DESIGN OF WAVELETS ADAPTED TO SIGNALS AND APPLICATION

Aude MAITROT, Marie-Françoise LUCAS, Christian DONCARLI

Institut de Recherche en Communication et Cybernétique de Nantes (IRCCyN)
1, rue de la Noë, BP 92101, 44321 Nantes Cedex 03, France - {Maitrot, Lucas, Doncarli}@irccyn.ec-nantes.fr

ABSTRACT

This paper addresses the design of wavelets adapted to the processed signals and the considered application. Our approach consists of parameterizing a mother wavelet, and defining a quality criterion for the optimization of the parameters, according to the context. The first parameterization, leading to orthogonal wavelets, considers the coefficients of the scaling filter as the parameters. A second parameterization, leading to semiorthogonal wavelets, consists of convolving an existing wavelet (or scaling function) by a given sequence. In this paper, we explore these two methods and apply them to the supervised classification of signals made of waveform trains.

1. INTRODUCTION

Selecting a mother wavelet for a given problem is a delicate issue. Instead of choosing a wavelet from a catalogue more or less arbitrarily, an alternative approach consists of defining a general framework that enables principled wavelet selection. Such a framework needs to include $1)$ a family of wavelets that may depend on unknown parameters and $2)$ a quality criterion, written according to the context, to be used to perform wavelet selection (*i.e.*, wavelet parameters optimization). Formally, this consists of solving a parametric optimization problem under wavelet admissibility (and possibly regularity) constraints. This general framework can be applied to problems such as signal compression, classification or detection, where wavelets can be relevant tools.

Concerning point $1)$, we focus on two methods to parameterize discrete wavelet bases, driven by the Multiresolution Analysis (MRA) framework. Actually MRA allows for perfect reconstruction with fast algorithms and parameterization of the mother wavelet. The first wavelet design scheme directly considers the coefficients of the decomposition filters (to be used in the Mallat's pyramidal algorithm) as the parameters. The second method builds on a work by Abry *et al.* [1, 2] where it is shown that convolving an existing wavelet (or scaling function) with an admissible sequence provides a new wavelet associated to the same original MRA. Here, the parameters are the convolving sequence coefficients. Of course we can initialize the second method with the resulting wavelet of the first one.

The following step of the design process, point $2)$, is the optimization of the parameters w.r.t. a quality criterion. This criterion must be defined according to the context, that is the key point: it determinates the ability of the design methods to provide a mother wavelet adapted to the considered problem. This paper is motivated by a supervised classification problem: a natural quality criterion is the probability of classification error estimated on a training set (regularized risk). Signals are made of waveform trains (as encountered, *e.g.*, in medical diagnosis based on electromyography) with unknown shapes repeating with unknown scales and occurrence instants. Each class corresponds to one type of the waveforms characterizing the signals. The feature space is thus built from the decomposition of the signals on a dyadic wavelet basis. We describe the way to choose an optimal mother wavelet for this basis.

This paper is organized as follows. Section 2 addresses the two methods to parameterize the mother wavelet. In section 3, we describe the quality criterion and the classification process. Section 4 shows a toy example of application of the method.

2. TWO WAYS OF PARAMETERIZING THE MOTHER WAVELET

In this section, we present two methods to parameterize the mother wavelet, within the multiresolution analysis (MRA) framework. We denote θ the design parameter vector.

The MRA framework allows to define a scaling function ϕ and its associated wavelet ψ from two filters h and g thanks to the two-scale recursive relations : $\phi(t/2) = \sqrt{2}\sum_n h[n]\phi(t-n)$ and $\psi(t/2) = \sqrt{2}\sum_n g[n]\phi(t-n)$.

2.1. Orthogonal wavelets: parameterization of h

In the case of orthogonal wavelets, g can be deduced from h: $g[n] = (-1)^n h[1-n]$. Consequenty, by restricting to the orthogonal case, we only need the scaling filter h to define ψ. However, to generate an orthogonal MRA wavelet, h must satisfy some conditions. For a FIR filter of length Lh, there are $Lh/2+1$ sufficient conditions to ensure the existence and orthogonality of the scaling function and wavelets [4, 5]. There remains $Lh/2-1$ degrees of freedom that can be used to design the filter h.

The lattice parameterization described in [7] offers the opportunity to design h *via* unconstrained optimization: the Lh coefficients of h can be expressed in term of $Lh/2-1$ new free parameters. For instance, if $Lh=6$, we need a 2 component design vector, $\theta=[\alpha,\beta]$, and h is given by:

$$\begin{aligned}
i=0,1:\ & h[i]=4\sqrt{2}[(1+(-1)^i\cos\alpha+\sin\alpha)(1-\\
& (-1)^i\cos\beta-\sin\beta)+(-1)^i 2\sin\beta\cos\alpha]\\
i=2,3:\ & h[i]=2\sqrt{2}[1+\cos(\alpha-\beta)+(-1)^i\sin(\alpha-\beta)]\\
i=4,5:\ & h[i]=1/\sqrt{2}-h(i-4)-h(i-2)
\end{aligned}$$

For other values of Lh, expressions of h are given in [3, 6].

