Notes for a course, Master di II livello in mathematica per le applicazioni

Alain Yger

June 8, 2004

1 Introduction

In order to write the sketch of this course, I inspired myself from the notes of two Master Courses I used to give in the last years in Bordeaux : "Traitement du Signal et Ondelettes" (for students in enginering) and "Traitement numérique et compression de données" (for students in pure mathematics or computer science interested into cryptography, codes, and security of information). I will try here to make some kind of synthesis of these two courses.

One could say that one of the major goals of this course is to propose a panel of methods (most of them inspired by orthogonality ideas) which combine Fourier analysis with wavelet or wavelet packets analysis (in order to supply the disadvantages of one method with the advantages of the other) towards the study of analogic or digital signals or images which are highly *non-stationary*, that is do not look as finite sums

$$t \to \sum_k a_k(t) \cos(\omega_k t) \qquad (\dagger)$$

(to simplifiy), where the coefficient functions a_k depend very moderately of t. Clearly, it is well known that in order to code music for example, one needs two parameters (and not one !), the "time" parameter (horizontal reading of the partition) and the "frequency" parameter (vertical reading of the partition) ; clearly, such a signal is not of the form (†) since there are "jumps" of frequencies everywhere. Most of the examples treated in the course will come from geology, electrocardiography, or also from the crucial problem of facing how could be decomposed the signals which are generated theoritically by the Navier-Stokes non linear system of differential equations (in order to understand turbulence phenomena).

The **MATLAB** routines which will be used in this course have been written taking essentially into account didactical aspects (not performances of the algorithms) since I am not neither a numerical analyst or a computer scientist ! They can be found (together with test-signals) on my web site :

https://www.math.u-bordeaux.fr/˜ yger

2 Discrete Fourier Transform or cosine transform

The *Discrete Fourier Transform* with order N is defined as the linear transform from $\mathbb{C}^N$ do $\mathbb{C}^N$:

$$\mathbf{DFT}_N \ : (s(1),...,s(N)) \in \mathbb{C}^N \to \Big[W_N^{kl}\Big]_{0\leq j,k\leq N-1} \bullet \begin{pmatrix} s(1) \\ \vdots \\ s(N) \end{pmatrix} \in \mathbb{C}^N ,$$

$$W_N := \exp(-2i\pi/N)\,.$$

The matrix of such a transform has the important property that

$$\Big[W_N^{kl}\Big]^{-1}_{0\leq k,l\leq N-1} = \frac{1}{N}\Big[\overline{W}_N^{kl}\Big]_{0\leq k,l\leq N-1}.$$

Such a matrix is nearly unitary and the unitary matrix

$$A_N := \frac{1}{\sqrt{N}}\Big[W_N^{kl}\Big]_{1\leq k,l\leq N}$$

satisfies $X^4 = \mathrm{Id}_N$, which means that its eigenvalues are $1, -1, i, -i$ (whatever the value of N is). There happens to be an ambiguity in choosing a basis of eigenvectors for the Discrete Fourier Transform.

On the $\mathbb{C}$-vectorial space $\mathbb{C}^N$ of digital signals with length N (coded in **MATLAB** as $s = [s(1),...,s(N)]$), equipped with the usual hermitian scalar product

$$\langle [x(1),...,x(N)]\,,\,[y(1),...,y(n)]\rangle := \frac{1}{N}\sum_{j=1}^{N} x(j)\,\overline{y(j)}$$

(corresponding to the discretized version of energy) there is a peculiar orthonormal basis which is directly related to A_N, therefore to the Discrete Fourier Transform with order N ; such a base is obtained taking the inverse image of the canonical orthonormal basis $e_j = \sqrt{N}(\delta_i^j)_{1\leq i\leq N}$, $j = 1, ..., n$, by the inverse of the linear map A_N. The vectors of this base are the N vectors

$$U_k \ := (\overline{\alpha}_k^0, ..., \overline{\alpha}_k^{N-1}) \,, \qquad \alpha_k := W_N^k \,, \ k = 0, ..., N-1 \,.$$

Since A_N is unitary, transforming a digital signal $[x(1), ..., x(N)]$ into $A_N[x]$ does not affect its energy

$$\frac{1}{N}\sum_{j=1}^{N} |x(j)|^2$$

but it can drastically change its Shannon entropy, which is defined, given some orthonormal basis $\mathcal{B}$ (here the base $e_1, ..., e_n$) and some digital signal with energy 1, as

$$\text{entr}(x, \mathcal{B}) = -\frac{1}{N}\sum_{j=1}^{N} |x(j)|^2 \log_2 \frac{|x(j)|^2}{N} \qquad (\dagger\dagger)$$

if $x(1), ..., x(N)$ are the coordinates of x in the usual canonical basis

$$(1, 0, ..., 0), ..., (0, ..., 0, 1) \,.$$

Note that this notion of entropy pairs the numerical value $|x_j|^2/N$ with $-\log_2(|x_j|^2/N)$, that is the number of digits which are necessary to represent $N/|x_j|^2$; in some sense, it is an indicator of the complexity of the signal : if there are many coefficients $|x_j|^2/N$ all equivalently small, the entropy will be much bigger than if there is just a few number of significative coefficients ! For example, the entropy of U_k equals $\log_2 N$ while the entropy of its image through A_N equals $\log_2 N/N$! Entropy being some indicator for chaos (such a notion comes from thermodynamics), checking for a basis $\mathcal{B}$ such that the entropy of some given information becomes minimal is a quite important challenge prior compression, identification and transmission of data. Comparing entropies when signals are expressed in different basis and profiting from such comparison will be a leitmotive in this course.

Within such a frame, one should say that the Discrete Fourier Transform of order N is the expected universal base change matrix (corresponding to the minimisation of the entropy) for any digital N-signal with energy 1 of the form

$$\sum_{j=1}^{r} a_j U_j \ ,$$

where r is "small" compare to N.

Some quite important remark respect to the vectors U_k, $k = 0, ..., N-1$, is the following : for any $k = 0, ..., N-1$, the cyclic matrix

$$B_{N,k} := \begin{pmatrix} 1 & \alpha_k & \dots & \alpha_k^{N-2} & \alpha_k^{N-1} \\ \alpha_k^{N-1} & 1 & \dots & \alpha_k^{N-3} & \alpha_k^{N-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_k & \alpha_k^2 & \dots & \alpha_k^{N-1} & 1 \end{pmatrix}$$

has rank equal to 1 and its only non trivial vector space corresponds to the eigenvalue N, with precisely eigenvector U_k.

When N is a power of two ($N = 2^p$), the computation of the action of the Discrete Fourier Transform can be realised with $N \log_2 N$ multiplications instead of N^2. This drastic improvement which is due to Cooley and Tuckey (1966) intitiated the "numerical revolution". Fourier transform which was before a theoretical tool (think about its role in the Radon transform introduced by J. Radon, together with inversion formulaes around 1917) became immediately a powerful operational technique (the developpment of Cat-Scanner theory and its variants for example, which is precisely based on the inversion of Radon transform and the theory Radon developped around 1917, took place about 70 years after Radon's pionneer work). Cooley Tuckey algorithm is based on the fact that to perform a 2^d Transform, one needs to do first two 2^{d-1} tranforms and then combine the outputs of these transform $(u(0), ..., u(2^{d-1}-1))$ and $(v(0), ..., v(2^{d-1}-1)$ by computing the vectors

$$[W_2^{kl}]_{0\leq k,l\leq 1} \bullet \begin{pmatrix} u(j) \\ v(j) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \bullet \begin{pmatrix} u(j) \\ v(j) \end{pmatrix}$$

for $j = 0, ..., 2^{d-1}-1$ (this does not "consume" any multiplication). The only thing which is needed (in order to initiate the procedure) is a re-classification of the input entries based on a re-indexation of the indices of the entries in terms of reversing their digits when they are expressed in binary system.

Since Cooley-Tuckey, S. Winograd proposed clever algorithms based of the Chinese Remainder Lemma in order to construct fast algorithms when is factorized with powers of small prime numbers. All such algorithms are now banal and used systematically as FFT (*Fast Fourier Transform*) algorithms.

The key operational role of the Discrete Fourier Transform is the following : if

$$\begin{aligned} P(X) &= \sum_{k=0}^{N-1} x(k) X^k \\ Q(X) &= \sum_{k=0}^{N-1} y(k) X^k \end{aligned}$$

are two polynomials (which in some sense "codify" the digital data that are $[x(0), ..., x(N-1)]$ and $[y(0), ..., y(N-1)]$), then a basic operation is the one that consists in associating to this pair of polynomials (or to the pair of corresponding digital data x and y) the digital output $[z(0), ..., z(N-1)]$ corresponding to the polynomial

$$R(X) = \sum_{k=0}^{N-1} z(k) X^k$$

which is the remainder of PQ in the Euclidean division by $X^n - 1$. One has

$$z(k) = \sum_{j=0}^{N-1} x(j) y(k-j)$$

(the input x and y being periodized with period N). This operation

$$(x, y) \to z = x * y$$

is the *cyclic correlation* and it will appear to be fundamental in the treatment of information. Through the Discrete Fourier Transform, it becomes just pointwise multiplication, namely

$$\mathrm{DFT}_N[x * y](k) = \mathrm{DFT}_N[x](k) \times \mathrm{DFT}_N[y](k)$$

for $k = 0, ..., N-1$, and is therefore much simpler to handle !

The Discrete Fourier Transform can easily be transposed to multi-variate analogs. Nevertheless, in image processing, one deals (for symmetry reasons, which are directly related to the fact that the transform must preserve symetries, note that cos is a "symetrization" of the exponential while sin does not correspond as such a symetrization, but more at some kind of disymetrization !) also with the $2D$- (N_1, N_2) Cosine Transform which transforms a real image $I(k_1, k_2)$ into the image $\widehat{I}(q_1, q_2)$, where

$$\widehat{I}(q_1, q_2) = 4 \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} I(k_1, k_2) \cos\Big[\frac{\pi q_1(2k_1+1)}{2N_1}\Big] \cos\Big[\frac{\pi q_2(2k_2+1)}{2N_2}\Big].$$

Such a transform refers to another orthogonal system for $\mathbb{R}^N$, the *cosine system*, which is generated by the vectors

$$V_q := \Big(2\cos\Big[\frac{\pi q(2j+1)}{2N}\Big]\Big)_{j=0,\dots,N-1}.$$

Such a transform is the basic mathematical in standard versions of JPEG, where the input image I (with size $2^p \times 2^p$, $p \geq 3$, is preliminary divided into 8×8 subimages, each subimage being transformed by the *Discrete Cosine Transform* into some 8×8 matrix (where the compression process will be performed) ; once such a "cleaning" of the spectrum of the image has been realized, 8×8 blocks (in the wave number domain) are treated *via* compression techniques and their size will be reduced ; after this is done for each 8×8 subimage of the spectrum of I, one recovers a compressed version of I thanks to the Inverse Discrete Cosine Transform.

Note that the orthogonal systems that lie behind the Discrete Fourier Transform or the Discrete Cosine Transform are (if we transfer discrete problems to continuous ones) respectively the orthonormal basis

$$\Big\{t \to \sqrt{1/T} e^{2i\pi k(t-a)/T}, \quad k \in \mathbf{Z}\Big\}$$

or

$$\Big\{t \to \sqrt{2/T} \sin\frac{\pi(k+1/2)(t-a)}{T}, \quad k \in \mathbb{N}\Big\}$$

as well as

$$\Big\{t \to \sqrt{2/T} \cos\frac{\pi(k+1/2)(t-a)}{T}, \quad k \in \mathbb{N}\Big\}$$

or

$$\Big\{t \to \sqrt{2/T} \cos\frac{\pi(k+1/2)\Big(t - \frac{2a+T}{2}\Big)}{T}, \quad k \in \mathbb{N}\Big\}$$

for $L^2([a, a+T])$ (with its usual scalar product corresponding to continuous energy). We will refer to such orthonormal basis in this course in order to understand the key relation between Discrete Fourier Transform, Discrete Cosine Transform and stationary signals (in the deterministic sense).

3 Deterministic notion of "stationarity" ; about sums of elementary harmonics

In the sequel, we will call *deterministic analogic signal* any locally integrable function on the real line ; in fact, one should think always such a signal as a distribution, since one needs to incorporate in this definition signals like $\delta(t-t_0)$, where δ denotes the Dirac mass ; nevetherless, such signals can be approximated by functions (for example $\delta(t) \simeq (1/\epsilon)\chi_{]-\epsilon/2,\epsilon/2[}$ for $\epsilon << 1$) and we will treat them as functions. *Digital deterministic signals* will be sequences of complex numbers indexed by $\mathbb{Z}$.

