#### 1. Introduction au Traitement du Signal

#### 1.1 Qu'est-ce qu'un signal?

Par définition, un signal est le *support physique d'une information*. Il s'agit donc d'une notion tout à fait générale que l'on peut rencontrer dans des domaines aussi variés que l'électricité, l'électronique, l'acoustique, l'optique, la mécanique, l'astronomie, la biologie, l'économie, etc... En fait, il y a signal dès qu'il y a mesure et/ou transmission d'information d'une source vers un destinataire.

Traiter un signal, c'est essentiellement en extraire l'information que l'on juge utile, la mettre en forme pour mieux l'analyser, la transmettre ou la stocker, la nettoyer de parasites éventuels.

# 1.2 Place de la discipline

De par ses objectifs, le traitement du signal est clairement à une situation carrefour. Il est en effet en contact étroit avec ses champs d'application (comme la physique au sens large, les communications, la biologie, etc...) dans lesquels il puise ses motivations et pour lesquels il élabore la spécificité de ses solutions. Il repose par ailleurs sur des outils mathématiques (comme l'analyse harmonique, l'algèbre linéaire ou les statistiques) qui lui sont essentiels. Il est enfin un passage obligé pour des tâches comme la classification ou la reconnaissance de formes.

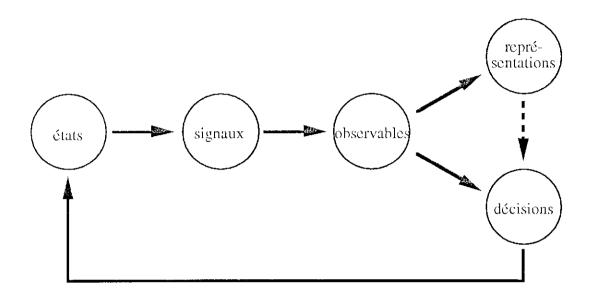
Le traitement du signal est cependant une discipline autonome, son originalité reposant d'une part sur l'universalité de son langage et d'autre part sur l'enrichissement qu'elle offre en retour aux disciplines qui la nourrissent (développement d'algorithmes en mathématiques appliquées, affinement de mesures en physique, etc...).

#### 1.3 Structure d'une chaîne de traitement

Le signal matérialise le voyage d'une information. On peut alors distinguer un certain nombre d'opérations ou de situations qui sont communes à tout chaîne de traitement. On peut en tout premier lieu distinguer les opérations

- de *pré*-traitement, qui sont essentiellement relatives à la prise d'information elle-même, aux capteurs et à leurs propriétés;
- de *traitement* proprement dit, qui constituent le cœur de ce que l'on appelle le traitement du signal;
- de *post*-traitement, qui incluent davantage des actions (on parle alors plutôt d'automatique) ou des méthodes symboliques ou de contexte (vision, intelligence artificielle, etc...).

D'une manière assez générale, on peut résumer une chaîne de traitement du signal par le schéma suivant



Structure générale d'une chaîne de traitement

Ce schéma met en jeu plusieurs espaces et des applications entre ces espaces. Les espaces sont ceux

- des *états* : c'est là que se trouve l'information utile.
- des signaux : ce sont eux qui matérialisent les états possibles.
- des *observables* : ce sont les seules grandeurs disponibles pour l'utilisateur.
- des *décisions* : c'est là que se matérialisent les opérations relatives aux signaux.
- des *représentations* : elles permettent de changer la façon dont on "regarde" un signal et peuvent soit aider à l'analyse seule, soit constituer un détour utile pour construire une décision.

Les liens qui unissent ces différents espaces correspondent à des opérations de

- codage entre états et signaux.
- transmission (au sens large) entre signaux et observables : on parle de canal.
- *règles de décision* entre observables et décisions (l'espace des décisions étant supposé *isomorphe* à l'espace des états).

On peut donc formuler une problématique générale du traitement du signal de la façon suivante :

Il s'agit d'élaborer la *meilleure* règle de décision, c'est-à-dire celle qui, à partir des seules données observables, permet de caractériser au mieux une information utile véhiculée par un signal.

Il est clair qu'un tel objectif nécessite de préciser ce que l'on entend par "meilleure" et par "au mieux". Il faudra donc définir des critères qu'il s'agira d'optimiser. Par ailleurs, l'objectif ne pourra être atteint que si un certain nombre d'informations (a priori) supplémentaires sont disponibles. Celles-ci pourront concerner

- la nature ou les états possibles d'un système.
- le type de codage.
- des informations relatives au canal de transmission et au bruit d'observation.
- ...

Ceci étant supposé, le schéma décrit ci-dessus permet d'englober plusieurs des tâches essentielles que l'on rencontre en traitement du signal :

- la *détection*: l'espace des états est alors discret (voire binaire) et le problème posé est de savoir si une observation donnée correspond à un signal utile mélangé à du bruit ou à du bruit seul.
- l'analyse et l'estimation (de paramètres) : il s'agit cette fois, et toujours à partir d'une observation imparfaite, de caractériser un signal par un jeu de paramètres qui le définissent.
- le *filtrage* : il s'agit alors d'estimer un signal dans son ensemble et non par l'intermédiaire d'un petit nombre de paramètres caractéristiques.

Ces tâches élémentaires apparaissent souvent couplées, les problèmes réels faisant fréquemment intervenir simultanément des opérations de filtrage et/ou de détectionestimation (par exemple en poursuite).

#### 1.4 Exemples

Les exemples d'applications pour lesquelles le traitement du signal est prépondérant abondent. On peut citer entre autres et en vrac :

- le traitement de la *parole* : analyse, synthèse, codage, transmission, reconnaissance, commande vocale.
  - Faudio (compact disc) et la video (TVHD) numériques.
  - les télécommunications: téléphone, radio-mobile, video-conférence.
  - le biomédical : ECG, EEG, EMG, échographie, imagerie RMN.
  - le radar et le sonar (passif ou actif).
  - le *contrôle non-destructif* (par ultrasons ou par courants de Foucault).
- l'astronomie : imagerie optique et radio, interférométrie des tavelures (speckle), optique adaptative.
  - la *physique* : turbulence, géophysique (interne et externe), sismique.
  - l'automobile et les industries mécaniques : diagnostic vibratoire.
  - le *nucléaire* : surveillance et détection précoce d'incidents.
  - *l'imagerie satellitaire* : SPOT.

#### 2. CLASSES DE SIGNAUX

L'objectif du traitement du signal étant d'extraire de l'information utile d'une version dégradée d'un signal et/ou de transformer cette information, il importe dans un premier temps de pouvoir décrire les signaux sur lesquels on désire travailler.

#### 2.1 Signaux 1D, multi-D et images

Un signal se présente le plus souvent sous la forme d'une grandeur de dimension 1 évoluant en fonction du temps : on parle alors de signal temporel et l'on note x(t),  $t \in \mathbf{R}$ . C'est typiquement la situation que l'on rencontre en sortie d'un microphone, d'un accéléromètre, etc...

La variable dont dépend un signal n'est cependant pas nécessairement le temps. Ce peut être par exemple une variable d'espace, que ce soit directement (imagerie) ou indirectement (comme dans le cas du balayage spatio-temporel d'une image de télévision). Ce peut également être une variable de fréquence (spectromètre) : on notera alors le signal X(v).

Lorsque plusieurs (disons N) capteurs temporels sont utilisés simultanément, la collection des différents signaux scalaires fournit un signal multidimensionnel ou vectoriel de dimension N+1. C'est en particulier le cas des antennes formées de microphones (acoustique aérienne), d'hydrophones (acoustique sous-marine) ou de géophones (sismique).

Si les dimensions qui constituent un signal multidimensionnel sont maintenant de même nature, on parle d'image, que ce soit en 2D ou 3D. Un certain nombre de problèmes de traitement d'images, dits "bas niveau" (filtrage, segmentation, etc.) se transposent directement à partir du cas 1D. Certains autres, dits "haut niveau" sont spécifiques aux images et requièrent des concepts d'interprétation, de "vision", voire d'intelligence artificielle.

#### 2.2 Continu et discret

Les capteurs physiques fournissent souvent une information dépendant d'un "temps" continu. Il peut arriver cependant que ce dernier soit discret, que ce soit par échantillonnage (prélèvement de valeurs à certains instants, ou à certaines positions spatiales, seulement) ou par nature. On parle dans ce dernier cas de séries temporelles ("time series") ou séries chronologiques : on les rencontre par exemple en économie. On note un signal à temps discret  $x[n], n \in \mathbb{Z}$ .

D'une façon analogue, les *valeurs* prises par un signal peuvent être continues ou discrètes. Là encore, le caractère discret (en amplitude) d'un signal peut provenir soit d'une *quantification* (le capteur ne délivre qu'un nombre fini ou une infinité dénombrable de valeurs), soit de la nature même de l'information considérée (par exemple une population).

Bien entendu, les différentes situations peuvent se mélanger, un signal pouvant être à temps continu ou discret, quantifié ou non.

#### 2.3 Déterministe et aléatoire

Une distinction importante, car elle est souvent liée à la différence existant entre signal et bruit, est celle que l'on peut faire entre déterministe et aléatoire.

La notion de signal déterministe est étroitement liée au concept de reproductibilité exacte, et celle de signal aléatoire à celui d'épreuve et de reproductibilité statistique. En supposant qu'il soit possible de rejouer les conditions expérimentales ayant donné naissance à un signal, le caractère déterministe se manifeste par une reproduction à l'identique de la situation observée. Ceci contraste avec le caractère aléatoire pour lequel chaque épreuve nouvelle réserve une part, imprévisible, de variations, seules les propriétés d'ensemble (c'est-à-dire les lois statistiques qui régissent le signal) ayant une permanence. Lorsque l'on veut expliciter le caractère aléatoire d'un signal et mettre en avant sa dépendance vis-à-vis d'une épreuve  $\omega$ , on note  $x(t, \omega)$ . D'une certaine façon, un signal déterministe (on dit encore *certain*), n'est qu'un cas particulier de signal aléatoire pour lequel la densité de probabilité est une masse ponctuelle.

Plusieurs remarques peuvent être faites quant à la distinction entre déterministe et aléatoire :

- dans un mélange "signal + bruit", le terme "signal" n'est pas nécessairement déterministe. On peut très bien s'intéresser à un bruit dans un bruit : c'est le cas par exemple des sons non-voisés en parole bruitée.
- la notion de *signal* déterministe ou aléatoire est souvent liée à une propriété de *système* générateur d'un signal : l'un crée de l'ordre, l'autre du désordre. Si déterministe est lié à reproductibilité, aléatoire se trouve implicitement lié à imprédicibilité.
- aléatoire et imprédicible ne sont cependant pas équivalents. Un système déterministe (au sens qu'il est régi par des équations déterminées fournissant des solutions théoriques reproductibles) mais présentant une dépendance sensitive aux conditions initiales peut générer des signaux certains (en théorie) mais imprédicibles (en pratique) : on parle alors de chaos déterministe.

Stationnarité. Une propriété importante des signaux aléatoires est la dépendance temporelle de leurs lois d'ensemble. Si ces dernières sont invariantes par translation, c'est-à-dire si elles ne dépendent pas d'un temps absolu mais ne sont relatives qu'à des différences de temps, on parle de stationnarité. Bien que la stationnarité puisse être définie à tous les ordres, c'est la stationnarité de second ordre qui est la plus importante en pratique. Elle stipule d'une part que

$$\mathbb{E}\{x(t,\,\omega)\}=m_{\rm r},$$

c'est-à-dire que la valeur moyenne soit une constante indépendante du temps, et d'autre part (en supposant le processus centré, ce que l'on peut toujours obtenir par soustraction de sa valeur moyenne) que

$$E\{x(t, \omega)x^*(s, \omega)\} = \gamma_r(t - s),$$

c'est-à-dire que la covariance soit une fonction, appelée *fonction de corrélation*, qui ne dépende que de la différence des dates considérées.

#### 2.4 Energie, puissance moyenne et puissance

Un signal certain est dit d'énergie finie si l'on a

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt < \infty.$$

On peut considérer que c'est toujours le cas en pratique dans la mesure où les observations sont finies en temps et les capteurs ont une dynamique limitée. Il est néanmoins utile d'introduire des idéalisations pour lesquelles l'énergie peut devenir infinie. C'est en particulier le cas des signaux "éternels" comme les ondes monochromatiques. De tels signaux peuvent alors être caractérisés par une puissance moyenne finie s'ils sont tels que

$$P_x = \lim_{T \to \infty} \frac{1}{T} \int_{(T)} |x(t)|^2 dt < \infty.$$

L'énergie peut être vue comme liée à une norme sur un espace des signaux. Celle-ci est alors associée au produit scalaire :

$$\langle x, y \rangle = \int_{-\infty}^{+\infty} x(t) y^*(t) dt$$
  $(\Rightarrow ||x||^2 = \langle x, x \rangle = E_x)$ 

qui permet de munir l'espace des signaux certains d'énergie finie d'une structure d'espace de Hilbert. La norme choisie permet en effet de définir une *distance* entre signaux

$$d(x, y) = \|x - y\| = \left(\int_{-\infty}^{+\infty} |x(t) - y(t)|^2 dt\right)^{1/2}$$

telle que

$$\lim_{n \to \infty} x_n(t) = x(t) \implies \lim_{n \to \infty} d(x_n, x) = 0,$$

ce qui assure la complétude de l'espace. La convergence est dite en moyenne quadratique.

