Stochastic Time-Frequency Dictionaries for Matching Pursuit

Piotr J. Durka, Dobiesław Ircha, Katarzyna J. Blinowska

IEEE Transactions on Signal Processing, Vol. 49, No. 3, pp. 507-510, March 2001.

Abstract— **Analyzing large amounts of sleep electroencephalogram (EEG) data by means of the matching pursuit (MP) algorithm, we encountered a statistical bias of the decomposition, resulting from the structure of the applied dictionary. As a solution we propose stochastic dictionaries, where the parameters of the dictionary's waveforms are randomized before each decomposition. The MP algorithm was modified for this purpose and tuned for maximum time-frequency resolution. Examples of applications of the new method include parametrization of EEG structures and time-frequency representation of signals with changing frequency.**

I. Introduction

THE ELECTROENCEPHALOGRAM (EEG), which is recorded from electrodes placed on the scalp, reflects the electrical activity of the brain. The signal is extremely complex and reveals both rhythmical and transient features, hence time-frequency methods are often considered an optimal choice for its analysis. Matching pursuit (MP, [1]) offers some extra advantages: high resolution, local adaptivity to transient structures, and compatibility of its time-frequency parameterization with definitions of EEG structures used in the traditional, visual analysis [2] [3] [4]. In spite of that, its widespread application is limited by the algorithm's complexity. This paper describes one of the basic problems encountered in practical applications of MP to large amounts of data: statistical bias of the parametrization. As a solution, we propose a modified version of the algorithm—MP with stochastic dictionaries, where the parameters of the dictionary's waveforms are randomized before each decomposition by drawing their values from continuous ranges.

Sections I-A and I-B review some basic ideas behind matching pursuit and the Gabor dictionary, as proposed in [1]. Section II introduces stochastic dictionaries and describes implementation of the corresponding MP algorithm. Section III illustrates the bias-free parametrization of EEG structures (III-A) and gives an example of another possible application of the proposed idea: time-frequency representation of structures of changing frequency (III-B).

Authors work at Laboratory of Medical Physics, Warsaw University, ul. Hoża 69, 00-681 Warszawa, Poland, http://brain.fuw.edu.pl, email: [durka rircha kjbli]@fuw.edu.pl. This work was supported by Polish State Committee for Scientific Research.

A. Matching pursuit

Given a set of functions (dictionary) $D = \{g_1, g_2, \ldots, g_n\}$ such that $||g_i|| = 1$, we can define an optimal M-approximation as an expansion, minimizing the error ϵ of an approximation of signal $f(t)$ by M waveforms:

$$\epsilon = \left\| f(t) - \sum_{i=1}^{M} w_i g_{\gamma_i}(t) \right\| \tag{1}$$

where $\{\gamma_i\}_{i=1..M}$ represents the indices of the chosen functions g_{γ_i}. Finding such an optimal approximation is an NP-hard problem [5]. A suboptimal expansion can be found by means of an iterative procedure, such as the matching pursuit algorithm (MP).

In the first step of MP, the waveform g_{γ_0} which best matches the signal $f(t)$ is chosen. In each of the consecutive steps, the waveform g_{γ_n} is matched to the signal $R^n f$, which is the residual left after subtracting results of previous iterations:

$$\begin{cases} R^0 f = f \\ R^n f = < R^n f, g_{\gamma_n} > g_{\gamma_n} + R^{n+1} f \\ g_{\gamma_n} = \arg\max_{g_{\gamma_i} \in D} | < R^n f, g_{\gamma_i} > | \end{cases} \tag{2}$$

Orthogonality of $R^{n+1} f$ and g_{γ_n} in each step implies energy conservation:

$$||f||^2 = \sum_{n=0}^{m-1} | < R^n f,\ g_{\gamma_n} > |^2 + ||R^m f||^2 \tag{3}$$

For a complete dictionary the procedure converges to f:

$$f = \sum_{n=0}^{\infty} < R^n f,\ g_{\gamma_n} > g_{\gamma_n} \tag{4}$$

From this equation we can derive a time-frequency distribution of the signal's energy, that is free of cross-terms, by adding Wigner distributions of selected functions:

$$Ef(t, \omega) = \sum_{n=0}^{M} | < R^n f,\ g_{\gamma_n} > |^2 \ W g_{\gamma_n}(t, \omega) \tag{5}$$