A deterministic analogic signal is stationary (in the deterministic sense) if it is in the closure of the $\mathbb{C}$ vectorial space generated by the

$$t \to \exp(i\omega t)$$

(such an "atom" being called an *elementary harmonic*) respect to the ergodic scalar product

$$\langle s_1\,,\, s_2\rangle_{\text{erg}} := \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} s_1(t)\overline{s_2(t)}dt\,.$$

Any stationary signal can be expressed as

$$s \simeq \sum_{\omega\in\Lambda(s)} a_\omega(s)\, e^{i\omega(\cdot)}\,,$$

where Λ is a <u>countable</u> set (called the *spectrum* of s), the convergence of the sum being in terms of the ergodic norm

$$\limsup_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} |s(t)|^2\, dt = \|s\|^2_{\text{erg}}\,,$$

so that one has the Plancherel formula

$$\|s\|^2_{\text{erg}} = \sum_{\omega\in\Lambda(s)} |a_\omega(s)|^2\,.$$

For such a signal, it is quite important to note that

$$(t_1, t_2) \to \langle s(\cdot + t_1)\,,\, s(\cdot + t_2)\rangle_{\rm erg} = \sum_{\lambda\in\Lambda(s)} |a_\omega(s)|^2 e^{i(t_1-t_2)\omega} = R_s(t_1 - t_2)$$

and therefore <u>depends only of $t_1 - t_2$</u>. Such a function R_s is called the *auto-correlation function* of s.

Let us point out another important point of view respect to the caracterization of particular stationary analogic signals which can be expressed as

$$s\ : t \to \sum_{j=1}^{m} a_j \exp(i\omega_j t)\,,$$

where $\omega_1, ..., \omega_r$ are real frequencies and $a_1, ..., a_m$ complex coefficients (the modulus of a_j is usually called the *amplitude* of s, the argument of a_j as the *phase* attached to the wave number ω_j). It comes back to Euler that such a signal s obeys to a m-order differential equation with constant coefficients

$$(D^m - A_1 D^{m-1} - \cdots - A_m)[s] \equiv 0\,,$$

where $D = d/dt$ and

$$X^m - A_1 X^{m-1} \cdots - A_m = \prod_{j=1}^{m} (X - i\omega_j)\,.$$

If the situation is now discretized (with a normalized sampling step equal to 1), one can view the differential operator d/dt as the analogic corresponding of (for example, there are other possible choices) of the operator

$$(x(k))_k \to (x(k) - x(k-1))_k\,;$$

then, the discretized version of the analogic signal s obeys a difference equation of the form

$$s(k) = \gamma_1 s(k-1) + \cdots + \gamma_m s(k-m)\,, \qquad (*)$$

which means that the digital signal $(s(k))_{k\in\mathbb{Z}}$ is correlated with all the digital signals $(s(k-1))_k$,...., $(s(k-m))_k$, which appear to be "*shifted versions*" of the original data s towards its past. As a matter of fact, such a correlation relation as $(*)$ controls (and in fact almost governs) the coherence of s ; looking just for the frequency contend of s (and not for the search for the coefficients a_j, which has to be done in some further step) is equivalent to looking for the optimal choice of parameters $\gamma_1, ..., \gamma_m$ such that $(*)$ "almost" holds, m being *a priori* fixed, which will appear as some stumbling block in the procedure (nevertheless, there will be some way to turn around it !).

4 Autocorrelation of digital signals and referent algorithms

4.1 The notion of autocorrelation matrix

Let $s = (s(k))_{k\in\mathbb{Z}}$ be a digital signal indexed by $\mathbb{Z}$ and N be some strictly positive integer ; in order to introduce some way to "explore" the digital signal s having in mind the idea that it may locally be expressed as the discretized version of a finite sums of harmonics, we profit from one of the two points of view we discussed in section 3 respect to the behavior of finite sums of exponentials.

In this first subsection, we will focuse mainly of the ergodic point of view (which was the first one we proposed in section 3). Therefore, it is natural to introduce some quantity to measure the "autocorrelation" of the digital signal.

In order to do that, one needs to introduce an integer N which is chosen intituively with the *a priori* idea that the digital signal remains essentially stationary on digital segments with length $2N-1$. Of course, such an hypothesis is an *a priori* hypothesis, but we will see later in this course how some kind of dichotomy argument may help towards the choice of N.

Once N has been choosen, the algorithmic procedure lies in the choice of two "windows" :

- a <u>temporal</u> window $g = [g(1), ..., g(N)]$ such that $g(1)+\cdots+g(N) = N$;
- a <u>spectral</u> window $h = [h(0), ..., h(N-1)]$, $0 \leq h(k) \leq 1$, such that $h(k) = h(N-k)$ for $k = 0, 1, ..., [N/2]$ and $h([N/2]) = 1$.

The reason for the terminology used here will be more transparent later on. Once g and h are choosen, one introduces naturally a fonction from $\mathbb{Z}$ to $\mathcal{M}_{N,N}(\mathbb{C})$ which is defined as

$$\begin{aligned} n \in \mathbb{Z} \quad &\to \quad \mathrm{AUTOCORR}_{g,h}[s\,;\,n] \\ &:= \quad \frac{1}{\|h\|_2^2}\left[\frac{h(q_1)h(q_2)}{N}\sum_{l=1}^{N} g(l)s(n+l+q_1)\overline{s(n+l+q_2)}\right]_{0\leq q_1,q_2\leq N-1}. \end{aligned}$$

Such a function is called *windowed autocorrelation matricial function* ; it depends on the choice of g and h ; the standard choices are $g = g_N :=$

$[1,1,...,1]$ and $h = h_N := [1/N,...,1/N]$; another choice is $g := [N,0,0,...,0]$ and h arbitrary (still with the restriction conditions imposed before).

For any $n \in \mathbb{Z}$, the hermitian matrix $\text{AUTOCORR}_{g,h}[s\,;\,n]$ has N complex eigenvalues $\lambda_1(n),...,\lambda_N(n)$, that for the sake of simplicity we will here suppose such that

$$|\lambda_1(n)| > |\lambda_2(n)| > \cdots > |\lambda_N(n)|\,.$$

(we decide to skip pathological values of n where this fails to happen). The **svd** command (*Singular Value Decomposition*) in **MATLAB** provides (in decreasing order) the modulus of the eigenvalues as well as corresponding normalized (respect to the euclidean norm) eigenvectors $v_1(n),...,v_N(n)$; of course, the search for such a basis of eigenvectors is only reliable when the eigenvalues $\lambda_1(n),...,\lambda_N(n)$ have distinct modulus, which we suppose here.

When $g = [1,...,1]$ and $h = [1/N,...,1/N]$, we will just write for the sake of simplification $\text{AUTOCORR}_{g,h}[s\,;\,n] = \text{AUTOCORR}[s\,;\,n]$. One can notice then that as soon as there exists integers $k_1,...,k_m$ between 0 and $N-1$ and complex coefficients a_{n,k_r} $(r = 1,...,m)$ such that

$$s(l+1) = \sum_{r=1}^{m} a_{n,k_r} \overline{W_N}^{k_r l} \qquad \forall l \in \{n,...,n+2(N-1)\}\,,$$

then, one has

$$\text{AUTOCORR}[s\,;\,n] = \frac{1}{N^2}\sum_{r=1}^{m} |a_{n,k_r}|^2\, B_{N,k_r}\,.$$

As a consequence of the key properties of the $B_{N,k}$ for any $k \in \{0,...,N-1\}$, we have, if $U_k := (1,\overline{\alpha}_k,...,\overline{\alpha}_k^{N-1})$,

$${}^tU_k \bullet \text{AUTOCORR}[s\,;\,n] \bullet \overline{U_k} = \sum_{r=1}^{m} |a_{n,k_r}|^2\, \delta(k-k_r)\,.$$

Coming back to the general situation where the temporal and frequential window g and h are now arbitrary ones (still with the imposed constraints $g(1)+\cdots+g(N) = N$, $0 \le h \le 1$, $h([N/2]) = 1$, $h(k) = h(N-k)$ for $k \le [N/2]$) the paragraph just above justifies the fact that, given N and the two windows g and h, the function

$$(n,k) \to {}^tU_k \bullet \text{AUTOCORR}_{g,h}[s\,;\,n] \bullet \overline{U_k}$$

may be called the *N-Windowed Power Spectral Density* of the signal s (relatively to the choice of the temporal and frequential windows g and h). This function will play an important role if we make the *a priori* assumption that the signal s remains stationary during time intervals with uniform lenght equal approximatively to N (which of course is never true in practice, but such an hypothesis is necessary to be assumed in order to draw conclusions on the spectal content of the signal). If s is highly non stationary, even locally, then of course, the study of the funtion

$$(n, k_1, k_2) \to {}^tU_{k_1} \bullet \mathrm{AUTOCORR}_{g,h}\,[s\,;\,n] \bullet \overline{U_{k_2}}$$

would be much more adequate (though much more difficult to handle since one represent as an image a fonction of two variables, not a function of three variables).

A statistical information respect to the global spectral content of the digital signal s is given by averaging the function

$$(n, k) \to {}^tU_k \bullet \mathrm{AUTOCORR}_{g,h}\,[s\,;\,n] \bullet \overline{U_k}$$

as a function of n ; such averaging leads to a function

$$k \in \{0, ..., N-1\} \to \mathrm{mean}_n \left[{}^tU_k \bullet \mathrm{AUTOCORR}_{g,h}\,[s\,;\,n] \bullet \overline{U_k}\right]$$

This function of k is the *Power Spectral Density* of the signal (computed respect to the Welch method) and the routine **spectrum** in **MATLAB** realizes such a computation (taking in the simplest case $g = [N, 0, ..., 0]$). Its graph provides a <u>statistical</u> information on the frequency content of a given signal (see the help of the command **spectrum** in **MATLAB** for practical details). Note that only positive frequencies are kept so that, if $\tau = 1$ corresponds to the normalized rate of sampling, the range of potential frequencies is $[-\pi, \pi]$ and that of positive frequencies is therefore $[0, \pi]$ (since $W_N^{N-k} = W_N^{-k}$ and $2k\pi/N$ lies between 0 and 2π when k varies between 0 and $N-1$). The **spectrum** command provides the value of the Power Spectral Density on $[0, \pi]$ only.

4.2 The spectrogram (*Widowed Fourier Transform*)

Clearly, the statistical information which is provided by the Power Spectral Density does not take in account the time-frequency evolution of the signal.

In order to take it into account, it is necessary to plot the image of the two-variable function :

$$(n, k) \to {}^tU_k \bullet \text{AUTOCORR}_{g,h}\,[s\,;\,n] \bullet \overline{U_k}\,,$$

instead of just its average in terms of n.

The simplest case is the case when $g = [N, 0, ..., 0]$ and h is arbitrary (but fulfills the required conditions). In this case, the function

$$(n, k) \to {}^tU_k \bullet \text{AUTOCORR}_{g,h}\,[s\,;\,n] \bullet \overline{U_k}$$

equals

$$(n, k) \to \frac{1}{\|h\|^2}\left|\frac{1}{N}\sum_{q=0}^{N-1} h(q)s(n+1+q)W_N^{kq}\right|^2,$$

and the square root of it, that is the function

$$(n, k) \to \frac{1}{\|h\|}\left|\frac{1}{N}\sum_{q=0}^{N-1} h(q)s(n+1+q)W_N^{kq}\right|$$

is called the *Spectrogram* of s, the complex valued function

$$(n, k) \to \frac{1}{N\|h\|}\sum_{q=0}^{N-1} h(q)s(n+1+q)W_N^{kq}$$

being the *Widowed Discrete Fourier Transform* of the digital signal s.

The spectrogram is a fundamental tool for example for speech processing ; its concept fits with the concept of musical coding. It provides precise information on the evolution of the spectrum of the signal as a function of the time. Of course, one has to assume that the signal is essentially stationary on any time interval with lenght N, therefore, appears the technical difficulty of choosing N. Note also that we face here the first crucial stumbling block of Fourier analysis : if N has to be chosen small (that is the evolution of the spectrum as a function of the time is fast), the number of frequency channels available is also equal to N (the Discrete Fourier Transform matrix is a square matrix !), that is the resolution in frequency is bad ; on the other hand, if N is large, the resolution in frequency is better, but the signal is not stationary anymore on windows with length N, so that the spectrogram does not take into account all components of the information. Here of course, the empirical search for some kind of "compromise" is crucial, as the examples proposed in the course illustrate it.

The **MATLAB** routine **wfft** provides the computation of the spectrogram, which has then to be read as an image (the horizontal axis represents the time evolution, the vertical axis from top to bottom the frequency range between 0 and π).

Note that, if s is the discretization of some analogic signal containing components $t \to e^{\pm i\omega t}$ with $|\omega| > \pi$, then we hit when we take the sampling at the rate $\tau = 1$ the *Undersampling Problem*, which is the second crucial stumbling block of Fourier Analysis. We will come back to this problem later on in this course.