L'ensemble de ces considérations peut se transposer au cas aléatoire moyennant que la notion d'énergie soit remplacée par celle de puissance, ce qui revient à prendre comme produit scalaire la quantité

$$\langle x, y \rangle = \mathbb{E}\{x(t) \ y^*(t)\}.$$

Un signal aléatoire centré est alors dit de second ordre si

$$var\{x(t)\} = E\{|x(t)|^2\} < \infty.$$

Géométrie. L'intérêt de disposer d'une structure d'espace fonctionnel de type Hilbert pour les signaux tant certains qu'aléatoires est de pouvoir se placer dans un cadre géométrique propice aux tâches d'approximation par des arguments de minimisation de distance.

Ainsi, on pourra dire objectivement qu'un signal x(t) appartenant à un espace E possède une "meilleure" approximation y(t) si d(x, y) est minimale. Ce critère est celui de l'erreur quadratique moyenne minimale (EQMM) dans le cas aléatoire. Si y(t) est contraint à vivre dans un sous-espace F de l'espace E, la solution est immédiate : elle est directement fournie par la projection orthogonale de x(t) sur F.

Il est à noter que, comme

$$d^2(x, y) = ||x||^2 + ||y||^2 - 2 \operatorname{Ré}\{\langle x, y \rangle\},$$

minimiser une distance revient à maximiser le produit scalaire  $\langle x, y \rangle$  qui joue ainsi un rôle de mesure de ressemblance entre signaux.

Les situations d'énergie et de puissance finies n'épuisent pas la totalité des idéalisations utiles en Traitement du Signal, et il est souvent utile d'introduire aussi des distributions, que l'on pourra dans certains cas considérer comme limite de fonctions usuelles.

#### 2.6 Exemples

Le prototype du signal "éternel" est la sinusoïde  $a \cos(2\pi vt + \varphi)$  définie par une amplitude a, une fréquence v (ou une pulsation  $\omega = 2\pi v$ ) et une phase  $\varphi$ . C'est un exemple de signal certain d'énergie infinie et de puissance moyenne finie (en l'occurrence  $a^2/2$ ).

Ce signal peut être rendu aléatoire de diverses façons en agissant sur a, v ou  $\varphi$ . Ainsi, si l'on suppose que la phase est une variable aléatoire équirépartie sur  $[0, 2\pi[$ , on a

$$\mathbb{E}\{a\cos(2\pi\nu t + \varphi)\} = a\int_{0}^{2\pi}\cos(2\pi\nu t + \varphi)\frac{\mathrm{d}\varphi}{2\pi} = 0.$$

Plus généralement, si l'on construit deux signaux

$$s_1(t) = a \cos(2\pi v t + \varphi)$$

$$s_2(t) = a \sin(2\pi v t + \varphi)$$

avec  $\varphi$  équirépartie sur  $[0, 2\pi]$ , il est facile de voir que

$$\langle s_1, s_2 \rangle = \frac{a^2}{2} \int_0^{2\pi} \sin(2(2\pi vt + \varphi)) \frac{d\varphi}{2\pi} = 0,$$

ce qui montre que  $s_1(t)$  et  $s_2(t)$  sont décorrélés au sens du produit scalaire sans pour autant être indépendants puisque, par construction,

$$[s_1(t)]^2 + [s_2(t)]^2 = a^2.$$

Une façon de restreindre le domaine d'existence d'un signal éternel (ou de rendre compte du caractère non-infini d'une observation) est de faire usage de pseudo-fonctions nulles pour certaines valeurs. Ainsi, l'échelon unité de Heaviside

$$u(t) = \begin{cases} 0, \ t < 0 \\ 1, \ t \ge 0 \end{cases}$$

permet de "causaliser" une observation en ne la faisant exister que pour les temps antérieurs au temps courant (le passé).

Combiner deux échelons unité permet de définir une fonction *porte* (ou indicatrice d'intervalle)

$$\pi_{T}(t) = \frac{1}{T} u(t + \frac{T}{2}) u(\frac{T}{2} - t) = \begin{cases} \frac{1}{T}, & |t| \le \frac{T}{2} \\ 0, & t > \frac{T}{2} \end{cases}$$

grâce à laquelle il est possible de prélever une portion de signal sur une durée finie T. Ainsi définie, la porte est d'aire unité et d'énergie 1/T. Sa limite lorsque T tend vers 0 n'est définie qu'au sens des distributions : il s'agit de l'impulsion de Dirac notée  $\delta(t)$ .

#### 3. TRANSFORMATIONS

Les transformations jouent un rôle crucial en Traitement du Signal. Elles correspondent non seulement à l'ensemble des modifications qu'un signal peut subir dans une chaîne de traitement mais encore à toute modification que l'on peut apporter aux données brutes afin de mieux les préparer à un traitement.

Quoiqu'une transformation puisse en principe être quelconque, les transformations *linéaires* sont parmi les plus importantes. Ce sont principalement celles-ci que nous considérerons dans un premier temps.

# 3.1 Concepts de base

Filtrage linéaire. Dans le cas général, une transformation linéaire est associée à l'action d'un opérateur caractérisé en temps continu par un noyau intégral

$$y(t) = \int_{-\infty}^{+\infty} h(t, u) \ x(u) \ du$$

et en temps discret par une matrice

$$y[n] = \sum_{m = -\infty}^{+\infty} h[n, m] x[m].$$

Par définition, on convient d'appeler filtre linéaire un opérateur linéaire invariant par translation temporelle. Il s'ensuit que le noyau h(t, u) d'un filtre linéaire n'est fonction que de t-u. En effet, l'invariance par translation impose que la filtrée de la translatée coïncide avec la translatée de la filtrée, ce qui est équivalent à écrire

$$\int_{-\infty}^{+\infty} h(t - T, u) x(u) du = \int_{-\infty}^{+\infty} h(t, u + T) x(u) du$$

avec T quelconque, soit encore

$$\int_{-\infty}^{+\infty} [h(t-T, u) - h(t, u+T)] \ x(u) \ du = 0.$$

Cette égalité devant être vérifiée pour tout signal x(t), on a nécessairement

$$h(t \mid u \mid) = h(t + T, u + T)$$

pour tout T. En particulier, si l'on choisit T = -u, il vient

$$h(t, u) = h(t - u, 0)$$

ce qui prouve le résultat annoncé. Par abus de notation, on notera généralement

$$h(t, u) = h(t - u).$$

Un filtre linéaire est donc caractérisé par une relation entrée-sortie de type convolutif associée à une réponse impulsionnelle h(t). Celle-ci tire son nom du fait qu'elle s'identifie à la sortie du filtre lorsqu'on l'excite en entrée par une impulsion de Dirac :

$$x(t) = \delta(t) \implies y(t) = \int_{-\infty}^{+\infty} h(t - u) \ \delta(u) \ du = h(t).$$

Le succès des filtres linéaires tient à nombre de propriétés dont certaines, fréquentielles, apparaîtront plus tard. On peut toutefois noter que, si peu de systèmes réels sont parfaitement linéaires, une approximation linéaire dans une plage de fonctionnement courant est souvent justifiée. La structure convolutive de la relation entrée-sortie des filtres linéaires permet alors, de par l'associativité et la commutativité de la convolution, de décrire commodément une chaîne dont les maillons sont tous les linéaires, de les regrouper et éventuellement d'en intervertir l'ordre.

Notons enfin que le filtrage linéaire conserve le caractère gaussien.

Transformations et représentations. Il est possible de faire une distinction entre transformations et représentations, ces dernières étant explicitement destinées à fournir un nouveau point de vue sur un signal, à le regarder sous un angle particulier, sans en altérer l'information. Ceci nécessite qu'une représentation permette de revenir au signal de départ de façon exacte. Si, dans le cas linéaire et inversible, les deux notions se confondent, il n'en est pas de même dans des cas non-linéaires comme par exemple une élévation au carré ou un redressement simple alternance.

Une représentation, qui change éventuellement la nature de l'espace dans lequel un signal vit, résulte souvent d'une *projection* de celui-ci sur un ensemble de fonctions (fixées *a priori* ou plus ou moins adaptées) qui constituent (ou jouent le rôle d') une *base* de décomposition.

Nous allons maintenant donner plusieurs exemples de transformations et de représentations dans des cas tant continus que discrets.

# 3.2 Une représentation continue de signaux continus : la transformation de Fourier

L'exemple de transformation-représentation le plus courant en traitement du signal est sans conteste la *transformation de Fourier*. Par définition, celle-ci et son inverse sont données par le couple

$$X(v) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi vt} dt$$
$$x(t) = \int_{-\infty}^{+\infty} X(v) e^{i2\pi vt} dv$$

dans lequel la variable v a un statut de *fréquence*. Quelques exemples simples et courants de paires de fonctions liées par transformation de Fourier :

$$\cos 2\pi v_0 t \qquad \frac{1}{2} \left[ \delta(v - v_0) + \delta(v + v_0) \right]$$

$$u(t) \qquad \frac{1}{2} \left[ \delta(v) + \frac{1}{i\pi v} \right]$$

$$\pi_I(t) \qquad \sin c(\pi v T) = \frac{\sin \pi v T}{\pi v T}$$

$$\delta(t) \qquad 1$$

$$e^{-\alpha t^2} \qquad (\pi/\alpha)^{1/2} e^{-\pi^2 v^2/\alpha}$$

Nous ne discuterons pas ici les nombreuses propriétés mathématiques de la transformation de Fourier que l'on pourra trouver un peu partout, ni ses conditions d'existence. Nous nous concentrerons sur quelques-unes spécifiques de son utilisation et de l'intérêt qu'elle offre en traitement du signal.

Transformation de Fourier et filtrage linéaire. La transformation de Fourier entretient des liens privilégiés avec le filtrage linéaire. En effet, elle transforme une convolution en produit, ce qui permet d'écrire la relation d'entrée-sortie sous la forme

$$Y(v) = H(v) X(v),$$

où la transformée de Fourier H(v) de la réponse impulsionnelle h(t) est appelée gain complexe. Il mesure, en module et en phase, la modification introduite par le filtre sur le spectre du signal d'entrée. Cette relation permet une évaluation particulièrement simple de l'influence d'une cascade de filtres dans une chaîne : les gains (réels) se multiplient et les phases s'ajoutent.

Une autre propriété relative aux filtres linéaires justifie l'utilisation intensive de la transformation de Fourier. Supposons en effet que l'entrée soit une exponentielle complexe, alors la sortie vaut

$$\int_{-\infty}^{+\infty} h(t - u) e^{i2\pi v_0 u} du = H(v_0) e^{i2\pi v_0 t},$$

ce qui n'est autre que l'entrée, à un facteur multiplicatif près. On voit ainsi que les exponentielles complexes sont *fonctions propres* des opérateurs filtres linéaires, la valeur propre associée étant le gain complexe à la fréquence correspondante. Par suite, la transformation de Fourier, qui décompose un signal sur la base de telles exponentielles complexes, est naturellement adaptée aux transformations dans les filtres linéaires.

Corrélation et densité spectrale d'énergie. De la même façon que  $|x(t)|^2$  définit une puissance instantanée que l'on peut interpréter comme une densité temporelle d'énergie,  $|X(v)|^2$  définit une densité spectrale d'énergie car (théorème de Parseval)

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(v)|^2 dv.$$

Calculons la transformée de Fourier de cette densité spectrale d'énergie. On obtient

$$\gamma_{x}(\tau) \equiv \int_{-\infty}^{+\infty} |X(v)|^{2} e^{i2\pi v \tau} dv = \int_{-\infty}^{+\infty} X(v) [X(v) e^{-i2\pi v \tau}]^{*} dv,$$

soit encore, en utilisant le fait que la transformation de Fourier soit une isométrie,

$$\gamma_{x}(\tau) = \int_{-\infty}^{+\infty} x(t) \ x^{*}(t - \tau) \ dt.$$

Cette dernière quantité est appelée fonction d'autocorrélation : elle mesure le degré de ressemblance qu'un signal partage avec ses translatées. Parce qu'elle est définie comme la transformée de Fourier d'une fonction non-négative (la densité spectrale d'énergie), une fonction d'autocorrélation est dite définie non-négative. Elle est de plus symétrique (en fait à symétrie hermitique dans le cas complexe)

$$\gamma_{r}(-\tau) = \gamma_{r}^{*}(\tau)$$

et maximale à l'origine. En effet, en appliquant l'inégalité de Cauchy-Schwartz à la définition, on obtient

$$|\gamma_x(\tau)|^2 \leq \int\limits_{-\infty}^{+\infty} |x(t)|^2 dt \cdot \int\limits_{-\infty}^{+\infty} |x(t-\tau)|^2 dt = E_x^2,$$

soit encore

$$|\gamma_{\rm r}(\tau)| \le \gamma_{\rm r}(0)$$

puisque  $\gamma_{\rm r}(0) = E_{\rm r}$ ...