B. Discrete Gabor dictionary

Gabor functions (sine-modulated Gaussian functions) provide optimal joint time-frequency localization. A real Gabor function can be expressed as:

$$g_\gamma(t) = K(\gamma) e^{-\pi \left(\frac{t-u}{s}\right)^2} \sin\left(2\pi \frac{\omega}{N}(t-u) + \phi\right) \tag{6}$$

N is the size of the signal for which the dictionary is constructed, $K(\gamma)$ is such that $||g_\gamma|| = 1$. $\gamma = \{u, \omega, s, \phi\}$ denotes parameters of the dictionary's functions (time-frequency atoms). The length of signal (N points) suggests ranges for the parameters of Gabor functions which can be reasonably fitted to such an epoch. However, within these ranges no particular sampling is *a priori* defined, and we are confronted with a three-dimensional[1] continuous space; this results in an infinite dictionary size. Therefore, in practical implementations, we use subsets of the possible dictionary functions.

In the dictionary implemented originally by Mallat and Zhang in [1], the parameters of the atoms are chosen from dyadic sequences of integers. Their sampling is governed by an extra parameter—octave j (integer). Scale s, which corresponds to an atom's width in time, is derived from the dyadic sequence $s = 2^j, 0 \leq j \leq L$ (signal size $N = 2^L$). Parameters u and ω, which correspond to an atom's position in time and frequency, are sampled for each octave j with interval $s = 2^j$, or, if oversampling by l is introduced, with interval 2^{j-l}.

II. Matching Pursuit with stochastic dictionaries

In forming the dictionary used for MP decomposition, using any *fixed* scheme of subsampling the parameter space introduces statistical bias in the resulting parametrization. This bias becomes apparent only when statistics comes into play, like in parameterization of large amounts of data—examples are given in the next section. Unbiased MP decompositions can be obtained by an analogue of Monte Carlo methods.

We propose MP with stochastic dictionaries, where the parameters of a dictionary's atoms are randomized before each decomposition. A stochastic dictionary D is constructed for a given signal length N and chosen "resolutions"[2] in time, frequency and scale ($\Delta t, \Delta\omega$ and Δs). The space of parameters t, ω and s is thereby divided into $\frac{\Pi}{\Delta\omega}\frac{N^2}{\Delta t \Delta s}$ bricks of size Δt by $\Delta\omega$ by Δs each, where $\omega \in (0, \pi)$. In each of those bricks, one time-frequency atom is chosen by drawing its parameters from flat distributions within the given ranges of continuous parameters.

Ater the first iteration, the dictionary D is reduced to a subset D_α, which is constructed by choosing an arbitrary percentage of atoms having the largest correlations with the original signal. In each iteration, parameters of the atom g_{γ_i} chosen from D_α (or from D in the first iteration) are further optimized by a complete search in a dense dictionary D_{γ_i} constructed from parameters in the neighborhood of γ_i.

[1] The phase ϕ is usually optimized in each step of the MP procedure separately for each fitted atom.

[2] The resolution of the matching pursuit is hard to define in general, since the procedure is nonlinear and signal-dependent. It should be related to the distance between neighboring dictionary waveforms available for decomposition. In the described procedure, this distance does not exceed twice the value of the corresponding parameter (Δt, $\Delta\omega$ or Δs).

Numerical implementation uses the formula for fast correlation updates given in [1] and calculation of some of the products in the FFT domain (for long atoms). In spite of that, the computational complexity of this aggregate procedure was empirically approximated as $\mathcal{O}\left(N \text{size}(D)\right)$ per iteration, where N and size(D) stand for sizes of the signal and the dictionary, correspondingly.