4.3 The MUSIC indicator of frequencies

There is another interesting way to profit from the information provided by the function

$$(n,k) \to U_k \bullet \mathrm{AUTOCORR}_{g,h}\,[s\,;\,n] \bullet \overline{U_k}\,.$$

Let m be fixed between 1 and M (m is intuitively the number of frequencies which we suppose are involved in the signal, even considered in the moving window). Then, if $[s(n+1),...,s(n+N)]$ was really a linear combination of some vectors $U_{k_1},...,U_{k_m}$, then these vectors (once normalized) would be the m eigenvectors $v_1(n),...,v_m(n)$ corresponding to the eigenvalues $\lambda_1(n),...,\lambda_m(n)$ of the matrix AUTOCORR $[s\,;\,n]$ (other eigenvectors would correspond to the eigenvalue 0) and then one should have :

$$1 - \frac{1}{N}\sum_{j=1}^{m}\left|\langle v_j(n)\,,\,U_k\rangle\right|^2 = \begin{cases} 0 \text{ if } k = k_1,...,k_m \\ 1 \text{ if } k \neq k_1,...,k_m\,. \end{cases}$$

Therefore, it is natural to consider as a potential indicator for the position of frequencies the function

$$(n,k) \to \frac{1}{1 - \frac{1}{N}\sum\limits_{j=1}^{m}\left|\langle v_j(n)\,,\,U_k\rangle\right|^2}$$

Such an indicator (constructed from the autocorrelation matrix built with $g = [1,...,1]$ and h which fulfills the required conditions to be a frequential window) is called a **MUSIC** indicator and we use the routine **music3** in **MATLAB** to compute it.

The use for such an indicator appears quite useful to separate freqencies that could be close ; it provides cleaner versions of the spectrogram (but only the support of the Power Spectral Density is figured, not the values of this function).

4.4 The search for the optimal autoregressive filter

An autoregressive filter (AR filter) is a linear operator from the space of complex sequences $\mathbb{C}^{\mathbb{Z}}$ into itself which transforms (formally) the input sequence $(e(k))_{k\in\mathbb{Z}}$ into the output sequence $(s(k))_{k\in\mathbb{Z}}$, such that

$$e(n) = s(n) - \gamma_1 s(n-1) - \cdots - \gamma_m s(n-m) \;;$$

in order to compute the values of the output for any $n \geq N_0$, it is necessary to fix initial values for $s(N_0-1), ..., s(N_0-m)$. The **MATLAB** routine **fir** computes the action of such a filter on some input data once the parameters γ_j, $j = 1, ..., m$, are fixed, as well as the initial values $s(N_0-1), ..., s(N_0-m)$.

Any discrete sum of exponentials satisfies a difference equation (as seen in section 3). Therefore, a digital stationary signal can be interpreted as the output through the action of such an autoregressive filter of a digital signal $(e(n))_{n\in\mathbb{Z}}$ which is completely decorrelated, that is, if the statistical mean of $(e(n))_{n\in\mathbb{Z}}$ is

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=-[N/2]}^{[N/2]} e(k) = m\,,$$

then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=-[N/2]}^{[N/2]} (e(k)-m)(\overline{e(k+l)-m}) = \begin{cases} \sigma_m^2 \text{ if } l=0 \\ 0 \text{ if } l \neq 0\,. \end{cases}$$

Such a digital signal $(e(k))_k$ is called a *white noise.*

Then, given a digital signal s that one supposes stationary on the time interval $\{1, ..., N\}$, it is interesting to choose $m << N$ and to look for the parameters γ_0,..., γ_m such that

$$\sum_{k=m+1}^{N} \left| s(k) - \gamma_0 - \sum_{j=1}^{m} \gamma_j (s(k-j) - \gamma_0) \right|^2$$

is minimal. This happens to be a problem of minimization in the sense of least squares ; one needs to look for the projection of the digital signal s on the vectorial subspace generated by its shifted versions $s(k-j)$, $j=1,...,m$, and by the constant signal $(1,...,1)$. Such a problem can be solved provided a certain Gram matrix (which one has to invert) is well conditionned. If it happens not to be well conditionned, a simple trick to turn around the difficulty is to add to the signal s a white noise (generated in **MATLAB** with the routine **random**).

Once this is done, a candidate for the Power Spectral Density of s (considered as stationary on $\{1,...,N\}$ will be the function

$$\omega \to \frac{\sigma_m^2}{|1-\gamma_1 e^{-i\omega} - \cdots - \gamma_m e^{-im\omega}|^2},$$

where σ_m^2 is the *residual variance*, that is

$$\sigma_m^2 = \frac{N-m}{N(N-2m-1)} \sum_{k=m+1}^{N} \Big| s(k) - \gamma_0 - \sum_{j=1}^{m} \gamma_j (s(k-j) - \gamma_0) \Big|^2 .$$

Such a method appears to be quite used for example in studying biological signals, where there happens to be a limited number of frequencies which fluctuate in rather well separated ranges ; for example, signals involved in the description of the cardiac rythm (the digital signal is then obtained taking the distance from one peak to the next one in some electrocardiogram) are of this form, in relation with the sympatico-vagal balance. For these reasons, it is quite important to sketch it here. Note that there is an ambiguity related to the choice of m ; theoretically, the function

$$m \to \sigma_m^2$$

should be convex and its minimum corresponds to the optimal choice of m (see the work of H. Akaike, "Fitting autoregressive models for prediction", Ann. Inst. Statist. Math. 21, 243-247, 1969, and G. Schwarz, "Estimating the dimension of a model", Ann. Statist. 6, 461-464, 1978) ; nevertheless, this is quite difficult to check in pratice and the choice of m is usually empirical.

5 The "chirp" model, the *Wigner-Ville Transform*

As a step from Fourier analysis towards Wavelet analysis, we may introduce another interesting dictionary of atoms than the atoms $t \to \exp(i\omega t)$ one is interested in when dealing with the (even local) stationarity hypothesis ; such atoms are the *Gaussian chirps*, of the form

$$t \to \exp(-\alpha(t-\beta)^2 + iP(t))\,,$$

where P is a polynomial with degree at most 2 with real coefficients, which can be written $P(t) = \rho(t-\gamma)^2 + \delta,\ \rho, \gamma, \delta \in \mathbb{R}$. We notice then that, for $\tau \in \mathbb{R}$,

$$\overline{s(t-\tau/2)}s(t+\tau/2) = \exp\Big(-2\alpha\Big[(t-\beta)^2 + \tau^2/4\Big]\Big) \times \exp(2i\rho(t-\gamma)\tau)\,,$$

which means that, if t is fixed, such a signal as an "instantaneous frequency" which is $2\rho(t-\gamma)$ (as a function of τ) ; it is this signal (function of τ) which is stationary, <u>not</u> the original gaussian chirp.

The continous transform which then seems of interest then is the transform

$$(t,\omega) \to \mathrm{WV}\,[s,s\,;\,t,\omega] := \frac{1}{2\pi}\int_{\mathbb{R}} \overline{s(t-\tau/2)}s(t+\tau/2)e^{-i\omega\tau}d\tau\,.$$

This transform is the *Wigner-Ville Transform* and its discrete version transform a signal $s = [s(1), ..., s(N)]$ into the image

$$(n,k) \to \sum_{l=0}^{N-1} s(n+l)\overline{s(n-l)}W_N^{kl}\,. \qquad (**)$$

The routine **wig0** under **MATLAB** will be used to generate the Discrete Wigner-Ville Transform ; note that the signal s has been extended by 0 outside $\{1, ..., N\}$ in order that computation $(**)$ is possible.

The fact that the Wigner-Ville Transform is quadratic (and non linear) explains why the images one obtains are blurred by some interference terms ; in fact

$$\begin{aligned}\mathrm{WV}\,[s_1+s_2, s_1+s_2\,;\,t,\omega] &= \mathrm{WV}\,[s_1,s_1\,;\,t,\omega] + \mathrm{WV}\,[s_2,s_2\,;\,t,\omega] \\ &\quad + \frac{1}{\pi}\,\mathrm{Re}\left[\int_{\mathbb{R}} \overline{s_1(t-\tau/2)}s_2(t+\tau/2)e^{-i\omega\tau}d\tau\right].\end{aligned}$$

Such interference terms are oscillating ones, so that local averaging of the image may help to attenuate then ; the routine **wiglis1h** does this job on our examples ; more subtile techniques (based on image processing, such that the one which consists in reassigning the value obtained at a given pixel at the local center of mass of the image close to this pixel) can be introduced, which ameliorate the lisibility of the Wigner-Ville images (see the book of P. Flandrin, *Temps-fréquences*, Hermès, Paris, 1993 and a method introduced by F. Auger and P. Flandrin, IEEE Transactions on Signal Processing 43,5, 1068-1089, May 1995).

Despite the difficulties linked with its treatment, the Wigner-Ville Transform (and the correlated ones) is a step from the analysis of stationary signals towards the analysis of signals where evolution of frequencies is linear (as function of the time). Note also that energy (hence orthogonality) is preserved in the following sense : if $s_1, s_2 \in L^2(\mathbb{R})$ are two analogic signals with finite energy, then

$$\left| \int_{\mathbb{R}} s_1(t)\overline{s_2(t)}\, dt \right|^2 = \frac{1}{2\pi} \iint_{\mathbb{R}^2} \mathrm{WV}\,[s_1, s_1\,;\, t, \omega]\, \overline{\mathrm{WV}\,[s_2, s_2\,;\, t, \omega]}\, dt d\omega$$

(this is known as *Moyal's formula*).

The Wigner-Ville Tranform realizes some intermediate step between Fourier analysis and Wavelet analysis (which is based on the simultaneous analysis in time ans scale, instead of time and frequency, and will be the next topic developped in the course).

6 The pyramidal algorithm of Burt & Adelson ; the concept of multiresolution analysis

6.1 Vision and pyramidal algorithm

Let $(s(n))_{n\in\mathbb{Z}}$ be a sequence of complex numbers ; one can also think it represents the graph of the piecewide linear function which interpolates these values.

Imagine an observator looking at this graph from far away ; the observator will get a synthetic view of the graph, for example, he will apprehend the digital signal $R_1[s]$ (R_1 for "résumé" or "summary") defined as

$$R_1[s](k) = \sum_{n\in\mathbb{Z}} s(n)h(n-2k)\,,$$

where $\sum\limits_n h(n) = 1$ and $\sum_n h(2n) = \sum_n h(2n+1) = 1/2$ corresponds to some averaging of the initial date s ; the first condition imposed on h correspond to the fact that each value $R_1[s](k)$ needs to be an average of values of s and the second one will be justified a bit later. Note that $R[s]$ has to be considered as a digital signal on $2\mathbb{Z}$, not on $\mathbb{Z}$ (there is change of scale since the observator apprehends the digital signal from far away).

Once such a resumed digital signal $R_1[s]$ is stocked in memory, the observator (this is now a brain effort) may reconstruct a blurred version of s, which is obtained as the sequence $\widetilde{s}$ defined by :

$$\widetilde{s}(n) = \sum_{k\in\mathbb{Z}} R_1[s](k)h(n-2k)\,;$$

this corresponds to a redistribution of the stocked information and the difference $(d_1[s](n))_{n\in\mathbb{Z}}$ defined as

$$d_1[s] = s - \widetilde{s}$$

corresponds to the details of s at the scale $2^0 = 1$, details that cannot be catched by the process that leads to the blurred version $\widetilde{s}$.

Note that the two operators

$$\begin{aligned} R\ &: (s(n))_n \ \to \ \left(\sum_{n\in\mathbb{Z}} s(n)h(n-2k)\right)_k \\ R^*\ &: (u(k))_k \ \to \ \left(\sum_{k\in\mathbb{Z}} u(k)h(n-2k)\right)_n \end{aligned}$$

are adjoints one to each other (as operators from $l^2(\mathbb{Z})$ into itself). It is the redistribution phase which make natural the conditions $\sum_k h(2k) = \sum_k h(2k+1)$ mentionned above : at all point, the sum of the averaging coefficients should be the same, that is here 1/2.

The procedure may be carried on, starting with $R_1[s]$, and the signal

$$d^{(2)}[s] = R_1[s] - \widetilde{R_1[s]}$$

corresponds to details of the digital signal s at the scale 2 ; of course $d^{(1)}$ and $d^{(2)}$ connot be added since they do not correspond to the same scaling of $\mathbb{Z}$ (1 is the scale step for $d^{(1)}$, 2 the scale step for $d^{(2)}$). Nevertheless, the list $d^{(1)}, d^{(2)}, ..., d^{(N)}, R_N[s]$ allows the iterative reconstruction of s thanks to the action of the operators R^* and the formula

$$R_k[s] = d^{(k)}[s] + 2R^*[R_{k+1}[s]] \,.$$

Such an algorithm (which announces the concept of multiresolution analysis) is called *Pyramidal Algorithm* because it compbines the two phases of averaging and redistributing the information in some kind of pyramidal way (draw a diagram to convince yourselves about it !).

Testing this algorithm for example on test-signals we used to illustrate the spectrogram will be done in the course thanks to the routine commands **rpyramid** (used for the phase of "stockage" *via* dyadic compression of the data) and **dpyramid** (used for the phase of redistribution of the information) ; note that both programs are very simple to implement, taking here for example $h(l) = 0$ for $|l| > 2$, $h(0) = a$, $h(1) = h(-1) = 1/4$, $h(2) = h(-2) = (1-2a)/4$, where a is a parameter to be choosen between 0 and 1 (this model corresponds to averaging with an order 3 spline) ; we will choose various values of a in the illustrations performed in the course.