Illustrons le comportement comparé d'un couple "autocorrélation-densité spectrale d'énergie" dans le cas d'un signal réel du type

$$s(t) = x(t) + x(t - T).$$

On obtient aisément

$$\gamma_s(\tau) = \gamma_x(\tau + T) + 2 \gamma_x(\tau) + \gamma_x(\tau - T).$$

On en déduit que la densité spectrale d'énergie a pour valeur

$$|S(v)|^2 = 2 |X(v)|^2 (1 + \cos 2\pi vT),$$

ce que l'on aurait pu calculer directement.

Puisqu'une fonction d'autocorrélation est maximale à l'origine, celle relative à la somme de deux signaux décalés dans le temps admet des maxima secondaires à t = +T et t = -T, ce qui fournit un moyen d'estimer le décalage. A T fixé, la séparation sera d'autant plus aisée que le pic de corrélation sera plus étroit, ce qui correspond à une bande spectrale du signal plus large.

Densité spectrale de puissance. Revenons au cas des processus aléatoires et considérons la construction d'une variable aléatoire s par une superposition du type

$$s = \sum_{i} \lambda_{i} x(t_{i})$$

où les  $\lambda_i$  et les  $t_i$  sont quelconques, les  $x(t_i)$  étant supposés stationnaires. Alors, nécessairement

$$E\{|s|^2\} \ge 0$$
,

soit encore

$$\mathrm{E}\{\mid s\mid^2\} = \sum_i \ \sum_j \ \lambda_i \ \lambda^*_i \ \gamma_s(t_i - t_j) \geq 0.$$

Tout comme dans le cas certain, on en déduit que la fonction de corrélation

$$\gamma_c(\tau) \equiv \mathbb{E}\{x(t) | x^*(t-\tau)\}$$

est définie non-négative, ce qui implique que sa transformée de Fourier est partout nonnégative : cette dernière, définie par

$$\Gamma_{x}(v) \equiv \int_{-\infty}^{+\infty} \gamma_{s}(\tau) e^{-i2\pi v \tau} d\tau,$$

est appelée densité spectrale de puissance car

$$\int_{-\infty}^{+\infty} \Gamma_x(v) \, \mathrm{d}v = \gamma_s(0) = \mathrm{E}\{|x(t)|^2\}.$$

Ce résultat fondamental constitue ce que l'on appelle le théorème de Wiener-Khintchine.

A titre d'exemple, considérons de nouveau le signal

$$x(t) = a\cos(2\pi v_0 t + \varphi)$$

avec  $\varphi$  équirépartie sur  $[0, 2\pi]$ . On obtient que

$$\gamma_s(\tau) = a^2 \, \mathbb{E} \{ \frac{1}{2} \left[ \cos 2\pi v_0 \tau + \cos (2\pi v_0 t - 2\pi v_0 t + 2\varphi) \right] \} = \frac{a^2}{2} \cos 2\pi v_0 \tau$$

et donc

$$\Gamma_{x}(v) = \frac{d^{2}}{4} [\delta(v - v_{0}) + \delta(v - v_{0})],$$

ce qui constitue un spectre de raies.

Les densités spectrales de puissance se transforment de manière particulièrement simple dans les filtrages linéaires. Si l'on considére en effet la situation où y(t) est la sortie d'un filtre de réponse h(t) soumis en entrée à l'excitation x(t), on a alors

$$\gamma_{y}(\tau) = \mathbb{E}\{y(t) | y^*(t - \tau)\}$$

avec

$$y(t) = \int_{-\infty}^{+\infty} h(t - u) \ x(u) \ du..$$

Par suite.

$$\gamma_{y}(\tau) = \mathbb{E}\left\{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(t-u) h^*(t-\tau-v)x(u) x^*(v) du dv\right\} = \int_{-\infty}^{+\infty} \gamma^*_{h}(\theta-\tau) \gamma_{x}(\theta) d\theta,$$

ce qui peut encore s'écrire, en utilisant l'identité de Parseval,

$$\gamma_{\mathbf{y}}(\tau) = \int_{-\infty}^{+\infty} |H(\mathbf{v})|^2 \Gamma_{\mathbf{x}}(\mathbf{v}) e^{\mathrm{i} 2\pi \mathbf{v} \tau} d\mathbf{v}.$$

On en déduit donc que

$$\Gamma_{v}(v) = |H(v)|^2 \Gamma_{x}(v).$$

Ce résultat se généralise au cas de deux filtrages en parallèle pour lesquels, par un calcul analogue, on aboutit au résultat

$$\Gamma_{12}(v) = H_1(v) H^*_2(v) \Gamma_x(v),$$

appelé parfois formule des interférences. Cette expression montre en particulier que si  $h_1$  et  $h_2$  sont deux filtres spectralement disjoints, l'intercorrélation

$$\gamma_{12}(\tau) \equiv \mathbb{E}\{y_1(t) \ y_2^*(t-\tau)\}$$

de leurs sorties est nulle : ils ne partagent pas d'énergie.

# 3.3 Représentations discrètes de signaux continus

Bien que les signaux soient le plus souvent associés à des évolutions naturelles à temps continu, les conserver sous une telle forme peut présenter des inconvénients évidents, en particulier au niveau de l'archivage. Il est donc intéressant de disposer de

représentations qui transforment un signal à temps continu (et donc décrit par une infinité non dénombrable de valeurs) en une collection discrète de nombres.

D'une façon tout-à-fait générale, l'approche la plus naturelle est du type "projection sur une famille de fonctions élémentaires"  $\psi_k(t)$ , si possible orthogonales entre elles  $(\langle \psi_k, \psi_l \rangle = 0$  si  $k \neq l)$ , les valeurs discrètes représentatives du signal n'étant alors autres que les poids associés à ces fonctions, tels qu'ils sont fournis par l'opération de projection  $\langle x, \psi_k \rangle$ .

Séries de Fourier. Un premier exemple de représentation discrète de signaux continus est donné par le développement en série de Fourier, qui s'applique tant aux signaux à support limité qu'aux signaux périodiques. Dans le premier cas, on considérera l'espace des signaux d'énergie finie sur [-T/2, +T/2] et dans le second sa périodisation sur tout R. Dans les deux cas, une base orthonormée est fournie (sur le support correspondant) par le système

$$\{\psi_k(t) = \frac{1}{\sqrt{T}} \mathrm{e}^{\mathrm{i} 2\pi k} \frac{t}{T} \ , \ k \in \mathbf{Z}\},\label{eq:psik}$$

puisqu'alors

$$<\psi_k, \ \psi_l> = \frac{1}{T} \int_{-T/2}^{+T/2} e^{-i2\pi(k-l)\frac{t}{T}} dt = \operatorname{sinc}(\pi(k-l)) = \delta_{kl}.$$

On en déduit les coefficients de série de Fourier définis par

$$X_k = \frac{1}{\sqrt{T}} \langle x, \psi_k \rangle = \frac{1}{T} \int_{-T/2}^{+T/2} x(t) e^{-i2\pi k \frac{t}{T}} dt,$$

d'où l'on déduit la représentation

$$x(t) = \sum_{k = -\infty}^{+\infty} X_k e^{i2\pi k \frac{t}{T}}.$$

Il s'en suit entre autres que (égalité de Parseval)

$$\frac{1}{T} \int_{-T/2}^{+T/2} |x(t)|^2 dt = \sum_{k=-\infty}^{+\infty} |X_k|^2.$$

Notons enfin que la transformée de Fourier peut être introduite (de manière heuristique) comme la limite des coefficients de série de Fourier lorsque  $T \rightarrow \infty$ .

Echantillonnage. Une série de Fourier permet de décrire un signal à temps continu par un ensemble discret de fréquences. Il est également possible de décrire ce même signal par un ensemble discret d'instants: l'opération correspondante est appelée échantillonnage. Nous ne considérerons ici que le cas de l'échantillonnage uniforme, qui consiste à prélever des valeurs du signal à des instants équidistants  $\{nT_e, n \in \mathbf{Z}\}$ , multiples d'un intervalle élémentaire  $T_e$  appelé période d'échantillonnage. On peut alors formaliser l'opération d'échantillonnage en écrivant

$$x[n] \equiv x(nT_e) = \int_{-\infty}^{+\infty} x(u) \, \delta(u - nT_e) \, du,$$

ou encore en disant que le signal échantillonné résulte de l'action d'un "peigne de Dirac" sur le signal à temps continu :

$$x_{\text{\'ech}}(t) = x(t) \sum_{n = -\infty}^{+\infty} \delta(t - nT_e).$$

Comme la transformation de Fourier transforme une multiplication temporelle en une convolution fréquentielle, et un peigne temporel de période  $T_e$  en un peigne fréquentiel de période  $1/T_e$ , il s'en suit que le spectre d'un signal échantillonné est le périodisé du spectre du signal à temps continu :

$$X_{\text{\'ech}}(v) = \frac{1}{T_e} \sum_{n=-\infty}^{+\infty} X(v - \frac{n}{T_e}).$$

Cette propriété fondamentale permet d'établir le théorème d'échantillonnage selon lequel un signal à bande limitée [-B, +B] est entièrement décrit par ses échantillons  $x(nT_e)$  si  $T_e \le 1/2B$ . Il suffit pour cela de remarquer que, comme

$$x(t) = \int_{-\infty}^{+\infty} X(v) e^{i2\pi vt} dv = \int_{-B}^{+B} X(v) e^{i2\pi vt} dv,$$

on a également

$$x(t) = \int_{-B}^{+B} X_{\text{\'ech}}(v) e^{i2\pi vt} dv$$

si la périodisation induite par l'échantillonnage n'altère pas le contenu spectral de la bande de base [-B, +B] (dans le cas contraire, on parle de *repliement spectral*, ou encore d'*aliasing*, selon la terminologie anglo-saxonne).

Dit d'une autre manière, comme X(v) est à bande limitée, on peut le développer en série de Fourier selon

$$X(v) = \sum_{k=-\infty}^{+\infty} c_k e^{i2\pi k} \frac{v}{2B}$$

avec

$$c_k = \frac{1}{2B} \int_{-B}^{+B} X(v) e^{-i2\pi k} \frac{v}{2B} dv = \frac{1}{2B} x(-\frac{k}{2B}).$$

Par suite,

$$x(t) = \int_{-B}^{+B} X(v) e^{i2\pi vt} dv = \sum_{n=-\infty}^{+\infty} x(\frac{n}{2B}) \int_{-B}^{+B} \frac{1}{2B} e^{-i2\pi(t-\frac{n}{2B})v} dv,$$

soit encore

$$x(t) = \sum_{n=-\infty}^{+\infty} x(\frac{n}{2B}) \operatorname{sinc}(2\pi B(t - \frac{n}{2B})).$$

Ce résultat fournit une formule d'interpolation pour les signaux à bande limitée puisque toute valeur du signal peut se déduire des seules valeurs discrètes résultant de l'échantillonnage. On choisit souvent de prendre comme unité de temps la période d'échantillonnage ( $T_e \equiv 1$ ), ce qui revient à normaliser la bande spectrale d'analyse à l'intervalle de fréquences réduites [-1/2, +1/2]. On écrit alors la formule d'interpolation (dite encore de reconstitution) selon

$$x(t) = \sum_{n = -\infty}^{+\infty} x[n] \operatorname{sinc}(\pi(t - n)).$$

La condition à remplir pour un échantillonnage correct (appelée condition de *Shannon* ou encore de *Nyquist*) est une cadence *minimum* liée à l'encombrement spectral du signal à analyser. Il est clair que, si un signal de bande spectrale [-B, +B] est correctement échantillonné à la fréquence de Shannon 1/2B, il l'est également à toute fréquence supérieure puisque

$$X(v) = \pi_{2R}(v) X(v) = \pi_{2R}(v) X(v)$$
 si  $B' \ge B$ .

On en déduit par transformation de Fourier que

$$x(t) = \int_{-\infty}^{+\infty} x(u) \operatorname{sinc}(2\pi B'(t-u)) du \quad \text{si} \quad B' \ge B,$$

ce qui signifie que, dans l'espace des signaux à bande limitée [-B, +B], tout sinus eardinal  $\operatorname{sinc}(2\pi B't)$  est *unité de convolution* (c'est-à-dire a une action équivalente à celle de la distribution de Dirac) si  $B' \ge B$ .