We must stress at this point that the whole idea was *not* driven by a quest for improving the speed/compression ratio. We are interested in attaining possibly exact parametrization of certain signal structures, free from bias and with a constant time-frequency resolution, primarily for research purposes. Routine applications, like *e.g.* parametrization of selected EEG structures for clinical purposes, must be built upon a solid framework, where speed improvements at the cost of introducing artifacts, such as the one presented in the next section, are out of the question. Therefore, optimization of this implementation was limited by the basic requirements: stability and uniform, predictable resolution. The latter led to a choice of uniform (before randomization) grids of parameters t, ω and s, contrary to the generally preferred structured dictionaries [1][6].

III. Results

A. MP parametrization of EEG structures

One of the first applications of MP in EEG analysis was detection and parametrization of sleep spindles—structures present in sleep EEG recordings [3][4]. The definition given in a classical reference [7] states: "*The presence of sleep spindle should not be defined unless it is at least 0.5 sec. duration, i.e. one should be able to count 6 or 7 distinct waves within the half-second period. (...) The term should be used only to describe activity between 12 and 14 cps.*" Although intended for standardization of visual detection, the above definition is almost directly translatable into the time-frequency parameters of MP atoms fitted to analyzed EEG. Since the shape of sleep spindles is quite similar to Gabor functions, we can basically assume a one-to-one relationship between sleep spindles and atoms fitted to EEG, conforming to the above definition. This yields an automatic parametrization of structures conforming to selected time-frequency criteria. Details of this procedure, as well as examples of parametrization of other EEG structures, are given *e.g.* in [4] and [8].

The middle and lower plots of Figure 1 present histograms of frequencies of sleep spindles as detected using the above procedure in an overnight EEG recording. MP decomposition was performed on consecutive 20-*sec* epochs of EEG from sleep stage II (above 3 hours), 12-bit ADC, and sampling frequency 128 *Hz*. The histogram in the middle plot is based on MP decomposition in a dyadic dictionary (section I-B) . Its most striking feature is a regular, repeating pattern superimposed on the overall trend of concentration spindles around 13 Hz. This regular structure is an artifact of the analysis method.

The time-frequency centers of the atoms from a dyadic dictionary form a regular lattice. The upper plot of Figure

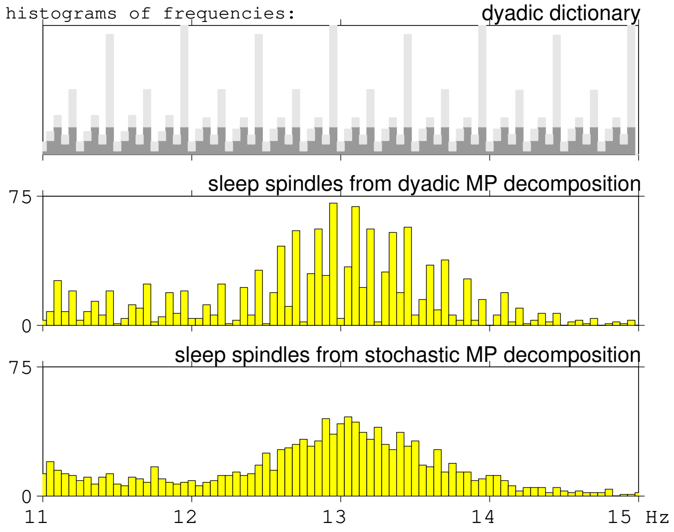

Fig. 1. Histograms of frequencies of sleep spindles from an overnight EEG recording, selected from MP parametrization in a dyadic (middle plot, section I-B) dictionary and stochastic Gabor dictionaries (lower plot, section II). The upper plot is a histogram of frequencies of atoms available for decomposition in the dyadic dictionary (for the frequencies of interest). The dark bars represent scales 64-256 chosen for sleep spindles, light gray is used for the rest of scales.