6.2 An orthogonal variant : Franklin-Strömberg decomposition

Another variant of the pyramidal algorithm is the quite useful (and also quite naïve) *Franklin Decomposition Algorithm*. The idea is to combine the synthetic description of the vision process in the discrete context in two steps as it is proposed by the pyramidal algorithm of Burt and Adelson with some interpretation of discrete signals as obtained by sampling analogic ones. This will happen to be a key idea later on, when we will have to understand how to choose a multiresolution analysis in order to treat some discrete information. In fact, a digital signal $(s(n))_n$ appears as the sampled version of the analogic signal

$$t \to \sum s(n)\Delta(t-n) \,,$$

where

$$\Delta \ : t \to \max(0, 1-|t|)$$

is the basic (here just continuous) spline function ; assuming more regularity *a priori* on the analogic model from which one assumes $(s(n))_n$ is a sampled version would lead to the choice of a spline of higher order instead of Δ ; let us take here Δ which is the simplest choice.

If $N = 2^q$, a basis for the $N+1$-dimensional $\mathbb{C}$-vectorial space $V_{0,N}$ of analogic signals s on $[0, N]$ which have nodes at $0, ..., N$ is the collection of functions $\Delta_{0,n}$, $n = 0, ..., N$, where

$$\Delta_{0,n}(t) = \max(0, 1 - |t - n|) = \Delta(t - n) \qquad \forall t \in [0, N]\,.$$

Such a vectorial space corresponds to the $N+1$-dimensional space of digital signals with lenght $N+1$. In order to construct a "blurred" version of such a digital signal $s = [s(1), ..., s(N+1)]$, one can do the following :

- form the analogic signal $S\ : t \in [0, N] \to \sum\limits_{j=1}^{N+1} s(j)\Delta_{0,j-1}(t)$;
- project orthogonally S on the $\mathbb{C}$-vectorial subspace of $V_{0,N}$ which is generated by the functions $\Delta_{1,n}$, $n = 0, ..., 2^{q-1}$, where

  $$\Delta_{1,n}(t) = \max(0, 1 - |t/2 - n|) \qquad \forall t \in [0, N]\,,$$

  which gives an analogic signal $R_1[S]$ which is piecewise linear with nodes at $0, 2, ...$.

The difference between S and such a blurred version $\widetilde{S}$ will correspond to the details $d_1[s]$ of the pyramidal algorithm, but such details are viewed this time as an analogic signal $d_1[S]$ on $[0, N]$ (piecewise linear with nodes at $0, 1, ...$). The procedure may be continued : one can project orthogonally $R_1[S]$ on the $\mathbb{C}$-vectorial space (with dimension $2^{q-2}+1$) of piecewise linear signals on $[0, N]$ with nodes at $0, 4, ...$, therefore obtain $R_2[S]$, and form the difference $d_2[S] = R_1[S] - R_2[S]$, which is now a piecewise linear signal on $[0, N]$ with nodes at $0, 2, ...$. Finally, we get

$$S = d_1[S] + d_2[S] + \cdots + d_k[S] + R_k[S]$$

(we can continue as soon as $k < q$), which is an <u>orthogonal</u> decomposition of the analogic signal S ; such an orthogonal decomposition can be also interpreted as a decomposition of s (we used capital letters for analogic signals, ordinary letters for digital signals), which is the digital signal $[s(1), ..., s(N+1)]$ which S interpolates on $[0, N]$ at the points $0, 1, ..., N$.

The same procedure can be carried with Δ replaced by a higher order elementary spline and it is known as the Franklin (or Strömberg) decomposition. It contains in germ the fundamental idea of *Discrete Multiresolution Analysis* that we will introduce next.

6.3 What is a *Discrete Multiresolution Analysis*, with its *father* ?

The concept we will introduce now is due to Stéphane Mallat who introduced it between 1980 and 1985, thus pursuing the renewal of the Haar analysis Yves Meyer proposed introducing the notion of "wavelet".

The fundamental pair of objects to start with is a pair $((V_j)_{j\in\mathbb{Z}}, \varphi)$, where the V_j, $j \in \mathbb{Z}$ are $\mathbb{C}$-linear subspace of the space $L^2(\mathbb{R})$ of analogic signal s with finite energy (that is such that $\int_{\mathbb{R}} |s(t)|^2\, dt < +\infty$) which are embedded one in each other, that is

$$\cdots V_j \subset V_{j-1} \subset \cdots \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots ,$$

together with a well localized element $\varphi \in V_0$ such that :

- V_j, $j \in \mathbb{Z}$, can be deduced from V_0 by

  $$V_j \ := \{s \in L^2(\mathbb{R})\,;\ s(2^j(\cdot)) \in V_0\}$$

  (that means that going from V_0 to V_j means just changing the scale at which informations are read : when $j > 0$, V_j is V_0 scaled with the rate $\tau = 2^j > 1$; when $j < 0$, V_j is V_0 scaled with the rate $\tau = 2^j < 1$) ;

- one has :

  $$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$$
  $$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})\,;$$

- the collection $(\varphi(t-k))_{k\in\mathbb{Z}} = (\varphi_{0,k})_{k\in\mathbb{Z}}$ is an Hilbert basis for V_0.

The function φ is called the *father* of the Discrete Multiresolution Analysis.

Of course, it is not completely evident to produce examples of Discrete Multiresolution Analysis. Nevertheless, there are two simple models corresponding to the simplest Discrete Multiresolution Analysis ; one is the more elementary model that could be imagined respect to time analysis of signal, the second is again the more elementary model that could be imagined respect to the analysis of signals based on the study of their frequency ranges.

The first model is the *Haar model* ; take V_0 as the set of analogic signals with finite energy which are almost everywhere constant on each interval $]k, k+1[$, the function φ being in this case $\varphi = \chi_{[0,1]}$.

The second model is the *Shannon model* : take V_0 as the set of analogic signals s with finite energy such that the spectrum of s, that is the signal (again with finite energy, equal to 2π times the energy of s)

$$\omega \to \int_{\mathbb{R}} s(t) e^{-i\omega t}\, dt$$

has support included in $[-\pi, \pi]$ (that is is almost everywhere equal to zero on the frequency domain $]-\infty, -\pi[\cup]\pi, +\infty[$) ; in this case, the function φ is the function

$$t \to \mathrm{sinc}\,(\pi t) := \frac{\sin(\pi t)}{\pi t}$$

whose Fourier transform is precisely $\chi_{[-\pi,\pi]}$.

To give other examples, it is important to notice that the key "cornerstone" in such a pair $((V_j)_j, \varphi)$ is not really the father φ neither the sequence $(V_j)_j$, but the 2π-periodic function $m_0 \in L^2(\mathbb{R}/\mathbb{Z})$ which satisfies :

$$\widehat{\varphi}(2\omega) = m_0(\omega)\widehat{\varphi}(\omega) \qquad \forall \omega \in \mathbb{R}\ ;$$

there exists such a function m_0 since $\omega \to \widehat{\varphi}(2\omega)$ is the Fourier Transform (in the analogic sense, that is the transform which associates to $s \in L^2(\mathbb{R})$ the limit, in $L^2(\mathbb{R})$, of the sequence of functions $\omega \to \int_{-N}^{N} s(t)e^{-i\omega t}\, dt$) of the function $t \to (1/2)\varphi(t/2)$, which is in V_1, therefore in V_0, and can be expressed as

$$(1/2)\varphi(t/2) = \sum_{k=-\infty}^{+\infty} 2^{-1/2} h(k) \varphi(t-k)\,,$$

with $\sum_k |h(k)|^2 < \infty$; the function m_0 is defined as

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} h(k) e^{-ik\omega}\,.$$

Examples. In the example of the Multiresolution Analysis of Haar,

$$m_0(\omega) = \frac{1 + e^{-i\omega}}{2} ;$$

in the example of the Multiresolution Analysis of Shannon,

$$m_0(\omega) = \chi_{[-\pi/2,\pi/2]}(\omega)$$

(periodized as a function with period 2π).

Once m_0 has been introduced, one can notice that the fact that the functions $t \to \varphi(t-k)$, $k \in \mathbb{Z}$, form an orthonormal system is equivalent to the fact that one has the following identity

$$\sum_{k \in \mathbb{Z}} |\widehat{\varphi}(\omega + 2k\pi)|^2 \equiv 1 ,$$

which implies the relation

$$|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 \equiv 1 \qquad (\dagger\dagger)$$

on the frequency space ; note also that if we make the additional assumption that $\widehat{\varphi}$ is continuous at $\omega = 0$ (which we will do here), then

$$\widehat{\varphi}(\omega) = \widehat{\varphi}(0) \times \lim_{N \to \infty} \prod_{j=1}^{N} m_0(\omega/2^j) , \qquad (\alpha)$$

(just iterating the formula $\widehat{\varphi}(\omega) = m_0(\omega/2)\widehat{\varphi}(\omega/2)$).

Note also that necesseraly $\widehat{\varphi}(\omega) \neq 0$ (otherwise $\widehat{\varphi} \equiv 0$ thanks to (α), which is absurd), then $m_0(0) = 1$ and $m_0(\pi) = 0$.

It was a clever idea due to Ingrid Daubechies to realize that the construction could indeed be reversed, that is a Discrete Multiresolution Analysis constructed from a given 2π periodic function $m_0 \in L^2(\mathbb{R}/\mathbb{Z})$

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} h(k) e^{-ik\omega} ,$$

satisfying

$$|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 \equiv 1 \qquad \forall \omega \in \mathbb{R} ,$$

$m_0(0) = 1$ (that is $\sum_k h(k) = \sqrt{2}$) and the techninal condition :

$$\exists \epsilon > 0 , \ \sum_{k \in \mathbb{Z}} |h(k)| \, |k|^{\epsilon} < +\infty$$

which ensures that the definition of the Riesz infinity product

$$\prod_{j=1}^{\infty} m_0(\omega/2^j)$$

is licit and provides an element in $L^2(\mathbb{R})$. In order to reconstruct the Discrete Multiresolution Analysis from m_0, one reconstructs φ such as

$$\widehat{\varphi}(\omega) = \prod_{j=1}^{\infty} m_0(\omega/2^j) \,,$$

then V_0 (spanned by the $\varphi(t-k)$, $k \in \mathbb{Z}$), then finally the V_j ($s \in V_j$ meaning that $s(2^j(\cdot)) \in V_0$).

What is essential is that for any value of $N > 0$, there is a unique trigonometric polynomial $m_{N,0}$ of the form

$$m_{N,0} = (1 + e^{-i\omega})^N (\sum_{l=0}^{N-1} a_{N,l} e^{-il\omega})$$

which satisfies $m_{N,0}(0) = 1$ and $|m_{N,0}(\omega)|^2 + |m_{N,0}(\omega + \pi)|^2 \equiv 1$ on the whole frequency line. Such a trigonometric polynomial $m_{N,0}$ generates the *Daubechies Discrete Multiresolution Analysis* with order N. We will use in this course the values $N = 1$ (this is the Haar analysis), $N = 2$ (when we will refer to **daub4** and $N = 4$ (when we will refer to **daub8**).

Other Discrete Multiresolution Analysis were introduced by Gilles Lemarié, taking as V_0 the $\mathbb{C}$-vectorial spaces generated by translated of the basic spline function of order p. Of course, here, the situation is somehow more involved since the translates $\varphi(t-k)$, $k \in \mathbb{Z}$, do not form usually an orthonormal system. For example, for 0-splines, with $\varphi(t) = \Delta(t) = \max(0, 1 - |t|)$, it is not true. The function φ (let us take here Δ as an example) has to be slightly changed in order to become a father for the Discrete Multiresolution Analysis. In order to to that, we define $\widetilde{\Delta}$ as

$$\widetilde{\Delta}(t) = \sum_{k \in \mathbb{Z}} a_k \Delta(t-k) \,,$$

where the 2π-periodic function

$$f \ : \omega \to \sum_{k=0}^{\infty} a_k e^{-ik\omega}$$

is such that

$$f(\omega) = \frac{1}{\sqrt{\sum_{k\in\mathbb{Z}} |\widehat{\Delta}(\omega + 2k\pi)|^2}} ;$$

it is therefore immediate to check that

$$\sum_{k\in\mathbb{Z}} |\widehat{\widetilde{\Delta}}(\omega + 2k\pi)|^2 \equiv 1 ,$$

which ensures that the collection $(\widetilde{\Delta}(t-k))_{k\in\mathbb{Z}}$ is an orthonormal system. In this particular case, f can be computed thanks to the Plancherel's formula and we have

$$f(\omega) = \sqrt{\frac{3}{2 + \cos\omega}} ;$$

the corresponding Discrete Multiresolution Analysis is the *Franklin Discrete Multiresolution Analysis.*

Discrete Multiresolution Analysis constructed from vectorial spaces generated by elementary spline functions appear to be of interest when treating for example sismic signals.

7 What to do with a Discrete Multiresolution Analysis ?

7.1 The *mother* of the analysis

Let $((V_j)_j, \varphi)$ a Discrete Multiresolution Analysis as before and m_0 the 2π-periodic function such that

$$\widehat{\varphi}(2\omega) = m_0(\omega)\widehat{\varphi}(\omega) ;$$

as we have seen, one can write

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} h(k) e^{-ik\omega} ,$$

where $(h(k))_k$ is some sequence in $l^2(\mathbb{Z})$.