Le théorème d'échantillonnage, établi dans le cas des signaux certains, s'applique également aux processus aléatoires stationnaires à bande limitée [-B, +B], l'erreur de reconstitution étant nulle au sens de la moyenne quadratique. Si l'on calcule en effet la quantité

$$\varepsilon^2 = \mathbb{E}\{|x(t) - \sum_{n=-\infty}^{+\infty} x(\frac{n}{2B})\operatorname{sinc}(2\pi B(t - \frac{n}{2B}))|^2\},\$$

on obtient après développement du module carré et application de l'opérateur d'espérance mathématique

$$\varepsilon^{2} = \gamma_{x}(0) - 2 \sum_{n=-\infty}^{+\infty} \gamma_{x}(t - \frac{n}{2B}) \operatorname{sinc}(2\pi B(t - \frac{n}{2B}))$$

$$+ \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \gamma_{x}(\frac{n-m}{2B}) \operatorname{sinc}(2\pi B(t - \frac{n}{2B})) \operatorname{sinc}(2\pi B(t - \frac{m}{2B})).$$

La fonction d'autocorrélation  $\gamma_x(\tau)$  étant déterministe et à bande limitée [-B, +B], il est possible de lui appliquer le théorème d'échantillonnage relatif aux signaux certains. Il s'en suit d'une part que

$$\sum_{n=-\infty}^{+\infty} \gamma_x(t - \frac{n}{2B}) \operatorname{sinc}(2\pi B(t - \frac{n}{2B})) = \gamma_x(0)$$

et d'autre part que

$$\sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \gamma_x(\frac{n-m}{2B}) \operatorname{sinc}(2\pi B(t-\frac{n}{2B})) \operatorname{sinc}(2\pi B(t-\frac{m}{2B}))$$

$$= \sum_{m=-\infty}^{+\infty} \operatorname{sinc}(2\pi B(t - \frac{m}{2B})) \sum_{n=-\infty}^{+\infty} \gamma_x(\frac{n - m}{2B}) \operatorname{sinc}(2\pi B(t - \frac{n}{2B}))$$

$$= \sum_{m=-\infty}^{+\infty} \operatorname{sinc}(2\pi B(t - \frac{m}{2B})) \gamma_x(t - \frac{m}{2B})$$

$$= \gamma_x(0),$$

ce qui entraîne la *nullité* de l'erreur quadratique moyenne  $\varepsilon^2$ .

Karhunen-Loève. Une série de Fourier représente un signal quelconque comme superposition de segments d'ondes monochromatiques. L'universalité de cette décomposition trouve son intérêt dans l'importance des exponentielles complexes et dans leur interprétation physique. En retour, sa limitation est d'imposer a priori des éléments de décomposition qui ne sont pas nécessairement en rapport avec la structure réelle du signal analysé, conduisant ainsi à l'éventualité d'un nombre important de coefficients pour atteindre une précision d'approximation donnée.

De la même façon, l'échantillonnage repose sur la seule fréquence maximale d'un signal, sans prendre en compte davantage d'informations quant à son contenu spectral à l'intérieur de la bande utile.

Une autre possibilité apparaît ainsi, qui consiste à ne pas fixer *a priori* le choix d'une base de décomposition mais à l'*adapter* au signal analysé, par exemple dans un but de parcimonie de la représentation.

Dans le contexte des signaux aléatoires, un exemple typique à cet égard est fourni par la *décomposition de Karhunen-Loève*, qui exploite la structure propre de la covariance du signal analysé. L'idée fondamentale est de disposer, dans l'espace des signaux aléatoires (centrés) à support limité [-T/2, +T/2], d'une représentation

$$x(t) = \sum_{k = -\infty}^{+\infty} x_k \ \psi_k(t) \ ; \ x_k = \int_{-T/2}^{+T/2} x(t) \ \psi_k^*(t) \ dt,$$

qui soit doublement orthogonale en ce sens que

$$\begin{cases} \int_{-T/2}^{+T/2} \psi_k(t) \ \psi_l^*(t) \ dt = \delta_{kl} \\ \mathbb{E}\{x_k x_l^*\} = \operatorname{var}\{x_k\} \ \delta_{kl} \end{cases}$$

Afin de construire une telle décomposition, partons de la quantité  $x_k^* x(t)$ , avec t dans l'intervalle [-T/2, +T/2]. Celle-ci peut s'écrire indifféremment

$$x_k^* x(t) = \int_{-T/2}^{+T/2} x(t) \ x^*(s) \ \psi_k(s) \ ds$$

Ou

$$x_k^* x(t) = \sum_{n=-\infty}^{+\infty} x_n x_k^* \psi_n(t).$$

Par suite, en égalant les deux membres de droite ci-dessus et en en prenant l'espérance mathématique, on obtient l'équation

$$\int_{-T/2}^{+T/2} E\{x(t) \ x^*(s)\} \ \psi_k(s) \ ds = var\{x_k\} \ \psi_k(t) \quad , \quad t \in [-T/2, +T/2].$$

Ceci signifie qu'une décomposition doublement orthogonale d'un processus aléatoire est caractérisée par le fait que ses fonctions de base sont les *fonctions propres* de la covariance du processus, les valeurs propres associées s'identifiant aux variances des coefficients de la décomposition. Dit d'une autre manière, la représentation (doublement orthogonale) de Karhunen-Loève est celle qui *diagonalise* l'opérateur de covariance.

En utilisant la décorrélation des coefficients relatifs au signal, on obtient facilement une décomposition de la covariance qui s'écrit (théorème de Mercer)

$$E\{x(t) | x^*(s)\} = \sum_{k = -\infty}^{+\infty} var\{x_k\} | \psi_k(t) | \psi_k^*(s).$$

Si le processus étudié est *stationnaire*, l'équation propre de Karhunen-Loève devient

$$\int_{-T/2}^{+T/2} \gamma_x(t-s) \ \psi_k(s) \ ds = \operatorname{var}\{x_k\} \ \psi_k(t) \ , \ t \in [-T/2, +T/2],$$

ce qui tend vers une équation de convolution dans la limite  $T \rightarrow \infty$ . Par transformation de Fourier, on obtient alors formellement l'équation équivalente

$$\Gamma_x(v) \Psi_k(v) = \text{var}\{x_k\} \Psi_k(v)$$

dont une "base" d'exponentielles complexes  $e^{i2\pi v_k t}$  est solution, les coefficients de la décomposition vérifiant alors

$$var\{x_k\} = \Gamma_x(v_k).$$

La décomposition de Karhunen-Loève se réduit ainsi, dans ce cas limite particulier, à une décomposition de Fourier. De manière plus particulière encore, si le processus étudié est *blanc*, c'est-à-dire si sa fonction de covariance est telle que

$$\mathrm{E}\{x(t)\;x^*(s)\}=\gamma_0\;\delta(t-s),$$

il est facile de vérifier que l'on a

$$\gamma_0 \ \psi_k(t) = \text{var}\{x_k\} \ \psi_k(t) \ , \ t \in [-T/2, +T/2],$$

ce qui montre que *toute base* convient, les coefficients de la décomposition ayant tous même variance : le niveau *constant* de la densité spectrale du bruit blanc.

# 3.4 Représentation continue de signaux discrets

Que ce soit par nature (séries temporelles) ou par échantillonnage, les signaux sont souvent disponibles sous forme discrète. On les notera génériquement x[n], la détermination principale de leur spectre X(v) occupant alors l'intervalle de fréquences réduites [-1/2, +1/2].

Transformation de Fourier à temps discret. Le spectre d'un signal à temps discret étant périodique (ici de période 1), il est possible de le développer en série de Fourier. Ceci conduit à la représentation

$$X(v) = \sum_{n=-\infty}^{+\infty} x[n] e^{-i2\pi nv},$$

dans laquelle les échantillons x[n] sont définis par

$$x[n] = \int_{-1/2}^{+1/2} X(v) e^{i2\pi nv} dv.$$

Cette paire d'équations constitue une transformation de Fourier des signaux à temps discret. Quoiqu'elle puisse être étudiée pour elle-même, il s'avère en fait qu'elle n'est qu'un cas particulier d'une transformation plus générale dans laquelle les exponentielles complexes sont remplacées par une variable complexe quelconque : c'est cette dernière que nous considérerons plus en détail.

*Transformation en z.* Par définition, la *transformée en z* d'un signal à temps discret x[n] est donnée par la relation

$$X(z) \equiv \sum_{n=-\infty}^{+\infty} x[n] \ z^{-n}$$

dans laquelle z est une variable complexe arbitraire. On vérifie que la transformée de Fourier à temps discret s'en déduit formellement comme cas particulier en posant

$$z = e^{i2\pi v}$$

c'est-à-dire en évaluant la transformée en z sur le cercle unité |z|=1, que l'on notera  $C_u$ .

Pour que cette identification ait un sens, il faut cependant que  $C_u$  soit à l'intérieur du domaine de convergence de la transformée. Dans le cas le plus général, celui-ci est un anneau compris entre deux cercles centrés à l'origine du plan complexe. En effet, comme

$$|X(z)| \le \sum_{n=-\infty}^{+\infty} |x[n]| |z|^{-n},$$

les limites du domaine de convergence sont des cercles |z| = cte. Séparant en outre la somme totale en deux sommes partielles, relatives l'une aux indices positifs ou nuls  $n \ge 0$  et l'autre aux indices négatifs n < 0, il est clair que la première (resp. la seconde) converge à l'extérieur (resp. à l'intérieur) d'un cercle.

En multipliant la définition par  $z^{k-1}$  et en intégrant le résultat sur un cercle quelconque  $C_+$  (centré à l'origine, inclus dans l'anneau de convergence et parcouru dans le sens positif), on obtient

$$\int\limits_{{\rm C}_+} X(z) \; z^{k-1} \; {\rm d}z = \sum_{n=-\infty}^{+\infty} x[n] \int\limits_{{\rm C}_+} z^{k-n-1} \; {\rm d}z.$$

En utilisant alors le théorème de Cauchy

$$\int\limits_{C_+} z^{k-n-1} \, \mathrm{d}z = \mathrm{i} 2\pi \, \delta_{kn},$$

ceci montre que la transformée en z s'inverse selon

$$x[n] = \frac{1}{i2\pi} \int_{C_{+}} X(z) z^{n-1} dz.$$

Il est donc possible d'évaluer cette quantité inverse comme somme des résidus associés à tous les pôles situés à l'intérieur du contour d'intégration.

Citons (sans les démontrer) quelques-unes des nombreuses propriétés de la transformée en z:

- conjugaison complexe:  $y[n] = x^*[n] \implies Y(z) = X^*(z^*)$
- translation en temps:  $y[n] = x[n-k] \implies Y(z) = z^{-k} X(z)$
- inversion du temps :  $v[n] = x[-n] \implies Y(z) = X(1/z)$
- dilatation:  $y[n] = a^n x[n] \implies Y(z) = X(a^{-1} z)$
- convolution:  $c[n] = \sum_{k=-\infty}^{+\infty} x[k] \ y[n-k] \implies C(z) = X(z) \ Y(z)$
- corrélation:  $\gamma[n] = \sum_{k=-\infty}^{+\infty} x[k] \ y^*[k-n] \implies \Gamma(z) = X(z) \ Y^*(1/z^*)$
- multiplication:  $p[n] = x[n] \ y[n] \implies P(z) = \frac{1}{i2\pi} \int_{C_+} X(z) \ Y(1/z) z^{-1} \ dz$

Donnons également quelques exemples de transformées en z usuelles. Soit ainsi le signal

$$x[n] = a^n u[n],$$

où u[n] est l'échelon unité. D'après la définition de la transformée en z, il vient

$$X(z) = \sum_{n=0}^{+\infty} a^n z^{-n} = \sum_{n=0}^{+\infty} (a z^{-1})^n.$$

Ceci constitue une série géométrique de raison  $a z^{-1}$ , dont la convergence est assurée si  $|a|z^{-1}| < 1$ , soit encore |z| > |a|. La somme de cette série vaut alors

$$X(z) = \frac{1}{1 - a z^{-1}}.$$

(Remarque : le cas particulier a = 1 fournit la quantité

$$X(z) = \frac{1}{1 - z^{-1}},$$

transformée en z de l'échelon unité u[n] (à laquelle est associée le domaine de convergence |z| > 1, c'est-à-dire l'extérieur du cercle unité.) Cette transformée en z peut s'écrire de manière équivalente

$$X(z) = \frac{z}{z - a}$$

ce qui, dans le plan complexe, la caractérise par deux singularités : un zéro en z = 0 et un  $p\hat{o}le$  en z = a.

Réciproquement, une caractérisation pôles-zéros peut permettre de définir, à une constante multiplicative près, une transformée en z et le signal qui lui est associé. Ainsi, la donnée de deux zéros en  $z_1=0$  et  $z_2=r\cos 2\pi v_0$ , et de deux pôles complexes conjugués l'un de l'autre en  $p_1={\rm e}^{{\rm i}2\pi v_0}$  et  $p_2={\rm e}^{-{\rm i}2\pi v_0}$  permet d'écrire

$$X(z) = \frac{(z - z_1)(z - z_2)}{(z - p_1)(z - p_2)},$$

soit, après quelques calculs,

$$X(z) = \frac{1 - r z^{-1} \cos 2\pi v_0}{1 - 2r z^{-1} \cos 2\pi v_0 + r^2 z^{-2}}, \quad |z| > r,$$

ce qui correspond au signal

$$x[n] = r^n \cos 2\pi v_0 n \ u[n].$$

Considérons enfin un autre exemple de calcul d'une transformée en z inverse. Soit à cet effet la fonction

$$X(z) = \frac{(a - 1/a) z}{(z - a)(z - 1/a)}, \quad |a| < |z| < 1/|a|.,$$

qui possède deux pôles simples, l'un en z=a et l'autre en z=1/a.. Pour  $n \ge 0$ , la fonction X(z)  $z^{n-1}$  a elle aussi un pôle simple à l'intérieur du cercle unité et le calcul du résidu associé donne  $R_a=a^n$ , ce qui conduit à

$$x[n] = \frac{1}{i2\pi} \int_{C_{n+}} X(z) z^{n-1} dz = \frac{1}{i2\pi} i2\pi R_a = a^n, \quad n \ge 0.$$

Par contre, pour n < 0, cette même fonction a un pôle de multiplicité n en z = 0. L'évaluation de x[n] est cependant possible en calculant soit le résidu correspondant, soit l'expression

$$x[n] = -\frac{1}{\mathrm{i} 2\pi} \int_{C_{n}} X(z) \ z^{n-1} \ \mathrm{d}z = \frac{1}{\mathrm{i} 2\pi} \, \mathrm{i} 2\pi \ R_{1/a} = a^{-n}, \quad n < 0.$$

On en conclut que la transformée en z inverse de la fonction considérée vaut

$$x[n] = a^{|n|}.$$

Il s'agit donc d'un signal exponentiel symétrique.