2.2. Semiorthogonal wavelets: wavelet linear combination

The second method starts from a given wavelet and builds a new wavelet by linear combination [1, 2]. The design parameters are the coefficients of the linear combination w.r.t. admissibility conditions.

Let ψ^o be the given original mother wavelet associated to an orthogonal or semiorthogonal MRA, with h^o and g^o its scaling and wavelet filters. We define a new wavelet ψ depending on a design admissible sequence p by:

$$\psi = p * \psi^o \tag{1}$$

The sequence p is admissible iff its Fourier transform $\hat{p}(f)$ satisfies:

$$A \le \underset{f\in[0,1/2]}{\text{ess-inf}}\,|\hat{p}(f)| \le \underset{f\in[0,1/2]}{\text{ess-sup}}\,|\hat{p}(f)| \le B \tag{2}$$

with constants $A,\ B>0$ [2]. Eq. (1) corresponds to a change of basis in the wavelet spaces. Consequently the approximation and wavelet spaces of the multiresolution are not changed (they remain orthogonal if they were), but an orthogonal wavelet ψ^o may lead to a semiorthogonal wavelet ψ (*i.e.*, the bases of the spaces are modified and may not be orthogonal anymore).

The filters associated to ψ are now:

$$\begin{aligned} h &= h^o \\ g &= \uparrow_2[p] * g^o \end{aligned} \tag{3}$$

where h^o and g^o are the filters associated to the original wavelet ψ^o. $\uparrow_2$ stands for upsampling by 2 (*i.e.*, $(\uparrow_2[x])(k)=x(k/2)$ if k is even, $(\uparrow_2[x])(k)=0$ if k is odd).

Here we need an existing MRA wavelet to initialize the method. The tuning parameters of ψ are the coefficients of the admissible sequence p. Thus the choice of the length Lp of p leads to the choice of the number of parameters to be optimized.

For our application, we chose to optimize a symmetrical sequence p of length $Lp=3$:

$$p=[\alpha\ \ 1\ \ \alpha] \tag{4}$$

in order to have only one component parameter vector: $\theta=[\alpha]$. In this case, we can easily write an analytic condition to satisfy the admissibility constraint (2): $|\alpha|<\frac{1}{2}$.

3. SUPERVISED CLASSIFICATION

We apply this design framework to a supervised classification problem of signals made of waveform trains. In this section, we present the feature space, derived from the Discrete dyadic Wavelet Transform (DWT). Then we define the decision rule and the quality criterion allowing to select the best parameters for the mother wavelet

3.1. Feature space

Given a mother wavelet ψ, the DWT decomposes the signal $x(t)$ on the corresponding discrete wavelet basis, where all the wavelets are dilated and translated versions of ψ. It provides a set of coefficients $d_x(j,k)=\langle x(t),\psi_{j,k}(t)\rangle$ where $\psi_{j,k}(t)=2^{-j/2}\psi(2^{-j}t-k)$. The N coefficients $d_x(j,k)$ of the decomposition of a discrete signal x of length N are computed by Mallat's pyramidal algorithm.

In order to make the representation space insensitive to the waveforms occurrence instants, we use the marginals of each level of the decomposition as the signals' features. We define the normalized marginals of the DWT by:

$$m_x(j) = \sum_{k=0}^{N/2^j-1} c_x(j,k), \qquad j=1,..,J \tag{5}$$

$$c_x(j,k) = \frac{|d_x(j,k)|}{\sum_{j=1}^{J}\sum_{k=0}^{N/2^j-1}|d_x(j,k)|} \tag{6}$$

where J is the deepest level of the decomposition ($J=\lfloor\log_2(N)\rfloor$). The features representing the signal x are the components of the vector $M_x=[m_x(1),..,m_x(J)]$. The vector M_x contains information on the distribution of the wavelet coefficients over J bands. It allows to represent the signal by the contributions of each frequency band (derived from a dyadic scale) computed with a chosen analyzing wave. Instead of offering this choice, a Fourier transform imposes sine waves.