If we set

$$\phi(t) := \sum_{k\in\mathbb{Z}} h(k)\varphi(t-k)\,,$$

one can see that

$$\phi(t) = \frac{1}{\sqrt{2}}\varphi(t/2)\,,$$

so that the list $(\phi(t-2k))_{k\in\mathbb{Z}}$ corresponds to the list

$$(\varphi_{1,k})_k := (2^{-1/2}\varphi(t/2-k))_{k\in\mathbb{Z}}$$

and therefore to the Hilbert basis for V_1 which is deduced from the Hilbert basis $(\varphi(t-k))_k$ of V_0. Moreover, if

$$s = \sum_{n\in\mathbb{Z}} s(n)\varphi_{0,n} \in V_0\,,$$

the orthogonal projection of s on V_1 can be expressed as

$$\begin{aligned}
\mathrm{Pr}_{V_1}[s] &= \sum_{k\in\mathbb{Z}}\left[\sum_{n\in\mathbb{Z}} s(n)\overline{h(n-2k)}\right]\varphi_{1,k}(t) \\
&= \sum_{k\in\mathbb{Z}}\left[\sum_{n\in\mathbb{Z}} s(n)\overline{h(n-2k)}\right]\phi(t-2k)\,.
\end{aligned}$$

Let now introduce the 2π-periodic function m_1 deduced from m_0 setting

$$m_1(\omega) = e^{-i(\omega+\pi)}\overline{m_0(\omega+\pi)} = \sum_{k\in\mathbb{Z}} \frac{g(k)}{\sqrt{2}}e^{-ik\omega}$$

where

$$g(k) = (-1)^k\overline{h(1-k)}\,.$$

Then, it is easy to check that the matrix

$$M(\omega) := \begin{pmatrix} m_0(\omega) & m_1(\omega) \\ m_0(\omega+\pi) & m_1(\omega+\pi) \end{pmatrix}$$

is a unitary matrix for all values of ω. If we define

$$\theta(t) := \sum_{k\in\mathbb{Z}} g(k)\varphi(t-k)\,,$$

then the collection of functions $(\theta(t-2k))_{k\in\mathbb{Z}}$ is a basis for the orthogonal complement W_1 of V_1 in V_0 ($V_0 = V_1 \oplus^{\perp} W_1$) and we have, for

$$s = \sum_{n\in\mathbb{Z}} s(n)\varphi_{0,n} \in V_0\,,$$

$$\begin{aligned}\mathrm{Pr}_{W_1}[s] &= \sum_{k\in\mathbb{Z}}\left[\sum_{n\in\mathbb{Z}} s(n)\overline{g(n-2k)}\right]\varphi_{1,k}(t)\\ &= \sum_{k\in\mathbb{Z}}\left[\sum_{n\in\mathbb{Z}} s(n)\overline{g(n-2k)}\right]\theta(t-2k)\,.\end{aligned}$$

The function $\psi \in V_{-1}$ which is defined (as φ) by the identity

$$\widehat{\psi}(2\omega) = m_1(\omega)\widehat{\varphi}(\omega)$$

(ψ is such that $\theta(t) = (1/\sqrt{2})\psi(t/2)$) is therefore such that for any $j \in \mathbb{Z}$, the system

$$\left(2^{-j/2}\psi(t/2^j - k)\right)_{k\in\mathbb{Z}}$$

is a Hilbert basis for the orthognal complement W_j of V_j in V_{j-1} ($V_{j-1} = V_j \oplus^{\perp} W_j$), so that, since

$$L^2(\mathbb{R}) = \bigoplus_j^{\perp} W_j$$

has as an Hilbert basis the basis

$$\left(2^{-j/2}\psi(t/2^j - k)\right)_{j,k\in\mathbb{Z}}$$

which is obtained from the "wavelet" ψ (note that ψ has average 0 since $m_1(0) = 0$) dilating, contracting and translating it. Such a function ψ is called the *mother* of the multiresolution analysis and, given $s \in L^2(\mathbb{R})$, the list of coordinates of s in this basis is the list of *wavelet coefficients* of s.

7.2 The wavelet analysis a a discrete signal against a Discrete Multiresolution Analysis

Let $((V_j)_j, \varphi)$ be a Discrete Multiresolution Analysis and $(h(k))_{k\in\mathbb{Z}}$, $(g(k))_{k\in\mathbb{Z}}$ the sequences of coefficients of the two 2π-periodic functions m_0 and m_1

attached to it as described above. The illustrations in this course will concern the case when the Discrete Multiresolution Analysis is the Ingrid Daubechies's one for respectively $N = 2$ (there are 4 parameters for g and h, $h(0), ..., h(3)$, $g(0), ..., g(3)$ to handle all computations) and $N = 4$ (there are 8 parameters for g and h, $h(0), ..., h(7)$, $g(0), ..., g(7)$).

Given a discrete signal $(s(n))_n$ indexed by $\mathbb{Z}$, one makes the *a priori* assumption that s is the sampled version at rate 1 of some analogic signal in V_0, namely the signal

$$S \ : t \to \sum_{n\in\mathbb{Z}} s(n)\varphi(t-n)\,;$$

the father φ of the analysis appears here, but in fact will never be explicited (there is no need for that, at least in this simple classical approach !). The list of wavelet coefficients $c_{1,k}$, $k \in \mathbb{Z}$, is obtained as we have seen before as the list of the coefficients

$$\sum_{n\in\mathbb{Z}} s(n)\overline{g(n-2k)}\,, \qquad k \in \mathbb{Z}\,.$$

At the same time, we compute also the list of coordinates of the orthogonal projection $\mathrm{Pr}_{V_1}[S]$ expressed in the Hilbert basis $(\varphi_{1,k})_{k\in\mathbb{Z}}$; this list corresponds to the discrete signal

$$\left(\sum_{n\in\mathbb{Z}} s(n)\overline{h(n-2k)}\right)_{k\in\mathbb{Z}}\,.$$

Such a signal $(R_1[s](n))_n$ is now treated as $(s(n))_n$ (except that the new scale is now $2 = 2^1$ instead of $1 = 2^0$). In order to compute the complementary list $(c_{2,k})_{k\in\mathbb{Z}}$ of wavelet coefficients of s, we note that this list corresponds to the list

$$\sum_{n\in\mathbb{Z}} R_1[s](n)\overline{g(n-2k)}\,, \qquad k \in \mathbb{Z}\,.$$

And so on... We can continue that way and get the lists $(c_{j,k})$, $j = 1, ..., N$, $k \in \mathbb{Z}$.

In fact, there are two operators from $l^2(\mathbb{Z})$ into itself that play a crucial role :

$$R \quad : \quad (s(n))_n \to \left(\sum_{n\in\mathbb{Z}} s(n)\overline{h(n-2k)}\right)_k$$

$$D \quad : \quad (s(n))_n \to \left(\sum_{n\in\mathbb{Z}} s(n)\overline{g(n-2k)}\right)_k .$$

These operators have to be paired with their adjoints

$$R^* \quad : \quad (u(k))_k \to \left(\sum_{k\in\mathbb{Z}} u(k)\overline{h(n-2k)}\right)_n$$

$$D^* \quad : \quad (u(k))_k \to \left(\sum_{k\in\mathbb{Z}} u(k)\overline{g(n-2k)}\right)_n .$$

These two operators satisfy the Bézout identity

$$R^*R + D^*D = \mathrm{Id}_{l^2(\mathbb{Z})}$$

which controls the mecanism of reconstition. They are called *mirror filters in quadrature.*

One can see the analogy between the algorithms being explicited here and those appearing in the pyramidal algorithm of Burt & Adelson.

Note that the information corresponding to the sequence $c_{1,k}$, $k \in \mathbb{Z}$, has to be thought as an information as an information on $2\mathbb{Z}$, that corresponding to the sequence $c_{2,k}$, $k \in \mathbb{Z}$, as an information on $4\mathbb{Z}$, etc. So, in graphical representation, we will keep track of this fact in order to represent each least as an histogram with width depending of the first index j in $c_{j,k}$.

Since the action of the operators R and D is usually numerically computed in terms of Discrete Fourier Transforms, the input signal $(s(n))_n$ needs in most of the practical algorithms to be a digital signal with lenght a power of two.

The presence of a significative wavelet coefficient at a temporal point $k2^j$ reveals the presence (about this point) of some atom fitting with a copy of some dilated (or contracted version) of the mother wavelet in the decomposition of the signal. Therefore, "accidents" of the signal are here taken into account. Note that the frequency aspect is lost in this decomposition (apart dealing expressely with a multi-resolution adapted to the frequency domain, such as the Shannon decomposition) : the treatment of some discrete information through a Discrete Multiresolution Analysis is a *Time-Scale* analysis, not a *Time-Frequency* analysis as studied before in this course ! It may be a quite interesting complement to a Time Frequency analysis, performed as described in section 4 of these notes.

We will try to correct this later on introducing some more subtil decomposition, namely the *Wavelet Packets Analysis*, in order to bring back the frequential aspect in the picture.

But, prior to do that, we will extend the concept of Discrete Multiresolution Analysis to the problem of treating images.

7.3 The wavelet decomposition of an image

Let $((V_j)_j, \varphi)$ be a Discrete Multiresolution Analysis and ψ be the mother wavelet ; note as $(\varphi_{j,k})_k$ the orthonormal basis of V_j deduced from the orthonormal basis $(\varphi(t-k))_k = (\varphi_{0,k})_k$ through the relations

$$\varphi_{j,k}(t) = 2^{-j/2}\varphi(t/2^j - k)\,, \qquad j \in \mathbb{Z}\,,\ k \in \mathbb{Z}\,.$$

Let $(\psi_{j,k})_k$ be the orthonormal basis of W_j (the orthogonal complement of V_j in V_{j-1}, *i.e.* $V_{j-1} = V_j \oplus^{\perp} W_j$) defined as

$$\psi_{j,k}(t) = 2^{-j/2}\psi(t/2^j - k)\,, \qquad j \in \mathbb{Z}\,,\ k \in \mathbb{Z}\,.$$

For any j in $\mathbb{Z}$, $V_j \times V_j$ is the orthogonal sum :

$$V_j \times V_j = (V_{j+1} \times V_{j+1}) \oplus^{\perp} (W_{j+1} \times V_{j+1}) \oplus^{\perp} (V_{j+1} \times W_{j+1}) \oplus^{\perp} (W_{j+1} \times W_{j+1})\,.$$

Given a digital image I in $V_j \times V_j$, its projection on $V_{j+1} \times V_{j+1}$ corresponds to a resumed version on this image ; its orthogonal projection on $W_{j+1} \times V_{j+1}$ should be an image putting in evidence the "horizontal" details of the image at the scale 2^j, since the first coordinate space is W_{j+1} and the first index in the matrix corresponding to the digital image is the line index ; its orthogonal projection on $V_{j+1} \times W_{j+1}$ should be an image putting in evidence the "vertical" details of the image at the scale 2^j, since the second coordinate space is W_{j+1} and the second index in the matrix corresponding to the digital image is the column index ; finally, its orthogonal projection on $W_{j+1} \times W_{j+1}$ should be an image putting in evidence the "oblic" details of the image at the scale 2^j.

The decomposition algorithm for a given image (with size $2^q \times 2^q$) goes then as follows. The image is thought as the analogic image

$$\mathcal{I} \;:\; (x,y) \rightarrow \sum_{k_1=0}^{2^q-1} \sum_{k_2=0}^{2^q-1} I(k_1,k_2)\varphi_{0,k_1}(x)\varphi_{0,k_2}(y)\,.$$

This analogic image is orthogonally projected of the four subspaces $V_1 \times V_1$, $W_1 \times V_1$, $V_1 \times W_1$ and $W_1 \times W_1$; in fact, one just compute the four digital images corresponding respectively :

- to the list of coordinates of $\mathrm{Pr}_{V_1 \times V_1}(\mathcal{I})$ in the orthonormal basis

$$(\varphi_{1,k_1}(x)\varphi_{1,k_2}(y))_{k_1,k_2}\;;$$

- to the list of coordinates of $\mathrm{Pr}_{W_1 \times V_1}(\mathcal{I})$ in the orthonormal basis

$$(\psi_{1,k_1}(x)\varphi_{1,k_2}(y))_{k_1,k_2}\;;$$

- to the list of coordinates of $\mathrm{Pr}_{V_1 \times W_1}(\mathcal{I})$ in the orthonormal basis

$$(\varphi_{1,k_1}(x)\psi_{1,k_2}(y))_{k_1,k_2}\;;$$

- to the list of coordinates of $\mathrm{Pr}_{W_1 \times W_1}(\mathcal{I})$ in the orthonormal basis

$$(\psi_{1,k_1}(x)\psi_{1,k_2}(y))_{k_1,k_2}\,.$$

Each of these lists corresponds to a $2^{q-1} \times 2^{q-1}$ image and the digital image I is thus decomposed in four digital images, the image corresponding to the resumed version (that is the projection of $\mathcal{I}$ on $V_1 \times V_1$) being at the top left. Such digital images are computed directly using the formulaes in section 7 (for $\mathrm{Pr}_{V_1}[S]$ and $\mathrm{Pr}_{W_1}[S]$), in terms of the lists $(h(k))_{k\in\mathbb{Z}}$ and $(g(k))_{k\in\mathbb{Z}}$ involved in the Fourier expansion of m_0 and m_1.