# 3.5 Représentation discrète de signaux discrets

Séries de Fourier discrètes. Considérons un signal x(t) de durée finie T. Par application du théorème d'échantillonnage dans le domaine des fréquences, son spectre X(v) est entièrement caractérisé par ses échantillons X(k/T). Considérons alors l'échantillonnage temporel d'un tel signal avec une période  $T_e$ , ce qui fournit N échantillons  $x[n] = x(nT_e)$  en supposant que  $T = NT_e$ . Il lui correspond un spectre  $périodique\ X_{\text{éch}}(v)$  de période  $1/T_e$ , mais celui-ci est encore caractérisé par ses échantillons

$$X[k] \equiv X_{\text{\'ech}}(k/T_e)$$

puisque x(t) étant de durée limitée, x[n] l'est aussi. En tenant compte à la fois de la période  $1/T_e$  de  $X_{\text{éch}}(v)$  et de la relation  $T = NT_e$ , on en déduit que le signal échantillonné, qui est constitué de N valeurs x[n] en temps, est également décrit par N valeurs X[k] en fréquence. Ces valeurs s'écrivent

$$X[k] = \int_{-\infty}^{+\infty} x_{\text{éch}}(t) e^{-i2\pi kt/T_e} dt$$

et, en introduisant la définition du signal échantillonné en temps

$$x_{\text{\'ech}}(t) \equiv \sum_{n=0}^{N-1} x[n] \, \delta(t - nT_e),$$

on obtient la définition de la transformation de Fourier discrète ou série de Fourier discrète:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i2\pi nk/N}$$
.

En utilisant d'une façon analogue des arguments de périodicité et d'échantillonnage, on peut montrer que cette transformée admet une formule d'inversion qui s'écrit

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i2\pi nk/N}$$
.

Il importe de noter que, si les x[n] correspondent exactement aux échantillons de x(t) en temps, les X[k] sont en fréquence des échantillons de  $X_{\text{éch}}(v)$  et ne sont donc qu'une approximation de ceux de X(v). En effet, x(t) étant supposé à durée limitée, X(v) ne peut être à bande strictement limitée, d'où un phénomène inévitable de repliement spectral.

Introduction à la transformation de Fourier rapide. Si l'on calcule la transformée de Fourier discrète d'un signal de N points en appliquant directement la définition, chaque point en fréquence requiert N multiplications complexes et, puisqu'il y a N points à calculer, la complexité résultante est de l'ordre de  $N^2$  opérations. Il est cependant possible de réduire le nombre des opérations nécessaires au calcul en introduisant des algorithmes dits rapides tels que la transformée de Fourier rapide (TFR en français et FFT, pour fast Fourier transform, en anglais).

Pour en expliquer succinctement le principe, réécrivons la définition de la transformée de Fourier discrète selon

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{nk},$$

en introduisant les racines n-ièmes de l'unité

$$W_N \equiv e^{-i2\pi/N}$$
.

Si l'on suppose, ce que l'on fera, que  $N = 2^M$ , on peut alors former à partir des N échantillons de x[n] deux séquences, de longueur N/2 chacune, correspondant aux échantillons d'indices respectivement pairs et impairs. Ceci permet d'écrire

$$X[k] = \sum_{m=0}^{(N/2)-1} x[2m] \ (W_N^2)^{mk} + W_N^k \sum_{m=0}^{(N/2)-1} x[2m+1] \ (W_N^2)^{mk}.$$

On reconnait en fait dans ces sommations partielles les transformées de Fourier discrètes de chacune des deux séquences (d'indices respectivement pairs et impairs) obtenues par *décimation* en temps puisque

$$W_N^2 = e^{-i4\pi/N} = e^{-i2\pi/(N/2)} = W_{N/2}.$$

Chacune de ces transformées de Fourier discrètes est décrite par N/2 échantillons seulement. Il est alors possible d'itérer le processus en scindant de nouveau chaque demiséquence en deux séquences décimées de longueur N/4 chacune, et ceci jusqu'à ce qu'il ne reste plus qu'à calculer N/2 transformées à deux points, ce qui ne met en jeu que des coefficients +1 et -1. Cet algorithme ne requiert donc au total que  $M = \log_2 N$  étapes, chacune reposant sur N multiplications complexes. Sa complexité totale est de l'ordre de  $N \log_2 N$  opérations, au lieu de  $N^2$  par l'application directe de la définition. Ainsi, si l'on prend N = 1024, soit M = 10, le calcul d'une transformée de Fourier rapide met en jeu environ  $10^4$  opérations, contre environ $10^6$  pour le calcul direct : le facteur de gain est donc de l'ordre de 100, soit deux ordres de grandeur.

#### 4. MODELISATIONS ET CARACTERISATIONS

#### 4.1 Idées générales de modélisation

On a jusqu'à présent considéré les signaux (déterministes ou aléatoires) d'un point de vue essentiellement *non-paramétrique*, c'est-à-dire sans contrainte forte quant à une possible structure sous-jacente. L'idée qui sous-tend la notion de *modélisation* d'un signal est au contraire de faire l'hypothèse d'une telle structure, descriptible si possible *via* un petit nombre de *paramètres* qu'il s'agit alors d'identifier.

D'une manière très générale, une modélisation considère un signal comme la sortie d'un système. De façon à mettre la complexité du signal dans le système, il est souhaitable de disposer d'une entrée neutre. Une caractéristique de neutralité de l'entrée est la nature microscopique de sa corrélation : dans un contexte déterministe, ceci correspond à une impulsion de Dirac (le signal n'est alors autre que la réponse impulsionnelle d'un système) tandis que, dans un contexte aléatoire, on prend pour entrée un bruit blanc. En outre, on souhaite également que la modélisation conduise à une description parcimonieuse, c'est-à-dire nécessitant un nombre de paramètres nettement moindre que la taille de l'échantillon analysé.

Les techniques de modélisation sont très importantes au niveau des applications (qu'il s'agisse de compression d'information, de codage, de reconnaissance) puisqu'elles transforment un signal en un petit nombre de paramètres qui en décrivent la structure, et sur lesquels il est possible ensuite d'agir. Il est clair que l'efficacité d'une modélisation est d'autant plus grande qu'elle est adaptée aux données, le gain en performances pouvant se comprendre comme lié à un prise en compte plus importante de l'information *a priori* sur le signal analysé. En retour, il ne faut pas oublier que l'identification d'un modèle fournira toujours des valeurs (et même des valeurs optimales!), y compris dans les cas où le modèle choisi est très éloigné de la réalité. Il importe donc de veiller à ce qu'une modélisation voit l'identification associée à une *validation* du modèle envisagé.

# 4.2 L'exemple du modèle ARMA

Les modèles les plus simples sont *linéaires* et à temps discret. Un modèle très répandu, parce que représentatif d'un grand nombre de situations physiques, est le modèle dit autorégressif à moyenne ajustée ou ARMA. Sa forme est celle d'une équation aux différences

$$x[n] + \sum_{k=1}^{p} a_k x[n-k] = \sum_{k=0}^{q} b_k e[n-k],$$

dans laquelle l'entrée e[n] est un bruit blanc, c'est-à-dire un processus aléatoire (centré) tel que

$$\mathbf{E}\{x[n]|x[m]\} = \sigma^2 \delta[n-m].$$

Structures. Le membre de gauche correspond à une structure récursive d'ordre p et celui de droite à une structure transversale d'ordre q. Notons que l'on peut, sans perte de généralité, supposer que  $b_0 = 1$ . Dans le cas le plus général où les deux structures coexistent, le processus x[n] est dit ARMA(p,q). Si q=0, le processus, réduit à sa partie récursive, s'écrit

$$x[n] + \sum_{k=1}^{p} a_k x[n-k] = e[n]$$

et est dit AR(p). Si au contraire p=0, le processus résultant, purement transversal, s'écrit

$$x[n] = \sum_{k=0}^{q} b_k e[n-k]$$

et est dit MA(q).

 $P\hat{o}les$  et zéros. Par transformation en z de l'équation aux différences, on obtient aisément que

$$X(z) \left[ 1 + \sum_{k=1}^{q} a_k z^{-k} \right] = E(z) \left[ 1 + \sum_{k=1}^{q} b_k z^{-k} \right]$$

d'où la fonction de transfert rationnelle en z entre l'entrée et la sortie :

$$H(z) = \frac{B(z)}{A(z)}$$

en notant

$$A(z) = 1 + \sum_{k=1}^{q} a_k z^{-k}$$
;  $B(z) = 1 + \sum_{k=1}^{q} b_k z^{-k}$ .

D'un point de vue système, un processus ARMA est donc associé à un *filtre pôles-zéros*. Pour que celui-ci soit *stable* et *causal*, il faut que tous les pôles de H(z), donc les zéros de A(z) soient à l'intérieur du cercle unité. En effet, la causalité correspond à la propriété

$$h[n] = 0$$
 ,  $n < 0$ ,

ce qui impose au domaine de convergence d'être l'extérieur d'un cercle de rayon r. Par ailleurs, la stabilité au sens BIBO ("bounded input - bounded output") est associée à la propriété

$$|x[n]| < +\infty \implies |y[n]| < +\infty$$

où y[n] est la réponse à une entrée x[n]. Dans le cas d'un système linéaire, cette contrainte est équivalente à

$$\sum_{n=-\infty}^{+\infty} |h[n]| < +\infty ,$$

ce qui implique que le cercle unité soit inclus dans le domaine de convergence puisque

$$|H(z)| \le \sum_{n=-\infty}^{+\infty} |h[n]| z^{-n}| = \sum_{n=-\infty}^{+\infty} |h[n]| \text{ si } |z| = 1.$$

Par suite, un système causal et stable doit avoir une fonction de transfert en z qui converge pour |z| > r < 1, ce qui impose que ses pôles (c'est-à-dire les zéros de A(z) dans le cas d'une fonction de transfert rationnelle) soient à l'intérieur du cercle unité. S'il en est de même des zéros de B(z), le filtre est aussi à inverse stable et causal. Un filtre stable et causal ainsi que son inverse est dit à minimum de phase.

Comme la densité spectrale de puissance (en fréquence réduite) est reliée à la fonction de transfert en z par

$$\Gamma_{x}(v) = \sigma^{2} H(z) H^{*}(\frac{1}{z^{*}}) \Big|_{z=e^{i2\pi v}},$$

on en déduit qu'elle vaut

$$\Gamma_{\Lambda}(\nu) = \sigma^2 \left| \frac{B(e^{i2\pi\nu})}{A(e^{i2\pi\nu})} \right|^2$$
.

Son comportement est de type "pics et vallées", associés respectivement aux pôles et aux zéros de la fonction de transfert. Ces pics et ces vallées sont d'autant plus accentués que les singularités qui les génèrent sont plus près du cercle unité. Ce point souligne la grande souplesse des processus ARMA pour approcher une densité spectrale quelconque moyennant une disposition judicieuse de ses singularités (et un nombre suffisant de celles-ci). Cette souplesse permet également d'approcher un processus AR (resp. MA) par un modèle MA (resp. AR) dont l'ordre tend vers l'infini.