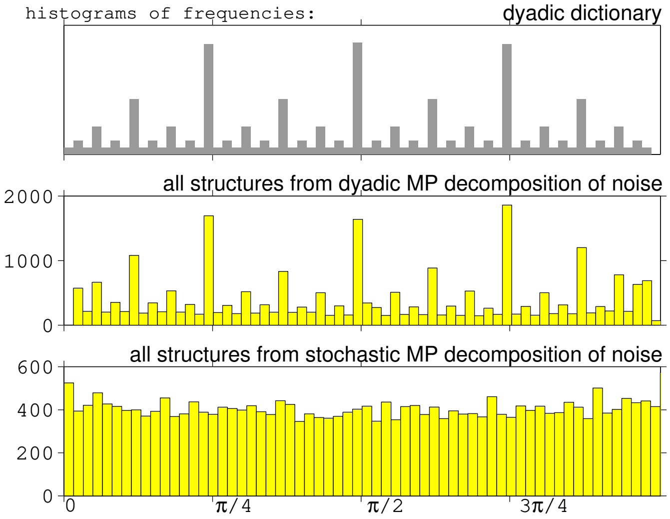

Fig. 2. Illustration of the effect from Figure 1 on simulated data. 200 realizations of 128-points white noise process were subjected to MP decomposition in dyadic (middle plot) and stochastic (bottom) dictionaries. The middle and lower plots present histograms of frequencies of all the structures fitted by MP. The upper plot is a histogram of frequencies of all the time-frequency atoms present in the dyadic dictionary used for the decomposition presented in the middle plot.

1 presents a histogram of frequencies of all the atoms from a dyadic dictionary, within the frequency range of interest. Its pattern can be recognized in the middle plot. A higher density of atoms available for decomposition at certain frequencies "attracts" results of the nonlinear MP procedure, and more waveforms are fitted at these frequencies.

The histogram in the lower plot of Figure 1 is constructed from a statistically unbiased MP representation, which was obtained by decomposition of the same EEG data as in the middle plot using stochastic dictionaries. Their size was equal to the size of the dyadic dictionary used to compute the middle plot's data (1.3×10^6 waveforms). We observe the absence of the artifactual maxima, conforming to physiological expectations.

The analogous effect is illustrated in Figure 2, where frequencies of atoms fitted to white noise reveal the structure of the dyadic dictionary (middle plot) or expected flat characteristics when obtained by MP with stochastic dictionaries (bottom).

B. Averaging time-frequency representations of one signal in several stochastic dictionaries

The proposed idea can be also applied to a single data epoch. Let's consider a signal simulated as

$$s(t) = \begin{cases} \sin\left(0.625\pi t \sin(0.002\pi t)\right), & t = 0 \dots 299 \\ \sin\left(0.007(t-300)^2\right), & t = 300 \dots 512 \end{cases} \tag{7}$$

The upper plot of figure 3 presents the result of decomposition of this signal over a large dictionary (7.5×10^5 Gabor functions). In spite of the high resolution of this decomposition, the changing frequency is represented by a series of structures since all the dictionary's functions have constant frequency. The middle plot of Figure 3 shows an average of 50 time-frequency representations constructed from decompositions over different realizations of small (1.5×10^4) stochastic dictionaries. Their size was optimized for this particular signal, and the number of averaged decompositions was chosen to make the computational costs of both representations equal (compare section II). The plot in the middle panel corresponds better to Equation 7. However, it is constructed from 50 times more waveforms than the upper plot, so the underlying parametrization is not compact.

IV. Conclusion

The statistical properties of matching pursuit parametrization in structured dictionaries result from complicated interactions between the properties of the decomposed signal and the structure of the applied dictionary. Bias-free parametrization can be obtained by hereby proposed MP with stochastic dictionaries. Our experience indicates that this approach yields a high-resolution, robust and reliable procedure. However, lacking *a priori* knowledge of the dictionary's structure makes numerical optimization difficult, in contrast to MP in structured dictionaries.

An implementation of the algorithm described in section II is available from `http://brain.fuw.edu.pl/~mp`.

Acknowledgements

We thank Prof. Waldemar Szelenberger, Dr. Michał Skalski, Dr. Wojciech Androsiuk and Dr. Szymon Niemcewicz from Warsaw Medical School for the EEG recordings and consultations. Thanks to Prof. Piotr Franaszczuk for consulting on an early manuscript.