The procedure can be continued, starting now from the $2^{q-1} \times 2^{q-1}$ image corresponding to the projection of I on $V_1 \times V_1$. This image can be splitted in four $2^{q-2} \times 2^{q-2}$ images corresponding to the splitting of $V_1 \times V_1$ as the orthogonal sum of $V_2 \times V_2$, $W_2 \times V_2$, $V_2 \times W_2$, $W_2 \times W_2$, and so on ... This decomposition algorithm can be tested (with the Daubechies Multiresolution Analysis corresponding $N = 4$) thanks to the command **daub8_2d**.

When we iterate this procedure, we compute in fact the coordinates of the analogic image $\mathcal{I}$ in a new orthonormal basis, which is formed by the functions

$$(\psi_{j,k}(x)\varphi_{j,l}(y))_{k,l}\,, \quad (\varphi_{j,k}(x)\psi_{j,l}(y)_{k,l}\,, \quad (\psi_{j,k}(x)\psi_{j,l}(y)_{k,l}\,, \qquad j=1,2,...$$

These coordinates are the wavelet coefficients of the analogic image $\mathcal{I} \in V_0 \times V_0$ (and by extension the wavelet coefficients of the digital image I to which I has been associated).

Of course, the digital image I can be reconstructed from the resumed version corresponding to the projection on $V_N \times V_N$ and the list of all wavelet coefficients corresponding to values of j between 1 and N. Therefore, some compression (eliminination of non significative coefficients for example) can be performed after the decomposition and a compressed image thus restaured.

Nevertheless, it is important to point out that wavelet coefficients are the coordinates of the image $\mathcal{I}$ (associated to the digital image I) in some very peculiar orthonormal basis. In the next section, we will enrich the set of possible orthonormal basis in which the decomposition is possible (and indicate how to pick up one such that the Shannon entropy of the image is minimal).

It is also noticable also that neither φ nor ψ appear in the iterative computations which lead to the decomposition of the digital image I ; of course, φ is indirectly there since it governs the choice of the analogic image $\mathcal{I}$ which is associated to I and which is in fact the image being decomposed in the process.

7.4 Different basis to decompose a digital signal respect to a multi-resolution analysis

Let again consider a Discrete Multiresolution Analysis $((V_j)_j, \varphi)$, together with the two digital signals $(h(k))_k$ and $(g(k))_k$ which govern the expressions of the 2π-periodic functions m_0 and m_1 associated to the multiresolution analysis.

Let us make here a crucial remark : consider a vectorial space U which is generated by the shifted versions $(\xi(t-k))_k$ of a given function ξ such that the collection $(\xi(t-k))_k$ is an orthonormal basis of U. Such is the case for $U = V_j$ (with a rescaling of $\mathbb{Z}$ and $\xi = \varphi_{j,0}$) but also for $U = W_j$ (still with a rescaling of $\mathbb{Z}$, with $\xi = \psi_{j,0}$). Consider the two functions

$$\begin{aligned} \eta(t) &:= \sum_{k\in\mathbb{Z}} h(k)\xi(t-k) \\ \tau(t) &:= \sum_{k\in\mathbb{Z}} g(k)\xi(t-k)\,. \end{aligned}$$

Then, one can show (under the conditions on m_0 and m_1) that if U_0 denotes the subspace of U which is generated by the functions $t \to \eta(t-2k)$, $k \in \mathbb{Z}$, and U_1 the subspace of U which is generated by the functions $t \to \tau(t-2k)$, $k \in \mathbb{Z}$, then U can be split as

$$U = U_0 \oplus^{\perp} U_1$$

and moreover the collections

$$\{t \to \eta(t-2k)\,;\, k \in \mathbb{Z}\} \quad , \quad \{t \to \tau(t-2k)\,;\, k \in \mathbb{Z}\}$$

are respectively orthonormal basis of U_0 and U_1, so that the procedure puts at our disposal two distinct orthonormal basis $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ of U :

- the original basis $\mathcal{B} := (t \to \xi(t-k))_k$ one was starting with ;
- the basis $\widetilde{\mathcal{B}}$ one obtains by concatening the two orthonormal systems $\{t \to \eta(t-2k)\,;\, k \in \mathbb{Z}\}$ and $\{t \to \tau(t-2k)\,;\, k \in \mathbb{Z}\}$.

Moreover, if

$$s = \sum_k s(k)\xi(t-k)\,,$$

the coordinate of s along $t \to \eta(t-2k)$, $k \in \mathbb{Z}$, is

$$\sum_{n\in\mathbb{Z}} s(n)\overline{h(n-2k)}$$

while the coordinate of s along $t \to \tau(t-2k)$, $k \in \mathbb{Z}$, is

$$\sum_{n\in\mathbb{Z}} s(n)\overline{g(n-2k)}\,.$$

One recovers here the formulaes involved in the search for wavelet coefficients ; nevertheless the point of view is different since U can be taken as V_j as well as W_j (U was just taken as V_j, $j = 0, ...$, in the wavelet decomposition of a digital signal). So, there is here some new idea one should indeed take advantage of.

Let us now explain how can be treated a digital signal with length $N = 2^q$ (continued by 0 outside the interval $\{1, ..., N\}$. One associates to the signal s the analogic signal (which is in V_0)

$$S \ : t \to \sum_{n \in \mathbb{Z}} s(n)\varphi(t-n)\,.$$

The space $U = V_0$ can be decomposed as $U = U_0 \oplus^\perp U_1$ (here in fact $V_0 = V_1 \oplus^\perp W_1$) and one can compute the two lists of coefficients (with length 2^{q-1}) r_k, $k = 1, ..., 2^{q-1}$ and d_k, $k = 1, ..., 2^{q-1}$, corresponding to the coordinates of the orthogonal projections of S respectively on U_0 and U_1 (equipped with their adequate orthonormal basis deduced from the basis $(\varphi(t-k))_k$ of $U = V_0$). Then one compares the Shannon entropies of S

$$\text{Entr}\,(S, \mathcal{B}) := -\sum_{k=1}^{2^q} |s(k)|^2 \log_2 |s(k)|^2$$

and

$$\text{Entr}\,(S, \widetilde{\mathcal{B}}) := -\sum_{k=1}^{2^{q-1}} |r_k|^2 \log_2 |r_k|^2 - \sum_{k=1}^{2^{q-1}} |d_k|^2 \log_2 |d_k|^2\,.$$

Now, we are in either one of the two situations :

- if
$$\text{Entr}\,(S, \mathcal{B}) \leq \text{Entr}\,(S, \widetilde{\mathcal{B}})\,,$$
there is no need to pursue the decomposition procedure and it stops here ; the basis $(\varphi(t-k))_k$ is considered as an optimal one in terms of the entropy criterion (minimizing the entropy). The wavelet decomposition may be interesting for itself, but not respect to minimizing the entropy ;

- if
$$\text{Entr}\,(S, \mathcal{B}) > \text{Entr}\,(S, \widetilde{\mathcal{B}})\,,$$
then the basis $\widetilde{\mathcal{B}}$ is more interesting than the basis $\mathcal{B}$; the decomposition is justified and one may continue, starting now with $U = U_0$ and $U = U_1$ in order to decide whether a splitting of U_0 or U_1 is necessary or not respect to the search for a basis in which the entropy of S is minimal, etc.

The algorithm above, which is basically due to Ronald Coifman and Victor Wickerhauser, is known as the *Split and Merge Algorithm.* It leads to the construction of an orthonormal basis of V_0 in which the Shannon entropy of the analogic signal S corresponding to s is minimal. The basis itself is composed by atoms which look like packets of wavelets (this is why the decomposition itself is called the *Wavelet Packet Decomposition*).

The routines **wpack4** and **wpack8** provide the construction of the decomposition in the optimal basis (with the Daubechies Multiresolution Analysis corresponding respectively to $N = 2$ and $N = 4$).

The procedure may of course been extended to the decomposition of digital images in most convenient orthonormal basis.

7.5 Some applications of the Wavelet Packets decomposition

We just propose here five applications of the *Split and Merge* algorithm ; of course, this list of potential applications is far from being exhaustive !

Compression of data

First, the search for an optimal basis (respect to the problem of the minimization of entropy) is interesting if one wants to compress the digital signal. This is done just by eliminating coordinates of the signal expressed in the optimal basis (such coordinates are given trough the decomposition algorithm) which have an absolute value below a certain *a priori* fixed value. It is of course more intelligent to make the compression once the signal is expressed in the optimal basis ! Then, of course, a compressed version of the signal may be reconstructed from preserved coefficients. Note that in the decomposition process, nor in the reconstruction process, the explicit functions φ and ψ never appear evidently (only the sequences $(h(k))_k$ and $(g(k))_k$ attached to m_0 and m_1 are involved) ; of course, φ is present since it governs the choice of the analogic signal S which is treated.

Classification of data

Given a Discrete Multiresolution Anlysis, the *Split and Merge* algorithm applied on digital signals with length $N = 2^q$ (or digital images with size $2^q \times 2^q$) generates a finite number of possible orthonormal basis.

The search for an optimal basis may be used also for classicication ; different signals may be classified in terms of the particular basis which is the best adapted to them (in the sense, the entropy is minimal) ; one can even refine the classification using different Discrete Multiresolution Analysis. For example, this technique was introduced for the classification of finger prints by the FBI.

Signature of a digital image and watermarking

Putting a signature on an image in order one could authentify it is an interesting challenge in cryptography ; of course, the major difficulty is that the signature needs to be undetectable by standard techniques and robust to filtering, compression, etc., all constraints being in general impossible to fulfill at the same time !

The *Split and Merge* algorithm provides some approach to this problem : the signature is put on a certain component of the decomposition of the image in some peculiar orthonormal basis from the family of basis the *Split and Merge* algorithm provides us. Then, the "key" is the algorithmic decision procedure (for example a comparaison of entropy criterion for a particular choice of entropy –non necesseraly the Shannon entropy– or another kind of criterion to be defined in some algorithmic way) which leads to a specific basis extracted from this family ; the signature is put on certain vectors from this peculiar basis. Of course, in order to detect the signature and then authentify the image, it is necessary to know the key.

Analysis of the correlation between data

In order to illustrate this application, we will focuse on an example of potential application taken from sismology.

Let $s = (s_1, s_2, s_3)$ be a vector of 3 digital signals corresponding to the registration in three directions (West-East, South-North, vertical) of a sismic event.

It happens to be quite important in sismology to be able to detect in such a sismic wave the components that propagate horizontally (the P-waves) and the ones that propagate vertically (the S-waves) and know in particular the instant of arrival of these waves (there may be successive ones, of course).

In order to do that, one can construct a digital signal, which is the *polariza-*

tion function of the sismic wave ; to the instant t, one associate the quantity

$$f(t) = 1 - \frac{\lambda_2[s](t)}{\lambda_1[s](t)}$$

where $\lambda_1[s](t) \geq \lambda_2[s](t) \geq \lambda_3[s](t)$ are the 3 eigenvalues of the positive symmetric matrix $M(t)$ corresponding to the instantaneous correlation matrix of (s_1, s_2, s_3) near the instant t. A peak of this function should materialize the instant of arrival of a P-wave. The points where $f(t)$ achieves a maximum correspond to the instants where the sismic wave is polarized and thus mark the instants of arrival of P-waves.

Instead of using f (and in order to reinforce the information about polarization), one can consider

$$\tilde{f}(t) = \prod_j \left(1 - \frac{\lambda_2[\sigma_j](t)}{\lambda_1[\sigma_j](t)}\right)$$

where the $\sigma_j = (\sigma_{j,1}, ..., \sigma_{j,3})$ are such that $\sigma_{j,1}, \sigma_{j,2}, \sigma_{j,3}$ are components of respectively of s_1, s_2, s_3 in the optimal basis (respect to some Multiresolution Analysis and entropy criterion) in which the three signals can be decomposed when these components correspond to the same label in the *Split and Merge* decomposition tree and analyze the local peaks of this function.

8 Continuous time-scale analysis

8.1 The *Continuous Wavelet Tranform* (CWT)

Orthogonal decomposition of informations may have some inherent defects : it makes fragile the treatment of the information through its decomposition (changing some component in the decomposition may change drastically the whole information). It happens in many situations that it is quite important to profit from the redundancy of the information, thus to obtain more robust decompositions ready for the treatment of the information (compression, separation signal-noise, watermarking,...) This is the reason why we introduce here a continuous transform, the CWT-Transform. The defect of such a transform (on the opposite of what happens to transforms leading to orthogonal decompositions such as in section 7) is the fact that it provides too much information at the same time, without any organization method to treat it ! We will try to correct that in section 9 when introducing the *Matching Pursuit* algorithm.