Identification AR. Soit à identifier un processus que l'on sait AR: il s'agit de trouver le jeu de paramètres  $(p, a_1 \dots a_p, \sigma^2)$  correspondant aux données. Supposons dans un premier que l'ordre p soit connu ou fixé a priori. En multipliant les deux membres de l'équation définissant le modèle par x[n-j] et en en prenant l'espérance mathématique, on obtient

$$\mathbf{E}\{x[n] \ x[n-j]\} + \sum_{k=1}^{p} \ a_k \, \mathbf{E}\{x[n-k] \ x[n-j]\} = \mathbf{E}\{e[n] \ x[n-j]\}.$$

Par suite, si l'on suppose que  $j \ge 1$ , la causalité du système permet d'annuler le membre de droite et d'obtenir le système, dit de Yule-Walker,

$$r_{x}[j] + \sum_{k=1}^{p} a_{k} r_{x}[j-k] = 0 , j \ge 1,$$

en notant  $r_x[j]$  la séquence des corrélations de x[n]. Ainsi, la connaissance de p valeurs consécutives de cette séquence permet d'identifier les p coefficients autorégressifs recherchés par résolution d'un système linéaire de p équations à p inconnues. La puissance  $\sigma^2$  peut quant à elle s'obtenir en faisant j=0 puisque l'on obtient alors

$$\sigma^2 = r_x[0] + \sum_{k=1}^{p} a_k r_x[k].$$

Le regroupement de ces deux équations d'identification conduit à l'équation matricielle unique

$$\mathbf{R} \ \mathbf{a} = \mathbf{p} \quad \text{avec} \quad \mathbf{a} = [1 \ a_1 \ \dots \ a_p]^{\mathrm{T}}, \quad \mathbf{p} = [\sigma^2 \ 0 \ \dots \ 0]^{\mathrm{T}} \quad \text{et} \quad [\mathbf{R}]_{ij} = r_x [i - j].$$

Prédiction linéaire. Soit maintenant une séquence de données quelconque x[n]. On peut se poser le problème de décrire chacun des termes de cette séquence par une combinaison linéaire de ses p prédecesseurs (avec des poids  $a_1 \dots a_p$ ), et ceci à une erreur de prédiction près, de puissance P: il s'agit alors de trouver le "meilleur" jeu de paramètres  $(p, a_1 \dots a_p, P)$  correspondant à ces données, c'est-à-dire celui qui minimise P. Par construction de cette prédiction linéaire, chaque valeur x[n] est estimée par une quantité

$$\hat{x}[n] = -\sum_{k=1}^{p} a_k x[n-k]$$

avec une erreur de prédiction

$$e[n] = x[n] - \hat{x}[n] = x[n] + \sum_{k=1}^{p} a_k x[n-k]$$
.

Si l'on compare cette équation à celle définissant un modèle AR, on voit que la prédiction finéaire est équivalente à l'ajustement d'un modèle AR aux données, à la condition que l'erreur de prédiction ait les propriétés d'un bruit blanc. La "meilleure" prédiction linéaire consisterait ainsi à identifier le filtre inverse à appliquer aux données pour les décorréler. Le critère usuellement retenu est en fait la minimisation de l'erreur quadratrique moyenne  $P = \mathbb{E}\{e^2[n]\}$ . Puisque l'estimation x[n] est engendrée linéairement à partir de p observations, l'estimée la meilleure (c'est-à-dire celle qui minimise p) n'est autre que la projection orthogonale de x[n] sur l'espace engendré par les observations. Le jeu des coefficients optimaux est donc celui pour lequel (principe d'orthogonalité) l'erreur est orthogonale aux observations :

$$E\{e[n] \ x[n-j]\} = 0 , j \ge 1$$

et la puissance de l'erreur minimum associée vaut

$$P = \mathbf{E}\{e[n] \ (x[n] - x[n])\} = \mathbf{E}\{e[n] \ x[n]\}.$$

Il résulte clairement de ces deux équations que l'on retrouve le système de Yule-Walker déjà rencontré. Algorithme de Levinson. Afin de résoudre le problème posé, la solution la plus immédiate serait d'inverser directement le système de Yule-Walker, système linéaire de p équations à p inconnues. Cette approche se heurte cependant à plusieurs difficultés :

- le coût calculatoire d'une telle inversion est élevé puisqu'il est en  $O(p^3)$ ;
- la séquence de corrélation, nécessaire pour l'inversion, n'est généralement qu'imparfaitement connue, ce qui peut résulter en une inversion numériquement instable;
- si l'ordre du modèle n'est pas connu *a priori*, plusieurs ordres doivent être essayés, mais au prix de recommencer l'inversion à chaque essai.

L'ensemble de ces remarques suggère de rechercher un algorithme d'inversion implicite qui soit *récursif* sur l'ordre, c'est-à-dire tel que la solution à l'ordre p puisse s'obtenir en fonction de la solution à l'ordre p-1 et de  $r_x[p]$ , l'information supplémentaire nécessaire. En explicitant la dépendance des paramètres par rapport à l'ordre du modèle, ceci revient à chercher une solution de la forme

$$\begin{cases} \{a_k(p)\} = f(\{a_k(p-1)\}, r_x[p]) \\ P(p) = g(P(p-1), r_x[p], \{a_k(p)\}) \end{cases}$$

Un bénéfice supplémentaire est de fournir pour chaque ordre une puissance d'erreur minimale, ce qui permet, en suivant cette puissance d'erreur en fonction de l'ordre, de déterminer un ordre optimal associé à la plus petite des erreurs minimales.

Une solution de ce type est fournie par l'algorithme de Levinson-Durbin, dont le principe repose sur la structure Toeplitz de la matrice de corrélation à inverser, et dont la complexité se trouve réduite à  $O(p^2)$ . L'algorithme se déroule de la façon suivante :

Initialisation 
$$a_{1}(1) = -\frac{r_{x}[1]}{r_{x}[0]} \; ; \; P(1) = r_{x}[0] \left(1 - a_{1}^{2}(1)\right)$$

$$p = 1$$

$$Récursion \qquad Q(p) = r_{x}[p+1] + r_{x}[p] \; a_{1}(p) + \dots + r_{x}[1] \; a_{p}(p)$$

$$k(p) = -\frac{Q(p)}{P(p)}$$

$$P(p+1) = P(p) \left(1 - k^{2}(p)\right)$$

$$a_{i}(p+1) = a_{i}(p) + k(p) \; a_{p-i+1}(p) \; \; , \; 1 \le i \le p$$

$$a_{p+1}(p+1) = k(p)$$

$$p = p + 1$$

Choix de l'ordre. L'algorithme de Levinson met en évidence le fait que l'erreur de prédiction est une fonction non croissante de l'ordre qui ne possède pas à proprement parler de minimum absolu permettant de sélectionner un ordre optimal qui serait associé à la plus petite erreur parmi tous les modèles possibles. Le suivi de l'erreur de prédiction en fonction de l'ordre fournit néanmoins des informations intéressantes.

Supposons tout d'abord que l'on cherche à identifier un processus effectivement AR, mais d'ordre r inconnu, et supposons de plus que sa fonction de corrélation soit parfaitement connue. On observera alors une décroissance de l'erreur de prédiction jusqu'à ce que l'on atteigne l'ordre effectif du modèle. Pour cette valeur, les coefficients identifiés sont alors exactement égaux aux coefficients effectifs et la puissance d'erreur vaut alors

$$P_r = \sigma^2$$
.

Si l'on fait alors croître davantage l'ordre, on aura pour p = r+1,

$$\begin{cases} a_i(r+1) = a_i(r) &, 1 \le i \le p \\ a_{r+1}(r+1) = 0 \end{cases}$$

et par suite, comme  $k(r) = a_{r+1}(r+1) = 0$ , on en déduit que

$$P(r+1) = P(r).$$

Il en sera de même pour toute valeur p > r: l'ordre du modèle peut ainsi être identifié en retenant la première valeur pour laquelle la puissance d'erreur se stabilise à une valeur minimum constante.

En fait, si le processus analysé n'est pas strictement AR, il est formellement représentable par un modèle AR d'ordre infini et le minimum de la puissance d'erreur est rejeté à l'infini. Il en est de même si l'identification est conduite à partir d'estimées des valeurs de la corrélation, ce qui est pratiquement toujours le cas. Dans ce dernier cas, lorsque l'on travaille sur une séquence d'échantillons de taille finie N, on est en fait confronté à un compromis biais-variance. En effet, plus l'ordre augmente, plus la puissance d'erreur diminue et plus la séquence prédite se rapproche de la réalisation observée (diminution du biais). Cependant, augmenter l'ordre du modèle conduit dans le même temps à utiliser des valeurs de la fonction de corrélation pour des retards de plus en plus grands, et donc de moins en moins bien estimées puisqu'elles le sont sur de moins en moins d'échantillons (augmentation de la variance). Une solution de compromis est

d'introduire une fonction de pénalité sur la puissance d'erreur de telle sorte que la fonction résultante passe par un minimum que l'on choisira comme ordre optimal. Un choix courant est le critère dit FPE (Final Prediction Error) d'Akaike qui s'écrit

$$FPE(p) = \frac{N+p}{N-p}P(p) = P(p) + 2\frac{p}{N-p}$$

ou encore le critère dit AIC (Akaike Information Criterion):

$$AIC(p) = N \ln P(p) + 2p$$
.

(On vérifie d'ailleurs que les deux critères tendent à être équivalents pour de grandes tailles d'échantillons puisque l'on obtient alors  $AIC(p) = N \ln FPE(p)$ .)

Modularité et treillis. Pour un ordre p donné, un modèle AR se caractérise par le jeu des coefficients autorégressifs  $\{a_i(p), 1 \le i \le p\}$ . Lorsque l'on change d'ordre pour passer de p à p+1, il existe bien une relation de passage qui permet de déduire la collection  $\{a_i(p+1), 1 \le i \le p+1\}$  de la précédente, mais les nouveaux coefficients individuels  $a_i(p+1)$  n'ont a priori rien à voir avec les anciens  $a_i(p)$ . En termes de système à structure récursive, tous les coefficients du filtre sont à changer lorsque l'on change d'ordre. Une caractérisation plus intéressante du même modèle est cependant possible par l'intermédiaire de la collection des coefficients de réflexion  $\{k(i) = a_i(i), 1 \le i \le p\}$ , conduisant cette fois-ci à une possibilité de structure modulaire puisqu'augmenter l'ordre correspond alors à ajouter un coefficient, sans rien changer aux précédents.

Cette structure est de type *treillis* et peut être consyruite de la façon suivante. On a jusqu'à présent considéré l'identification d'un modèle AR en liaison avec le problème de prédiction à un pas, celle-ci étant par nature supposée *directe*, c'est-à-dire du passé vers le présent :

$$\hat{e}_p[n] = x[n] - \hat{x}_p[n] = x[n] + \sum_{k=1}^p \hat{a}_k(p) x[n-k].$$

On peut cependant imaginer également une erreur de prédiction *rétrograde*, dirigée cette fois du futur vers le présent :

$$\overset{\checkmark}{e}_{p}[n] = x[n] - \overset{\checkmark}{x}_{p}[n] = x[n] + \sum_{k=1}^{p} \overset{\checkmark}{a}_{k}(p) \ x[n-k] .$$

Il suit alors de la stationnarité du processus que la minimisation de la puissance de ces erreurs de prédiction définit en fait le même jeu de coefficients que nous noterons simplement  $a_k(p)$ . En explicitant les erreurs de prédiction, tant directe que rétrograde, on peut montrer que l'on obtient les deux équations

$$\hat{e}_p[n] = \hat{e}_{p-1}[n] + k(p-1) \, \stackrel{\checkmark}{e}_{p-1}[n-p]$$

$$\overset{\checkmark}{e}_{p}[n-p] = \overset{\checkmark}{e}_{p-1}[n-p] + k(p-1) \overset{\land}{e}_{p-1}[n]$$

Avec l'initialisation  $\hat{e}_0[n] = \stackrel{\checkmark}{e}_0[n] = x[n]$ , ceci définit une structure modulaire de filtre (inverse) en treillis, caractérisée par la mise en cascade de cellules toutes identiques et paramétrées chacune par un seul coefficient, le coefficient de réflexion associé au numéro de la cellule. On peut définir de la même façon une structure de filtre direct

$$\stackrel{\wedge}{e}_{p-1}[n] = \stackrel{\wedge}{e}_{p}[n] - k(p-1) \stackrel{\vee}{e}_{p-1}[n-p]$$

$$\overset{\checkmark}{e}_{p-1}[n-p] = \overset{\checkmark}{e}_{p}[n-p] - k(p-1) \, \overset{\wedge}{e}_{p-1}[n]$$

qui, avec l'initialisation  $\stackrel{\wedge}{e}_p[n] = e[n]$ , bruit blanc, génère en sortie la séquence  $x[n] = \stackrel{\wedge}{e}_0[n] = \stackrel{\vee}{e}_0[n]$ .

#### 4.3 Eléments d'analyse spectrale

L'analyse spectrale a pour but de caractériser le contenu fréquentiel d'un signal. Si, dans le cas des signaux certains, l'estimation de la densité spectrale d'énergie passe simplement par le calcul d'une transformée de Fourier (suivie d'un module carré), dans le cas des signaux aléatoires, il s'agit d'estimer la fonction de densité spectrale de puissance, quantité liée à une moyenne d'ensemble.

Dans le cas d'un signal aléatoire à temps discret x[n], il est possible de réécrire la définition de la densité spectrale de puissance selon

$$\Gamma_{\mathcal{X}}(v) = \lim_{N \to \infty} \mathbf{E} \left\{ \frac{1}{2N+1} \left| \sum_{n=-N}^{+N} x[n] e^{-i2\pi nv} \right|^2 \right\}$$

Ergodisme. Le plus souvent, les moyennes d'ensemble sont inaccessibles dans la mesure où l'on ne dispose que d'une réalisation observée d'un processus aléatoire. Afin de remonter à des caractéristiques d'ensemble à partir de la seule connaissance d'une réalisation, il est nécessaire de faire une hypothèse quant au processus analysé: on devra supposer que le processus est *ergodique*, ce qui signifiera que la distribution instantanée de toutes les valeurs possibles du processus pourra se retrouver par moyennes temporelles sur un temps suffisamment long pour qu'une réalisation particulière ait le temps de "visiter" ces mêmes valeurs avec une distribution identique.

