R-D optimization, we used a training set to select a reduced set of good parameter quantizers. Simulation results have shown that the proposed scheme achieves near-optimum R-D performance with low computational complexity.

References

[1] O. K. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor, "Video compression using matching pursuits," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 123–143, Feb. 1999.

[2] M. M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal model based on recursive filter banks," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1890–1902, Jul. 1999.

[3] K. Skretting, K. Engan, and J. H. Husoy, "ECG compression using signal dependent frames and matching pursuit," in *Proc. IEEE ICASSP*, Mar. 2005, vol. 4, pp. 585–588.

[4] L. Lovisolo, E. A. B. da Silva, M. A. M. Rodrigues, and P. S. R. Diniz, "Efficient coherent adaptive representations of monitored electric signals in power systems using damped sinusoids," *IEEE Trans. Signal Processing.*, vol. 53, no. 10, pp. 3831–3846, Oct. 2005.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.

[6] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[7] P. Frossard, P. Vandergheynst, R. M. Figueras, I. Ventura, and M. Kunt, "A posteriori quantization of progressive matching pursuit streams," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 525–535, Feb. 2004.

[8] K. Sayood, *Introduction to Data Compression*, 2nd ed. San Francisco, CA: Morgan Kaufman, 2000.

[9] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[10] W. Xu, "Component modeling issues for power quality assessment," *IEEE Power Eng. Rev.*, vol. 21, no. 11, pp. 12–15, Nov. 17, 2001.