3.2. Decision rule

Consider a two-class problem. The training set is splitted in ω_a and ω_b. We use the decision rule of the nearest representative briefly recalled: let R_a (R_b) the representative of

ω_a (ω_b), *i.e.* the average of M_x, x lying in ω_a (ω_b). Define a distance in the feature space, denoted $d(M_x, M_y)$, and an assigning variable $f(x) = d(M_x, R_a) - d(M_x, R_b)$. Then the decision rule writes:

$$\text{Assign } x \text{ to "}\omega_a\text{" ("}\omega_b\text{") if } f(x) < 0 \quad (f(x) > 0)$$

As the normalization (6) ensures:

$$\begin{cases} m_x(j) \geq 0, & j = 1, .., J \\ \sum_{j=1}^{J} m_x(j) = 1 \end{cases}$$

we can use a Kullback distance for $d(M_x, M_y)$:

$$d(M_x, M_y) = \sum_{j=1}^{J} \left[m_x(j) \log \frac{m_x(j)}{m_y(j)} + m_y(j) \log \frac{m_y(j)}{m_x(j)} \right] \tag{7}$$

As the decompositions depend on the design parameters gathered in vector θ, the superscript θ will appear in the following for all the concerned notations.

3.3. Quality criterion

A misclassified signal x of ω_a (ω_b) corresponds to a positive (negative) occurrence of $f^\theta(x)$:

$$f^\theta(x) = d(M_x^\theta, R_a^\theta) - d(M_x^\theta, R_b^\theta) \tag{8}$$

We assume that f^θ follows a Gaussian distribution (consequence of the central limit theorem and verified on a toy example). The overall probability of classification error (regularized risk) is:

$$\begin{aligned} P_e^\theta &= \frac{1}{2} \left(P_e^\theta(\omega_a) + P_e^\theta(\omega_b) \right) \qquad (9) \\ P_e^\theta(\omega_a) &\approx \frac{1}{\sigma_a^\theta \sqrt{2\pi}} \int_0^{+\infty} \exp\left(-\frac{1}{2} \left(\frac{z - \mu_a^\theta}{\sigma_a^\theta} \right)^2 \right) dz \\ P_e^\theta(\omega_b) &\approx \frac{1}{\sigma_b^\theta \sqrt{2\pi}} \int_{-\infty}^{0} \exp\left(-\frac{1}{2} \left(\frac{z - \mu_b^\theta}{\sigma_b^\theta} \right)^2 \right) dz \end{aligned}$$

where $\mu_a^\theta, \sigma_a^\theta$ ($\mu_b^\theta, \sigma_b^\theta$) are the empirical means and standard deviations of the assigning variable $f^\theta(x)$ estimated on the training signals of ω_a (ω_b).

We propose to optimize θ by minimizing the criterion (9) (exhaustive search : sampling of θ on a grid and computation of the criterion on each node). Note that the optimal criterion value can be used to assess the relevance of the choice of Lh or Lp (the number of design parameters) as it provides information on the classification error.

4. A TOY EXAMPLE

In this section, we apply the method on academic simulations and, for reasons of computing time, we restrict the optimization procedure to filters of length $Lh = 6,\ 8$ and/or

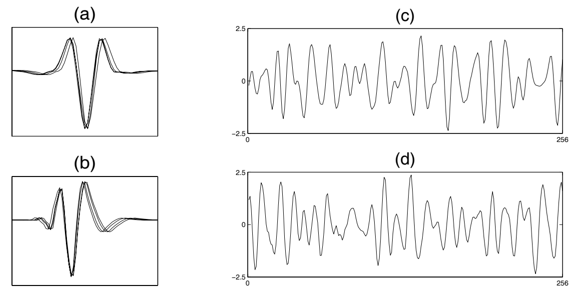

Fig. 1. Signals generation. The four dilated versions (non-integer dilation factor) of the waveform characterizing (a): the class ω_a, (b): the class ω_b. A signal realization (c): of class ω_a, (d): of class ω_b (each pulse of a random train is convolved by one of the four dilated versions of the waveform, randomly selected).

convolving symmetrical sequences of length $Lp = 3$ (see (4)). We present the signals generated as waveform trains, and discuss the classification results.

4.1. Signals

We generate two classes of signals. In each class, the signals are the result of the convolution of a pulse train with a waveform characterizing the class. More precisely, each pulse is convolved by one of four dilated versions of the waveform, randomly selected (non-integer dilation factor). Pulse occurrences are random for each signal, they appear with a probability $q = 0.2$. The signal's length is $N = 256$, and the waveforms' length is about 40. We generate 50 signals per class for the training set, and 1000 signals per class for the test set. Figure 1 shows an instance of signal generation for each class.

Notice that the waveforms used to generate these signals are supposed to be unknown. We do not try to detect these particular waveforms in the signals.