Méthodes "Fourier". Une première famille de méthodes d'estimation spectrale consiste à utiliser les définitions de la densité spectrale basées sur la transformation de Fourier. Par application du théorème de Wiener-Khintchine, ceci peut revenir à calculer la transformée de Fourier d'une estimée de la fonction de corrélation (on parle alors de corrélogramme), ou encore à estimer directement la moyenne d'ensemble inaccessible mentionnée précédemment (on parle alors de périodogramme), sous hypothèse dans les deux cas d'ergodisme et de stationnarité.

En supposant que l'observation consiste en N valeurs x[n],  $n = 0 \dots N-1$ , et en négligeant l'opération de moyenne d'ensemble, le *périodogramme* est défini par la densité empirique

$$P_{\mathcal{X}}(v) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-i2\pi nv} \right|^2.$$

Si l'on obtient bien le résultat souhaitable

$$\lim_{N \to \infty} \mathbf{E} \left\{ P_{\chi}(v) \right\} = \Gamma_{\chi}(v)$$

(ce qui signifie qu'en augmentant le nombre des observations, l'estimée tend en moyenne vers la valeur vraie), on peut montrer que l'on a par contre

var 
$$\left\{ P_{\chi}(v) \right\} \approx \Gamma_{\chi}^{2}(v)$$
,

et ceci quel que soit le nombre des observations, aussi grand soit-il. Ce résultat signific que le périodogramme n'est pas un estimateur consistant de la densité spectrale, ce qui est bien sûr dû au fait que la moyenne d'ensemble, nécessaire à la définition, n'a pas été prise en compte dans l'estimation.

Une façon d'améliorer cette situation de grande variabilité statistique (l'écart-type est de l'ordre de grandeur de la quantité à estimer!) est de faire usage de l'hypothèse d'ergodisme et de moyenner plusieurs périodogrammes calculés sur des sous-segments (supposés indépendants) prélevés dans les données. En supposant pour simplifier que l'on dispose de N = KL données et que l'on choisisse de les segmenter en K blocs de la forme  $x_k[n] = x[n+kL]$ ,  $n = 0 \dots L-1$ ;  $k = 0 \dots K$ , le périodogramme moyenné est alors défini par

$$PM_{X}(v) = \frac{1}{KL} \sum_{k=0}^{K-1} \left| \sum_{n=0}^{L-1} x_{k}[n] e^{-i2\pi nv} \right|^{2}.$$

Comme on pouvait s'y attendre, on peut montrer que cet estimateur modifié est de variabilité plus réduite, puisque l'on a

$$\operatorname{var} \left\{ PM_{\mathcal{K}}(v) \right\} = \frac{1}{K} \operatorname{var} \left\{ P_{\mathcal{K}}(v) \right\} \ ,$$

ce qui signific que la variance est d'autant plus réduite que le nombre K de blocs (indépendants) est plus grand. On ne peut cependant impunément augmenter K car on a dans le même temps

$$\mathbf{E} \left\{ P M_{\chi}(v) \right\} = \int_{-1/2}^{+1/2} W_L(v - \xi) \ \Gamma_{\chi}(\xi) \ \mathrm{d}\xi \ ,$$

avec

$$W_L(v) = \frac{1}{L} \left( \frac{\sin \pi L v}{\sin \pi v} \right)^2,$$

ce qui montre que le biais augmente puisque, à taille d'observation N = KL fixée, augmenter K revient nécessairement à diminuer L, et donc à dégrader la résolution fréquentielle de l'estimateur. L'estimation spectrale par périodogrammes moyénnés est donc confrontée à un compromis biais-variance.

Méthodes "non-Fourier". L'analyse spectrale classique (de type périodogramme ou corrélogramme) repose de manière fondamentale sur les propriétés de la transformation

de Fourier, ainsi que sur les hypothèses qui valident son emploi. D'une manière schématique, on peut considérer que les performances de ces méthodes traditionnelles deviennent insuffisantes dans deux circonstances principales :

- lorsque les signaux analysés ne sont connus que sur une courte durée d'observation, ou sur un nombre réduit d'échantillons;
  - lorsque les signaux analysés sont *non-stationnaires* .

En effet, dans le premier cas, disposer de peu de données réduit la résolution fréquentielle et augmente la variabilité de l'estimation puisque moins de moyennes temporelles indépendantes sont disponibles.

Dans le deuxième cas, c'est la définition même des notions de fonction de corrélation et de densité spectrale qui n'est plus pertinente, puisqu'elle ne peut prendre en compte une évolution temporelle des caractéristiques de deuxième ordre ou fréquentielles du signal. Pour un signal faiblement non-stationnaire, une solution intuitive est de s'intéresser à des estimations locales. Cependant, dès que les non-stationnarités deviennent importantes, la faible longueur des segments temporels sur lesquels l'hypothèse de quasi-stationnarité peut être considérée comme valide ramène au problème des échantillons courts.

Les améliorations à apporter à l'analyse spectrale classique, de façon à prendre en compte effectivement les situations décrites précédemment, sont de deux types :

- pour les échantillons courts, incorporer les *a priori* que l'on peut avoir sur la *structure* des signaux analysés, information qui n'est pas prise en compte par les analyses non-paramétriques à base de transformée de Fourier : ce sont les méthodes paramétriques à base de *modèles*;
- pour les signaux non-stationnaires, modifier les définitions existantes pour les adapter à la *nature* des signaux analysés : ce sont les méthodes *temps-fréquence*, sur lesquelles on reviendra plus longuement au § 5.

L'exemple de l'estimation spectrale par modèle AR. Dans le cas du périodogramme, l'estimateur de Fourier fait implicitement l'hypothèse que, en-dehors de l'intervalle d'observation, les données sont nulles. Cette hypothèse se transporte à la fonction de corrélation qui, en-dehors d'un support lié à la taille des données, est elle aussi supposée nulle. C'est cette situation en fait que l'on peut tenir pour responsable de la faible résolution fréquentielle du périodogramme dans le cas d'échantillons courts. Les seules améliorations possibles nécessitent de rajouter de l'a priori sur la structure des données, ce que l'on peut faire par exemple en ajustant un modèle paramétrique sur

les données observées et en faisant l'hypothèse que le même modèle reste valide endehors de l'intervalle d'observation.

On a vu précédemment que, dans le cas d'un modèle AR d'ordre p, les p premières valeurs de la fonction de corrélation suffisent pour identifier le modèle. De plus, si l'on ne connaît que ces p premières valeurs (cas des échantillons courts), l'identification des coefficients AR permet une *extrapolation* de la fonction de corrélation au-delà du retard p puisque le système de Yule-Walker reste valable pour j > p, conduisant à :

$$r_{x}[p+j] = -\sum_{k=1}^{p} a_{k} r_{x}[p+j-k]$$
 ,  $j \ge 1$ ,

La fonction de corrélation ainsi extrapolée assure donc à l'estimation spectrale correspondante une meilleure résolution fréquentielle.

#### 5. FOURIER ET SES LIMITATIONS

### 5.1 Difficultés d'interprétation

Quoique parfaite du point de vue mathématique, la transformation de Fourier conduit bien souvent à des difficultés d'interprétation physique. Ainsi, si l'on s'en reporte à la définition usuelle, il est clair que le calcul d'une valeur fréquentielle X(v)nécessite la connaissance de toute l'histoire temporelle du signal, éventuellement de -∞ à +∞. Réciproquement, la transformation inverse de Fourier montre que toute valeur du signal à un instant t peut être vue comme superposition infinie d'exponentielles complexes, c'est-à-dire d'ondes éternelles parfaitement délocalisées dans le temps. Si ce point de vue mathématique peut dans certains cas (situations "quasi monochromatiques", de régime permanent, ...) révéler des propriétés effectives d'un signal, il peut également, dans d'autres, travestir la réalité physique. C'est ce qui se passe par exemple dans le cas de signaux transitoires que l'on sait être nuls en dehors d'un certain support temporel (par exemple par l'action de brancher puis débrancher un appareil). La nullité de ces valeurs est certes bien rendue par l'analyse de Fourier, mais d'une manière artificielle : elle correspond au résultat de la superposition d'une infinité d'ondes virtuelles qui interféreraient entre elles de telle sorte qu'elles s'annulent mutuellement. Sur le support où le signal est nul, on a donc une situation de zéro "dynamique" (des ondes existeraient, qui interféreraient pour fournir une contribution résultante nulle), en contradiction

d'interprétation avec la situation physique réelle d'un zéro "statique" (le signal n'existe pas).

D'une façon liée, la représentation de Fourier décompose un signal d'énergie finie en une combinaison linéaire d'éléments constitutifs dont aucun n'est d'énergie finie.

Ce type de problèmes suggère, dans les situations évolutives pour lesquelles la transformation de Fourier perd de sa pertinence physique, de lui chercher un substitut qui par exemple puisse concilier description fréquentielle et localisation.

# 5.2 Inégalités temps-fréquence

On a déjà évoqué ce que l'on appelle la dualité temps-fréquence dans le cas de la transformée de Fourier d'une gaussienne et, à la limite, d'une distribution de Dirac. Ce comportement est en fait tout-à-fait général et rentre dans le cadre des inégalités de type Heisenberg-Gabor qui bornent inférieurement le produit durée-bande d'un signal.

Afin de quantifier cette affirmation, considérons un signal centré en ce sens que

$$\int_{-\infty}^{+\infty} t |x(t)|^2 dt = \int_{-\infty}^{+\infty} v |X(v)|^2 dv = 0$$

et définissons respectivement sa durée et sa bande utiles par les épanouissements temporel et fréquentiel

$$\Delta t^2 = \int_{-\infty}^{+\infty} t^2 |x(t)|^2 dt \quad \text{et} \quad \Delta v^2 = \int_{-\infty}^{+\infty} v^2 |X(v)|^2 dv.$$

Si l'on forme alors la quantité

$$I \equiv \int_{-\infty}^{+\infty} t \, x^*(t) \, x'(t) \, dt$$

où x'(t) désigne la dérivée de x(t), il est facile de vérifier en utilisant l'inégalité de Cauchy-Schwartz que

$$|I|^2 \le \int_{-\infty}^{+\infty} t^2 |x(t)|^2 dt \cdot \int_{-\infty}^{+\infty} |x'(t)|^2 dt = \Delta t^2 \cdot 4\pi^2 \Delta v^2.$$

On a par ailleurs

$$|I|^2 = (R\acute{e}{I})^2 + (Im{I})^2 \ge (R\acute{e}{I})^2$$

et, en intégrant par parties,

$$I = -E_x + I^*.$$

On en déduit que

$$Ré\{I\} = -\frac{E_x}{2}$$

et, par suite,

$$\Delta t \ \Delta v \ge \frac{E_x}{4\pi}$$
.

La borne inférieure est atteinte lorsque l'inégalité de Cauchy-Schwartz utilisée se transforme en égalité, c'est-à-dire lorsque les signaux sur lesquels elle porte sont proportionnels. Ceci correspond à l'équation différentielle

$$t x(t) = k x'(t)$$

dont la solution est fournie par la classe des signaux gaussiens.

La signification la plus simple de l'inégalité d'Heisenberg-Gabor est que la description temporelle et fréquentielle d'un signal, telle qu'elle est fournie par la transformation de Fourier, ne peut être arbitrairement précise dans les deux domaines.

Il est à noter que d'autres formes d'inégalités temps-fréquence existent, qui concernent l'impossibilité qu'il y a pour un signal de confiner parfaitement son énergie sur des supports temporel et fréquentiel compacts, même arbitrairement grands (théorie de Slepian-Landau-Pollak).

# 5.3 Fréquence instantanée

La notion de stationnarité est bien définie pour les processus aléatoires. Dans le cas des signaux certains, une notion apparentée existe, qui est en fait liée à une idée de *régime* permanent et de stabilité d'une fréquence au cours du temps. Sa définition nécessite de recourir au concept de *signal analytique* qui s'appuie lui-même sur une transformation spécifique : la *transformation de Hilbert*.

Par définition, la transformation de Hilbert est donnée par la formule

$$H\{x(t)\} \equiv \frac{1}{\pi} \operatorname{vp} \int_{-\infty}^{+\infty} \frac{x(u)}{t-u} \, \mathrm{d}u,$$

où l'intégrale doit être prise en valeur principale de Cauchy. Cette transformation constitue une représentation dans la mesure où

$$H\{H\{x(t)\}\} = -x(t).$$

La réponse impulsionnelle d'un filtre de Hilbert est la fonction  $1/\pi t$ , d'où l'on déduit que son gain complexe vaut -i sgn v. Le résultat en est qu'un signal monochromatique cos  $2\pi v_0 t$  voit son spectre

$$\frac{1}{2}\left[\delta(v-v_0)+\delta(v+v_0)\right]$$

transformé en

$$-\frac{i}{2} [\delta(v - v_0) - \delta(v + v_0)] = \frac{1}{2i} [\delta(v - v_0) - \delta(v + v_0)],$$

ce qui est le spectre de sin  $2\pi v_0 t$ . Le filtre de Hilbert agit ainsi comme un déphaseur pur de  $\pi/2$ : c'est un filtre de quadrature.