The basic "atom" involved in such a redundant decomposition will be some "wavelet" ψ with finite energy such that $\widehat{\psi}$ vanishes at $\omega = 0$, more precisely

$$\int_{\mathbb{R}} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} d\omega < +\infty \,.$$

The fact that the Fourier Transform of the atom ψ vanishes at $\omega = 0$ indicates that convolution with ψ can be interpreted as pass-band filtering (low frequencies of signals are cut through convolution with ψ). Of course, the more $\widehat{\psi}$ vanishes at $\omega = 0$, the more this pass-band filtering property in enhanced ; when ψ is in the Schwartz space $\mathcal{S}(\mathbb{R})$ (the space of of C^∞ functions which decrease to zero at infinity faster than any $|t|^{-N}$ for any N, as well as all their derivatives), the fact that $\widehat{\psi}$ vanishes at the order $q+1 \in \mathbb{N}^*$ at $\omega = 0$ (that is is such that all derivatives of $\widehat{\psi}$ or order less than q are zero at $\omega = 0$) is equivalent to the fact that ψ is orthogonal to polynomial functions with degree less than q. Such is the case for example when ψ is the $q+1$-th derivative of the Gaussian function $t \to (2\pi)^{-1/2} \exp(-t^2/2)$. We will use models of such kind later on. The ideal situation (respect to this pass-band filtering property) takes place when the Fourier transform of the wavelet is identically zero near $\omega = 0$; such is the *Gabor wavelet* (which is the fundamental tool in *Gabor analysis*) where

$$\widehat{\psi}(\omega) = \exp(-(\omega - \omega_0)^2/2) \,,$$

with $\omega_0 := 5.33644$ being chosen such the ratio between the two first local maxima of ψ on $[0, +\infty[$ equals $1/2$; note that the Gabor wavelet is the modulated gaussian

$$t \to \frac{1}{\sqrt{2\pi}} \exp(-t^2/2 + i\omega_0 t) \,.$$

The Fourier transform of the Gabor wavelet being not exactly zero at $\omega = 0$, it was corrected by J. Morlet who introduced the *Morlet wavelet* with Fourier transform

$$\omega \to \exp(-(\omega - \omega_0)^2/2) - \exp(-\omega_0^2/2) \exp(-\omega^2/2) \,.$$

The *Continuous Wavelet Transform* corresponding to any such wavelet ψ is the map which transforms an analogic signal $s \in L^2(\mathbb{R})$ into the analogic image

$$(a,b) \in]0,+\infty[\times\mathbb{R} \to \mathrm{CWT}_\psi\,[s;a,b] := \frac{1}{\sqrt{a}}\int_{\mathbb{R}} s(t)\overline{\psi((t-b)/a)}\,dt\,.$$

Note that the atom

$$t \to \frac{1}{\sqrt{a}}\psi((t-b)/a)$$

is a contracted or dilated version of the "wavelet" ψ which is translated from $t=0$ to the point $t=b$ (it is also normalized in order to keep the same energy). So that the *Continuous Wavelet Transform* appears as the test for the correlation of the information against a "dictionary" built from a reference wavelet which is either contracted or dilated, then translated, on the whole time interval. It follows from Plancherel's formula that

$$\mathrm{CWT}_\psi\,[s;a,b] = \frac{\sqrt{a}}{2\pi}\int_{\mathbb{R}} \widehat{s}(\omega)\overline{\widehat{\psi}(a\omega)}\,e^{ib\omega}\,d\omega\,,$$

which means that numerical computations are made thanks to DFT algorithm. The **MATLAB** routines **cwt** and **gaussq** (when the wavelet is the q-derivative of a the standard gaussian) provide the numerical computation starting with a digital signal s with length $N=2^q$. The scale axis is equipped with a logarithmic scaling : there are $q-1$ dyadic octaves (scale between 2^{-2} and 2^{-1},..., scale between 2^{-q} and 2^{-q+1}) which are each divided in m "voices" (note the analogy with the vocabular used in music coding) ; we assume here that the digital signal is rated at the scale 2^{-q} (one could as well have decided to fix the smaller scale at 1).

The Continuous Wavelet Transformation can be easily inverted, provided the Fourier transform of the wavelet satisfies

$$\int_0^{+\infty}\frac{|\widehat{\psi}(\omega)|^2}{|\omega|}d\omega = \int_0^{+\infty}\frac{|\widehat{\psi}(-\omega)|^2}{|\omega|}d\omega = C(\psi) < +\infty\,.$$

The inversion formula is

$$s = \frac{1}{C(\psi)}\lim_{\substack{L^2\\ \epsilon\to 0^+,A,T\to+\infty}}\int_{a\in[\epsilon,A]}\int_{b\in[-T,T]}\mathrm{CWT}\,[s;a,b]\frac{\psi((\cdot-b)/a)}{a^2}\,\frac{dadb}{\sqrt{a}}\ ;$$

this follows from the inversion formula for the Fourier Transform from $L^2(\mathbb{R})$ into itself which is given as

$$s = \frac{1}{2\pi} \lim_{\substack{L^2 \\ T \to +\infty}} \int_{-T}^{T} \widehat{s}(\omega) e^{i\omega(\cdot)} \, d\omega \,,$$

the Fourier Transform $\widehat{s}$ being defined as

$$\widehat{s} = \lim_{\substack{L^2 \\ T \to +\infty}} \int_{-T}^{T} s(t) e^{-i\omega(\cdot)} \, dt \,.$$

So the inverse CWT is as easy to realize as the transform itself. The **MATLAB** routine which does the job is **icwtg**. Be careful that the inversion formula does not hold for digital signals which are constant (such a signal does not have a finite energy !) so it is secure to take the mean value of the digital signal equal to zero in order to apply the *Continuous Wavelet Transform* and its inversion (in fact, the inversion formula holds numerically when tested on the space of digital signals with mean value equal to zero).

The fact that such a redundant transform can be easily inverted may be of interest respect to compression of data or steganography (we will develop applied examples in these directions in the course). Since the transformation is highly redundant, a modification of the signal through a modification of its *Continuous Wavelet Transform* resists to the treatment of the signal by standard techniques (compression, filtering, *etc.*) and thus may be of interest. Note also that there is much more freedom in chosing the wavelet than it happens to be for the construction of a Discrete Multiresolution Analysis. Nevertheless, it is interesting to compare on examples the results of Discrete (thus orthogonal) and Continuous (thus redundant) wavelet decompositions.

We will brievely describe in the subsequent sub-sections three important applications of Continous Wavelet Transform.

8.2 Multifractal analysis

Singularities of an analogic signal can be "classified" thanks to the local Lipschitz exponent ; such an exponent is defined as the largest positive real number $\gamma(t_0)$ such that, for t_1, t_2 in some infinitesimal neighborhood of t_0,

$$|f(t_1) - f(t_2)| \leq K(t_0)|t_1 - t_2|^{\gamma(t_0)} \,.$$

If q is an integer and $\alpha \in [0,1[$, the point t_0 is called a point of $q+\alpha$ regularity if s is q-times differentiable at t_0 and the local Lipschitz exponent is $q+\alpha$ at t_0.

Some information respect to the classification of real points in terms of their regularity respect to some anlogic signal s can be derived from the analysis of the positive images

$$(a,b) \to |\mathrm{CWT}_\psi [s;a,b]|$$

where ψ_{q+1}, $q = 0,1,...$, is the $q+1$-derivative of a Gaussian. Since ψ_{q+1} is orthogonal to polynomial functions of degree less than q (that is to the Taylor polynomial of degree q of s if s is q-times differentiable at t_0), it follows that the *Continuous Wavelet Transform* isolates and enhances the error term in Taylor-Young formula at $t = t_0$, so that, if there exists an interval $I(t_0)$ containing t_0 and some $\epsilon > 0$ such that, for any $a \in]0,\epsilon[$, the signal

$$b \to |\mathrm{CWT}_{\psi_{q+1}} [s;a,b]|$$

do not present any local maximum in I, then the point t_0 is a point of β-regularity with $\beta \geq m$, which means that the singularity at t_0 is a gentle one (the more q for which this happens is large, the more "gentle" the singularity is, so that when q increases, there are less and less singularities to be seen).

Moreover, if there is a conic sector $\{a \in]0,\epsilon[\,,\ |b-t_0| < C(t_0)a\}$ such that all points where horizontal sections of the image $(a,b) \to |\mathrm{CWT}_\psi [s;a,b]|$ admit a local maximum remain in this cone when they are close to $(0,b)$, then all points near t_0 (except t_0 itself) are β-regular points, with $\beta \geq m$. The point t_0 itself is a point of γ-regularity for some $\gamma \in [0,+m[$ if

$$|\mathrm{CWT}_\psi [s;a,b]| \leq K(t_0)a^{\gamma+1/2}$$

along all descending curves connecting points which are in the cone and where local maxima of horizontal sections are achieved (for some uniform constant $K(t_0) > 0$).

Therefore, the study of maxima of horizontal sections of the image

$$(a,b) \to |\mathrm{CWT}_\psi [s;a,b]|$$

(and the behavior of "descending" curves connecting the points where such maxima are achieved) is an important tool for the classification of singularities (and what is called the multi-fractal analysis). It has been extensively developped by mathematicians (S. Jaffard and Y. Meyer) as well as by physicists (A. Arnéodo). It is also an important tool to analyse the coherence of an analogic signal in terms of the scale level at which it is analysed ; this will happen to be also an interesting tool for noise extraction keeping track of the singularities of the signal (which is usually something quite hard to do !).

8.3 Separation signal-noise

Any deterministic analogic signal obeys to some deterministic coherence which reflects either in time-frequency analysis when the signal is stationary or at least piecewise stationary or in time-scale analysis when the signal has some fractal or multifractal structure. It is usually paired (when transmitted or even measured) with a stochastic error (a *noise*) which obeys to some stochastic coherence.

Being able to separate noise and signal without eliminating significative components of the signal (corresponding either to high-frequency components or to details referent to a small scale) is a delicate challenge. The standard technique of filtering (which consists in the action on the signal of a low-pass filter) "cuts" high frequency components and therefore the noise ; but it may also eliminate significative components of the signal itself !

Let us explain here shortly how *Continuous Wavelet Transform* may help to such separation, assuming for the sake of simplicity that the noise is a white noise, that is a stochastic process with mean 0 and autocorrelation function $(t,s) \to E[X_t \overline{X_s}] = \sigma^2 \delta(t-s)$, where δ is the Dirac mass at the origin ; of course, we will deal only with digital signals, so that our process will be a discrete process $(X_n)_{n\in\mathbb{Z}}$, all alea X_n being in the same $L^2(\Omega, \mathcal{T}, P)$, where $(\Omega, \mathcal{T}, P)$ is the reference probability space.

If s is a deterministic signal in $L^2(\mathbb{R})$ and $B = (B_t)_{t\in\mathbb{R}}$ a white noise with variance σ^2 (as described above), then one has, for any $a > 0$ and $b \in \mathbb{R}$ (given some wavelet ψ) :

$$E[|\mathrm{CWT}_\psi[s(t) + B_t; a, b]|^2] = |\mathrm{CWT}_\psi[s; a, b]|^2 + \frac{\sigma^2}{a}\|\psi\|_2^2 \, ;$$

this explains why the presence of the additional noise (above some deterministic signal $s \in L^2(\mathbb{R})$) is responsible for some "explosion" in $1/a$ in the brilliance of the positive image corresponding to the modulus of the *Continuous Wavelet Transform* when one goes towards small scales ($a \to 0$).

Moreover, the average number of local extrema for the horizontal section $b \to \mathrm{CWT}_\psi[B_t; a, b]$ is proportional to $1/a$, which means (statistically) that one local extremum over two is lost going up from the dyadic scale level $a = 2^j$ to the dyadic scale level $a = 2^{j+1}$.

Taking into account these two remarks (how the stochastic coherence of the noise behaves respect to time-scale analysis), one may propose some methods for separation signal-noise.

Here is a very naïve one : if 2^j is a scale level, one may decide that the extrapolation at b_0 (using for example linear regression, which is easy to realize with **MATLAB** and for example the routine **polyfit**) of the function $b \to \mathrm{CWT}_\psi[s + B; a, b](2^j, b)$ from the average slope of

$$a(\theta) \to \log |\mathrm{CWT}_\psi[s + B; a(\theta), b(\theta)]|$$

along "crete" curves in the image $|\mathrm{CWT}_\psi[s + B; \cdot, \cdot]|$ (lying in a conic sector above the point $(0, b_0)$ between the scales 2^j and 2^{j+1}) is a reasonable candidate for a predicted value of $\mathrm{CWT}_\psi[s + B; 2^{j-1}, b_0]$. One may decide to affect this value at the point $(2^{j-1}, b_0)$ and to take into account (at the scale level 2^{j-1}) only points b_0 which correspond to a local maximum of $b \to |\mathrm{CWT}_\psi[B_t; 2^j, b]|$. Thus, the "blurred" level $a = 2^{j-1}$ (blurred because of the presence of the noise) is replaced by a new "predicted" level which should be closer to what should be the scale level $a = 2^{j-1}$ if the noise was absent. The routines **gauss2q** and **gauss3q** realize this quite naïve (though interesting) approach.