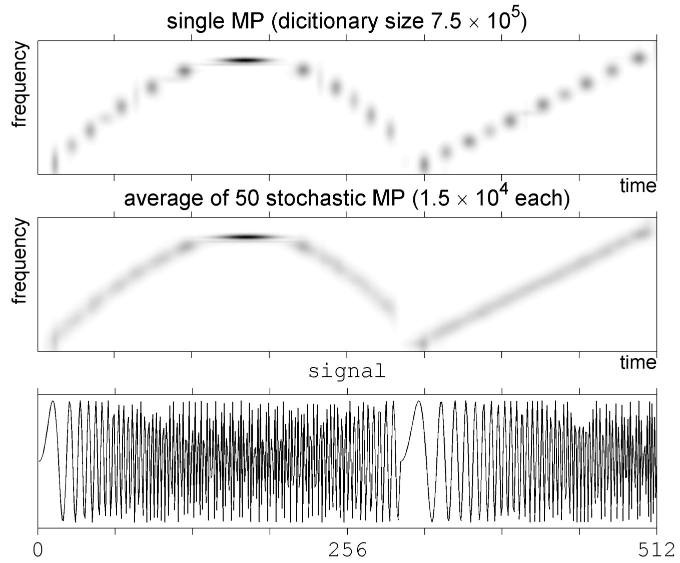

Fig. 3. Energy density ($Ef(t, \omega)$, eq. (5), proportional to shades of gray) of a simulated signal (bottom plot), calculated from single MP decomposition over a dictionary containing 7.5×10^5 waveforms (top) and averaged over 50 decompositions in different realizations of stochastic dictionaries, containing 1.5×10^4 atoms each (middle plot).

References

[1] Stéphane Mallat and Zhifeng Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, Dec 1993.

[2] Ernst Niedermayer and F. Lopes Da Silva, *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, Williams & Wilkins, fourth edition, 1999, p. 1154.

[3] P. J. Durka and K. J. Blinowska, "Analysis of EEG transients by means of matching pursuit," *Annals of Biomedical Engineering*, vol. 23, pp. 608–611, 1995.

[4] Piotr Jerzy Durka, *Time-frequency analyses of EEG*, Ph.D. thesis, Warsaw University, Jun 1996, ftp://brain.fuw.edu.pl/pub/DissertationPJDurka.ps.Z.

[5] Geoff Davis, *Adaptive Nonlinear Approximations*, Ph.D. thesis, New York University, 1994, ftp://cs.nyu.edu/pub/wave/report/DissertationGDavis.ps.Z.

[6] Michael M. Goodwin and Martin Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, July 1999.

[7] A. Rechtschaffen and A. Kales, Eds., *A manual of standarized terminology, techniques and scoring system for sleep stages in human subjects*, US Government Printing Office, 1968.

[8] P. J. Durka and K. J. Blinowska, "In pursuit of time-frequency representation of brain signals," in *Time frequency and wavelets in biomedical signal processing*, Metin Akay, Ed., IEEE Press Series in Biomedical Engineering, pp. 389–406. IEEE press, New Jersey, 1997.

Piotr J. Durka received his MSc and PhD degrees in Physics from Warsaw University. Since defending his dissertation on time-frequency analysis of EEG in 1997 he has continued working there as an Assistant Professor in the Laboratory of Medical Physics. His research is aimed at introducing methods of analysis for biomedical signals, especially EEG, which are not only advanced but useful. More detailed information and some of his papers are are at `http://brain.fuw.edu.pl/~durka`

Dobiesław Ircha received his MSc in 1997. Actually he is a graduate student at the Laboratory of Medical Physics, Institute of Experimental Physics, Warsaw University. His research interests include: source localisation, inverse problems, signal processing—wavelets, time-frequency, matching pursuit, artificial neural networks—for analysis of electroencephalogram (EEG), genetic algorithms, artificial intelligence.

Katarzyna J. Blinowska received MSc and in 1969 PhD degree in experimental physics from Faculty of Physics, Warsaw University, where she has been employed since graduation till now. In 1979 she received Sc.D. (doctor habilitatus) degree for her research on biological signal analysis. During 1982-1983 she was invited professor at the University of Southern California. Since 1984 she is a head of Medical Physics Laboratory at Warsaw University. In 1994 she became Full Professor. Her main research interests concern investigation of biophysical basis of electrical activity of central nervous system by modeling and signal analysis.