4.2. Results

The results of the classification of these academic signals are summarized in Tables 1 and 2. Table 1 deals with classification using, in the first column, a standard wavelet (*i.e.*, not optimized) and in the second one, an optimized orthogonal wavelet (direct optimization of the filter h, see Section 2.1). Table 2 shows the results concerning the linear combination approach (we convolve an existing wavelet with a sequence p and we optimize p, see Section 2.2). In the first column, we start from the previous *standard* wavelet. In the second column, we start from the previous *optimal* wavelet. Figure 2 shows the wavelet corresponding to the optimization of h in the case $Lh = 6$, and its optimal linear combination.

	Daubechies		Optimization of h	
Lh	Crit.	Mc rate	Crit.	Mc rate
6	0.26	23.4 %	0.10	11.6 %
8	0.13	10.8 %	0.05	4.1 %

Table 1. Results of the classification of academic signals, using orthogonal wavelets. The first column corresponds to the classification using a wavelet selected from a catalogue, here Daubechies 6 and 8. For the second column, the mother wavelet has been optimized according to the parameterization seen in Section 2.1. The notation "Crit." means the criterion value, "Mc rate" stands for misclassification rate.

	Init: Daubechies		Init: optimal h	
Lh	Crit.	Mc rate	Crit.	Mc rate
6	0.14	11.5 %	0.07	5.6 %
8	0.04	4.9 %	0.02	2.5 %

Table 2. Results of the classification of academic signals, using semiorthogonal wavelets. Here we use the second parameterization (see Section 2.2). In the first column, we optimize p by starting from a Daubechies 6 or 8. In the second column, we optimize p by starting from the optimal wavelet stemmed from the previous result (second column of Table 1). The notation "Crit." means the optimal criterion value, "Mc rate" stands for misclassification rate.

Three main points arise by analyzing these results. Firstly, Table 1 shows that classification performances are significantly improved by the mother wavelet opimization. Secondly, observing Table 2 after Table 1 reveals that optimizing linear combinations of wavelets also provides better classification results. Thirdly, as proved in Table 2, the best result is achieved by combining the two methods: optimizing the filter h in a first step and then, optimizing a linear combination of this resulting optimal wavelet. Comment: the computing time needed to successively optimize a filter h of length $Lh = 6$ and then a symmetrical sequence p of length $Lp = 3$ is much lower than optimizing a filter h of length $Lh = 8$, and leads to close results.

We can also notice the good suitability between the optimal criterion values and the misclassification rates: it proves that the regularization (proposed in Section 3.3) provides a classifier with good generalization capability. Moreover, the criterion value, computed on the training set, can be used to assess the relevance of the choice of Lh or Lp, before launching the classification procedure of the test set. By way of comparison, we have also performed classification tests with a Fourier spectral method instead of using wavelets (Kullback distances between normalized power spectra): it failed with around 40% of misclassified signals, confirming that a sine basis is not adapted to this type of classification problem.

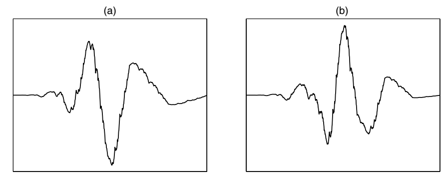

Fig. 2. Optimal wavelets stemmed from (a): the optimization of h with $Lh = 6$, (b): the optimal linear combination of the wavelet (a).

5. CONCLUSION

In this paper, we have defined a framework to generate mother wavelets adapted to signals and application. We have shown its efficiency for a classification problem of signals made of waveform trains (inpired by electromyogram signals, next forecast application). The mother wavelets are designed either from MRA filters, either from linear combinations of existing wavelets, but other ways are under study: linear combination of scaling function, lifting...

6. REFERENCES

[1] P. Abry, A. Aldroubi, "Designing Multiresolution Analysis-Type Wavelets and Their Fast Algorithms", Journal of Fourier Analysis and Applications, vol. 2(2), pp. 135-159, 1995.

[2] A. Aldroubi, M. Unser, "Families of multiresolution and wavelet spaces with optimal properties", Numerical Functional Analysis and Optimization, vol. 14(5-6), pp. 417-446, 1993.

[3] C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to Wavelets and Wavelet Transforms - A Primer", Prentice Hall, 1997, pp. 53-66.

[4] W. Lawton, "Tight frames of compactly supported affine wavelets", Journal of Mathematical Physics, vol. 31(8), pp. 1898-1901, 1990.

[5] W. Lawton, "Necessary and sufficient conditions for constructing orthonormal wavelet bases", Journal of Mathematical Physics, vol. 32(6), pp. 1440-1443, 1991.

[6] I. W. Selesnick, "Maple and the parameterization of orthogonal wavelet bases", Online, Oct. 1997. Available: http://taco.poly.edu/selesi/theta2h/

[7] P. P. Vaidyanathan, "Multirate Systems and Filter Banks", Wellesley-Cmabridge Press, 1996.