We will mention some less naïve approach at the end of the next subsection.

8.4 Reconstructing s from local extrema of the Continuous Wavelet Transform

Let s be a real signal and ψ a real wavelet. It may be of interest to notice that, just keeping for each value of $a > 0$ the local extrema of $b \to \mathrm{CWT}_\psi[s; a, b]$ and interpolating them linearly (this of course may correspond to some significative compression of the information), one may reconstruct (thanks to the *Inverve Continuous Wavelet Transform*) a quite reasonable approximation of the signal.

Such a remark was pointed out by S. Mallat and may be helpful, though one may show that, mathematically speaking, the reconstitution of $s \in L^2(\mathbb{R})$

from the local "horizontal" extrema of its Continuous Wavelet Transform is impossible (as shown by Y. Meyer). The method is thus more empirical than theoretical.

It may also be used as follows : once the local extrema of $b \to \mathrm{CWT}_\psi[s; 2^j, b]$ have been detected (position x_{jk}, value s_{jk}, $k \in \mathbb{Z}$) one looks for the signal $\tilde{s}$ in $L^2(\mathbb{R})$ such that, for each dyadic level 2^j (where the local extrema of $b \to \mathrm{CWT}_\psi[s; a, b]$ are (x_{jk})), the H^1-norm

$$\sum_k \left|\frac{d}{db}(\mathrm{CWT}_\psi[\tilde{s}; 2^j, x_{jk}])\right|^2 + |\mathrm{CWT}_\psi[\tilde{s}; 2^j, x_{jk}] - s_{jk}|^2$$

is minimal. Once this is done, s is reconstructed by the iterated projection algorithm assuming its orthogonal projection on the vectorial subspace spanned by the functions

$$b \to 2^{-j/2}\psi((t-b)/2^j) = \psi_{j,b}$$

is

$$\sum_{k \in \mathbb{Z}} \left\langle \tilde{s}\,,\ \psi_{j,x_{jk}} \right\rangle \psi_{j,x_{jk}}\,.$$

Such method (introduced by Mallat and Zhang) can be paired with the method proposed in the last subsection for the separation signal-noise.

9 Matching Pursuit algorithms and referent ideas inspired by statistics

9.1 *Matching Pursuit* algorithm

As we pointed it out, the disadvantage of redundant decompositions such as the ones which comes from CWT is that they provide all together (in fact through just one image) too much informations ; there is no efficient tool which is provided to "read" in some intelligent way these informations, as for example the *Split and Merge* algorithm presented in section 7.

A parade to such a disadvantage consists in using instead of such transforms the idea based of the fact one tries to "pursue" some given information against a "dictionary". Such "matching" techniques are quite familiar to computer scientists or statisticians. They provide naïve, but quite robust, algorithms which can be used as soon as one has some *a priori* idea about the "atoms" that should appear in the decomposition of a signal or an image. Good examples are provided by medical imaging, where one is often confronted to the problem of pursuing an image (eventually pathologic) along a dictionary of pathological test images.

We present here the idea of the *Matching Pursuit* algorithm (which can be implemented for example with the dictionary of atoms $(a^{-1/2}\psi((t-b)/a))_{a>0,b\in\mathbb{R}}$ involved in the definition of the *Continuous Wavelet Transform* associated to a wavelet ψ), as being renewed by S. Mallat (it is certainly a very old idea, inspired by statistics, as we will see later on).

The idea is very simple : suppose that $\mathcal{D}$ is a dense subset of $L^2(\mathbb{R})$ (one can replace $L^2(\mathbb{R})$ by some Hilbert space) whose elements d have energy equal to one, that $s \in L^2(\mathbb{R})$, and that $(d_k)_{k\in\mathbb{N}^*}$ is a sequence of atoms in $\mathcal{D}$ defined inductively (together with the sequence of coefficients $(\alpha_k)_{k\in\mathbb{N}}$) as :

$$\begin{aligned}
|\langle s\,,\,d_1\rangle| &= \max_{d\in\mathcal{D}} |\langle s\,,\,d\rangle| \\
\alpha_1 &= \langle s\,,\,d_1\rangle \\
\forall n>1\,, \left|\left\langle s-\sum_{k=1}^{n}\alpha_k d_k\,,\,d_{n+1}\right\rangle\right| &= \max_{d\in\mathcal{D}}\left|\left\langle s-\sum_{k=1}^{n}\alpha_k d_k, d\right\rangle\right| \\
\alpha_{n+1} &= \left\langle s-\sum_{k=1}^{n}\alpha_k d_k\,,\,d_{k+1}\right\rangle;
\end{aligned}$$

then the sequence $(s_n)_{n\in\mathbb{N}^*}$, where

$$s_n := \sum_{k=1}^{n}\alpha_k d_k\,,$$

converges towards s in $L^2(\mathbb{R})$. We refer to the work of S. Mallat and Z. Zhang, [*Matching pursuit algorithms with time frequency dictionaries*, IEEE Transactions on Signal Processing 12, 1993, pp. 3397-3415], for a proof of this result (involving basically the completeness of $L^2(\mathbb{R})$ and the density of the dictionary).

Such an algorithm is implemented under the routine **mpurs1** under the **MATLAB** environment.

In order not to use repeatedly the same atom (note that in the above procedure, the signal has to be explored with the whole dictionary ay any step of the algorithmic procedure), one may introduce an "orthogonalized" version of the *Matching Pursuit Algorithm*, as proposed for example by Patti and Krischnaprasad. This algorithm, which is described here in eight steps, is implemented as the **MATLAB** routine **mpurs2**.

- 1. *Initialization at step n.* One starts with a "resumed" version R_n of the input signal s, which has been obtained *via* n iterations of the algorithm :
$$R_n = \sum_{k=1}^{n} a_k^{(n)} d_k \,.$$
- 2. *Computation of the Gram matrix G_n of the first n atoms slected :* this computation will be done recursively, as we will see next.
- 3. *Detection of the $n+1$-th atom d_{n+1} following the principle of the Matching Pursuit algorithm :* if $r_n = s - R_n$,
$$|\langle r_n, d_{n+1}\rangle| = \max_{\substack{d \in \mathbf{D} \\ d \neq d_1,\ldots,d_n}} |\langle r_n, d\rangle| \,.$$
- 4. *Computation of correlations of the new selected atom d_{n+1} with atoms $d_1, \ldots, d_n$ which have been selected before :* such a vector is denoted as
$$C_n := \begin{pmatrix} \langle d_{n+1}, d_1\rangle \\ \vdots \\ \langle d_{n+1}, d_n\rangle \end{pmatrix} .$$
- 5. *Computation of $B_n = G_n^{-1} C_n$.*
- 6. *Computation of the coefficient which affects this new atom :* such a computation is done *via* the formula
$$a_{n+1}^{(n+1)} = \frac{\langle r_n, d_{n+1}\rangle}{1 - \sum\limits_{k=1}^{n} B_n(k)\langle d_k, d_{n+1}\rangle} \;;$$
note that it may also be interesting to keep track of the value $\langle r_n, d_{n+1}\rangle$ which corresponds to the coefficient of d_{n+1} prior the re-orthonormalization of the system.

- 7. *Final computation of the resumed version* R_{n+1} *prior applying the next step of the algorithm :* one has the formula

$$R_{n+1} = \sum_{k=1}^{n} (a_k^{(n)} - a_{n+1}^{(n+1)} B_n(k)) d_k + a_{n+1}^{(n+1)} d_{n+1} \,.$$

- 8. *One starts the procedure at step* 1*, starting with* R_{n+1}.

Note that G_n^{-1} may be computed in some recursive way, as

$$G_{n+1}^{-1} = \begin{bmatrix} G_n^{-1} + \rho_n B_n B_n^* & -\rho_n B_n \\ -\rho_n B_n^* & \rho_n \end{bmatrix} ,$$

where

$$\rho_n := \frac{1}{1 - C_n^* B_n} \,.$$

All such operations realize a combination between the standard *Matching Pursuit Algorithm* and the Gram-Schmidt orthonormalization procedure ; the convergence of the sequence $(r_n)_n$ towards zero is still fulfilled in the dictionary $\mathcal{D}$ is dense in $L^2(\mathbb{R})$.

As mentioned above, such algorithms are good canditates for treating physiological signals ; they may be also combined with statistical technics (such as *Neural Networks* or *Proper Orthogonal Decomposition* which are inspired by statistics) which help to the construction of adequate dictionaries (see the next subsection).

Respect to time-scale or time-frequency analysis, the most standard dictionaries that can be used (such as proposed by David Donoho in the logicial **Wavelab** which has been developped at Stanford University) are the dictionary of atoms $(a^{-1/2}\psi((t-b)/a))_{a>0, b\in\mathbb{R}}$, where ψ is an atom which generates a CWT analysis, the collection $(2^{-j/2}\psi(t/2^j - n))$, $j > 0$, $n, j \in \mathbb{Z}$, where ψ denotes the mother of a *Multiresolution Analysis*, the collection of local harmonics

$$\Big\{ t \to \sqrt{\frac{2}{b-a}} \sin \frac{\pi(k+1/2)(t-a)}{b-a} \,, \quad k \in \mathbb{N} \,,\ b > a \Big\} \,,$$

or even a discrete orthonormal subfamily constructed from a segmentation of the time line $\mathbb{R}$, $\cdots a_j < a_{j+1} < \cdots$, multiplying each function

$$t \to \sqrt{\frac{2}{a_{j+1} - a_j}} \sin \frac{\pi(k+1/2)(t-a_j)}{a_{j+1} - a_j} \,, j \in \mathbb{Z},\ k \in \mathbb{N}$$

by a smooth cut-off function θ_j localized in $[(a_{j-1}+a_j)/2, (a_{j+1}+a_{j+2})/2]$ (such as proposed by R. Coifman and Y. Meyer in *Remarques sur l'analyse de Fourier à fenêtre*, Comptes Rendus Acad. Sc. Paris, 312, Série I, 1991, 259-261.

9.2 The *Proper Orthogonal Decomposition*

To conclude this course (and to illustrate *Matching Pursuit Algorithm*), we present here some important idea (inspired by statistics) in order to construct a dictionary ready for the matching pursuit test (and adapted to some given information).

Let s be a digital information (indexed by $\mathbb{Z}$) corresponding to the measurement of some phenomenon. Suppose that we have at our disposal, for some temporal interval with length M, N realizations $s^{(1)},...,s^{(N)}$ of s (for example N samples of s which correspond to distinct temporal intervals $I_1,...,I_N$, all with lenght s) ; such realizations are sometimes called "snapshots" of the signal s. The correlation matrix $[R_{ij}]_{1\leq i,j\leq N}$, where

$$R_{ij} = \frac{1}{M}\sum_{m=1}^{M} s^{(i)}(m)\overline{s^{(j)}(m)}\,, \quad 1\leq i,j\leq N$$

induces some hermitian operator on $l^2(\{1,...,N\})$:

$$\mathcal{R}\,:\Phi = (\Phi(1),....,\Phi(N)) \mapsto \left(\sum_{j=1}^{N} R_{1j}\Phi(j)\,,\,...\,,\,\sum_{j=1}^{N} R_{Nj}\Phi(j)\right);$$

the eigenvalues of this positive operator can be organized in decreasing order : $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_N$. A unitary vector $\Phi_{\rm pp}=(\Phi_{\rm pp}(1),...,\Phi_{\rm pp}(N))\in l^2(\{1,...,N\})$ such that $\langle\mathcal{R}\Phi_{\rm pp}\,,\,\Phi_{\rm pp}\rangle$ is maximal (that is a unitary eigenvector $\mathcal{R}$ associated to λ_1) provides a principal direction for the subspace of $l^2(\{1,...,M\})$ generated by the snapshots $s^{(1)},...,s^{(N)}$, and the realization corresponding to this direction (which provides the best statistical least square approximation of the signal s on a digital temporal window with length M) is

$$S_1 = \sum_{j=1}^{N} \Phi_{\rm pp}(j)s^{(j)}\,.$$

One can thus construct a collection $S_1, S_2, ...,$ from the eigenvalues $\lambda_1, ...,$ (and corresponding normalized eigenvectors $\Phi_{pp,j}$). Such digitals signals $S_1, S_2, ...,$

are called *Proper Modes* of s (on temporal windows with lenght M) and (once translated) can be used as the list of atoms of a dictionary ready for the *Matching Pursuit* algorithm.

Such an idea, that we exposed here very brievely, is the basis for the *Proper Orthogonal Decomposition*, which is quite performing respect to matching problems in statistics.

10 Conclusion (so many things missing !)

We just pretended in this course propose a panel of tools inspired by Fourier analysis, Wavelet analysis, matching ideas coming from statistics, in order to treat digital informations. Of course, such a panel is far from beeing exhaustive ! We did not had time to speak for example of combination of entropy criterion and using of local harmonics to treat speech signals (as proposed by E. Wesfreid and V. Wickerhauser in *Adapted Local Trigonometric Transforms and Speech Processing*, IEEE Trans. on Signal Processing, 41 (12), 1993, 3596-3600, and there are so many other important technics we did not mention, which combine the diverse points of view presented here.