Ceci permet de définir de manière canonique des notions comme celles d'amplitude ou de fréquence instantanée(s) d'un signal modulé. Considérons en effet le signal monochromatique

$$x(t) = a \cos 2\pi v_0 t$$
.

Il est possible de lui associer une représentation complexe grâce au vecteur tournant

$$z_{\nu}(t) = a e^{i2\pi\nu_0 t}$$

dans lequel la composante imaginaire est en quadrature avec la composante réelle :

$$Im\{z_x(t)\} = a \sin 2\pi v_0 t = H\{a \cos 2\pi v_0 t\} = H\{R\acute{e}\{z_x(t)\}\}.$$

L'intérêt de cette représentation est de fournir des indications relatives à l'amplitude et à la vitesse de rotation (pulsation) du vecteur tournant de manière unique grâce respectivement au module et à la phase

$$a = |z_r(t)|$$
;

$$v_0 = \frac{1}{2\pi} \frac{\mathrm{d}}{\mathrm{d}t} \arg \{z_x(t)\}.$$

Intuitivement, lorsqu'un signal est modulé, c'est-à-dire lorsque son amplitude et sa fréquence sont réputées varier au cours du temps, on pourrait être tenté de généraliser le cas monochromatique en remplaçant  $a\cos 2\pi v_0 t$  par  $x(t)\equiv a(t)\cos \varphi(t)$ . Ceci ne constitue cependant pas une représentation admissible d'un signal quelconque. Pour s'en convaincre, il suffit d'introduire une fonction arbitraire b(t) telle que 0 < b(t) < 1. On a alors tout aussi bien

$$x(t) = \frac{a(t)}{b(t)}b(t)\cos\varphi(t),$$

ce que l'on peut écrire

$$x(t) \equiv a'(t) \cos \varphi'(t) \quad \text{avec} \quad \begin{cases} a'(t) = \frac{a(t)}{b(t)} \\ \varphi'(t) = \arccos (b(t) \cos \varphi(t)) \end{cases}$$

Une solution à ce problème consiste à généraliser, non pas directement la définition d'une onde monochromatique, mais la façon dont on peut déduire ses caractéristiques d'une représentation complexe convenable. Ainsi, si l'on part du modèle du vecteur tournant en le rendant dépendant du temps et si l'on impose que ce modèle complexe satisfasse à des relations de quadrature entre parties réelle et imaginaire, la solution consiste à complexifier un signal réel x(t) en lui adjoignant une partie imaginaire égale à son signal en quadrature :

$$z_x(t) = x(t) + i H\{x(t)\}.$$

Le signal résultant est dit *analytique*. Ainsi défini, il admet une écriture *unique* en termes de module et de phase et il devient possible de définir les notions d'*amplitude* et de *fréquence instantanées* par respectivement

$$a(t) = |z_{\mathbf{r}}(t)|;$$

$$v_t = \frac{1}{2\pi} \frac{\mathrm{d}}{\mathrm{d}t} \arg \{z_x(t)\}.$$

D'un point de vue spectral, le signal analytique est relié au signal réel dont il est issu par

$$Z_x(v) = 2 u(v) X(v).$$

Ceci revient à supprimer les fréquences négatives du spectre original, ce qui ne diminue en rien l'information puisque, pour un signal  $x(t) \in \mathbb{R}$ , on a la relation

$$X(-v) = X^*(v).$$

Si, par structure, la partie réelle d'un signal analytique

$$z(t) = a(t) e^{i\varphi(t)}$$

s'écrit naturellement

$$R\acute{c}\{z(t)\} = a(t)\cos\varphi(t),$$

il importe d'avoir présent à l'esprit que le signal analytique associé à un signal modulé selon a(t) cos  $\varphi(t)$  n'a a priori aucune raison d'avoir une écriture exponentielle dans laquelle les mêmes a(t) et  $\varphi(t)$  seraient respectivement les termes de module et de phase. D'un point de vue physique, on se rapprochera cependant d'autant plus de cette situation que les effets de modulations seront faibles. En particulier, si a(t) est de type passe-bas tel que son spectre soit limité à [-B, +B] et si cos  $\varphi(t)$  est de type passe-haut tel que son spectre occupe le domaine  $]-\infty, -B'] \cup [B', +\infty[$  avec B' > B, alors (théorème de Bedrosian)

$$H{a(t)\cos\varphi(t)} = a(t) H{\cos\varphi(t)}.$$

Ces seules considérations spectrales ne garantissent cependant pas l'écriture exponentielle du signal analytique puisqu'il faudrait en outre que

$$H\{\cos \varphi(t)\} = \sin \varphi(t),$$

ce qui n'est pas vérifié dans le cas général mais l'est approximativement dans le cas quasi-monochromatique pour lequel la largeur de bande est très faible vis-à-vis de la fréquence centrale du spectre.

# 6. TEMPS-FREQUENCE

# 6.1 Approches linéaires

L'analyse de Fourier décompose *linéairement* un signal sur des éléments parfaitement localisés en fréquence (les ondes monochromatiques), l'échantillonnage sur des éléments parfaitement localisés en temps (les impulsions idéales). Ce sont deux approches *a priori* qui, à l'inverse des décompositions de type Karhunen-Loève, ne cherchent pas à s'adapter à un signal particulier, mais davantage à être d'un emploi "universel". Tout en conservant ce point de vue méthodologique, il est alors possible d'imaginer d'autres décompositions *a priori*, qui par exemple cherchent à concilier les approches temporelle et fréquentielle.

Le point de vue général est de projeter le signal analysé sur un signal analysant  $h(u; t, \lambda)$  dépendant de deux variables:

- le *temps t*, puisque le signal est supposé non-stationnaire;
- une variable auxiliaire  $\lambda$  dont on veut suivre l'évolution temporelle :

$$x(u) \rightarrow \langle x(.), h(.;t,\lambda) \rangle$$
.

Si la transformation ainsi construite est une représentation admissible du signal, on peut espérer qu'elle admette une inversion suivant :

$$x(u) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle x(.), h(.;t,\lambda) \rangle h(u;t,\lambda) d\mu(t,\lambda) ,$$

où l'intégration est faite relativement à une mesure naturelle associée à la transformation.

L'interprétation de cette opération est que le signal analysé admet une représentation comme superposition de "briques élémentaires" (les signaux analysants *h*) dont le poids est égal à la projection.

De façon à réduire la complexité de la décomposition, le signal analysant est généralement construit à partir d'un signal élémentaire unique, h(u), par l'action d'un groupe de transformation. Le choix du signal élémentaire peut répondre à plusieurs objectifs :

- on possède un modèle vraisemblable pour les composantes élémentaires du signal : le choix naturel est d'y adapter h(u), ce qui devrait minimiser le nombre des projections non-négligeables ;
- on ne possède pas un tel modèle : le choix le plus neutre est d'utiliser un signal d'encombrement minimum dans le plan transformé.

Les différentes décompositions possibles dépendent du choix du groupe de transformation. Nous nous intéresserons ici à deux cas particuliers : la transformée de Fourier à court terme et la transformée en ondelettes.

Transformée de Fourier à court terme. Comme premier exemple, considérons le groupe des translations en temps et en fréquence (ou groupe de Weyl-Heisenberg), caractérisé par :

$$h(u; t, v) = h(u - t) e^{i2\pi vu}.$$

Dans ce cas, la variable auxiliaire  $\lambda = v$  a un statut physique de *fréquence* et tous les signaux analysants se déduisent du signal élémentaire h(u) par une translation dans le plan temps-fréquence. La projection associée prend la forme:

$$F_X(t, v) = \int_{-\infty}^{+\infty} x(u) h^*(u - t) e^{-i2\pi v u} du$$

ce qui correspond à une opération de transformée de Fourier à court terme si le signal élémentaire h(u) est une fonction de pondération de support faible devant la durée totale du signal.

Cette opération représente bien une décomposition (continue) du signal puisque, si h(u) est d'énergie unité, il est possible de l'inverser selon:

$$x(u) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_X(t, v) h(u - t) e^{i2\pi v u} dt dv.$$

La transformée de Fourier à court terme considère implicitement un signal non stationnaire comme une succession de situations quasi stationnaires, à l'échelle de la fenêtre à court terme h(u). La résolution temporelle d'une telle analyse est fixée par la largeur de cette fenêtre, la résolution fréquentielle étant fixée par la largeur de sa transformée de Fourier. Ces deux largeurs étant antagonistes, on se trouve donc en présence d'un compromis entre les résolutions temporelle et fréquentielle. En effet :

• pour un signal fortement non-stationnaire, une bonne résolution temporelle est requise, ce qui impose de travailler avec un fenêtre h(u) courte, limitant en retour la résolution fréquentielle;

• réciproquement, si une analyse fréquentielle fine est nécessaire, une fenêtre h(u) longue doit être utilisée, ce qui a le double effet de moyenner les contributions fréquentielles sur la durée de la fenêtre et de réduire la résolution temporelle.

Bien que fournissant une représentation admissible d'un signal non stationnaire, on voit donc que la transformée de Fourier à court terme ne permet pas une analyse à la fois *locale* en temps et *précise* en fréquence.

*Transformée en ondelettes*. Considérons maintenant l'exemple du groupe des translations et changements d'échelle en temps (ou *groupe affine*), caractérisé par :

$$h(u; t, a) = \frac{1}{\sqrt{a}} h\left(\frac{u - t}{a}\right), a > 0.$$

La variable auxiliaire  $\lambda = a$  a cette fois le statut physique d'un facteur d'échelle et tous les signaux analysants de déduisent de l'ondelette h(u) par l'intermédiaire d'une translation en temps et d'un changement d'échelle.

La projection associée prend la forme :

$$T_X(t, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u) h^* \left(\frac{u - t}{a}\right) du,$$

quantité qui est appelée transformée en ondelettes du signal.

D'après cette définition, les fréquences élevées sont associées à des échelles d'observation petites et elles sont obtenues à partir de h(u) par une opération de compression. Ceci nécessite donc que l'ondelette élémentaire comporte au moins quelques oscillations. En fait, sa structure imposée est de correspondre en fréquence à un signal de type passe-bande :

$$H(0) = 0$$
 ;  $\int_{-\infty}^{+\infty} |H(v)|^2 \frac{dv}{|v|} = 1$ .

Si cette condition d'admissibilité est remplie, la formule d'inversion s'écrit :

$$x(u) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} T_{\chi}(t, a) \frac{1}{\sqrt{a}} h\left(\frac{u-t}{a}\right) \frac{\mathrm{d}t \, \mathrm{d}a}{a^2}.$$

On retrouve le rôle de "briques de base" joué par les différentes ondelettes analysantes, celles-ci se déduisant d'une ondelette élémentaire par, outre des translations temporelles, des dilatations et des compressions.

Bien qu'elle ne fasse pas usage de la transformation de Fourier, l'analyse en ondelettes a aussi des résolutions temporelle et fréquentielle liées et antagonistes. La différence principale avec la transformée de Fourier à court terme est que ces résolutions ne sont pas identiques en tous les points du plan temps-fréquence :

- dans le cas de changements brusques ou de structures très localisées, la transformée en ondelettes existera essentiellement dans le domaine des petites échelles, aptes à "voir" des détails fins du signal. Cependant, puisque ces petites échelles sont traduites par une ondelette analysante de support temporel réduit, il s'en suit que son support fréquentiel est étendu, ce qui limite la résolution fréquentielle absolue;
- réciproquement, une résolution fréquentielle importante n'est possible qu'avec une ondelette analysante longue, soit à de grandes échelles d'observation. Dans la zone du plan temps-fréquence où ce gain en résolution fréquentielle est possible, celui-ci se fait donc encore au détriment de la résolution temporelle.

L'analyse en ondelettes présente cependant l'avantage de ne pas correspondre à une résolution fixée *a priori* par le choix du signal élémentaire : c'est fondamentalement une analyse multi-échelles.

Interprétation par bancs de filtres. Si l'on revient à la définition de la transformée de Fourier à court terme, il est possible d'en donner une formulation fréquentielle équivalente selon :

$$F_X(t, v) = \int_{-\infty}^{+\infty} [X(\xi) e^{i2\pi\xi t}] H^*(\xi - v) d\xi$$

L'interprétation de cette écriture est que, dans une transformée de Fourier à court terme, le signal est analysé par passage à travers d'un banc de filtres uniforme, chaque filtre du banc (continu) se déduisant d'un filtre passe-bas unique par translation fréquentielle, ou hétérodynage. La bande passante absolue de chaque filtre est la même, d'où une résolution constante en tout point de l'axe des fréquences.

Dans le cas de la transformée en ondelettes, une interprétation en termes de banc de filtres est également possible. En effet, la formulation fréquentielle de la définition s'écrit :

$$T_X(t,a) = \int_{-\infty}^{+\infty} [X(\xi) e^{i2\pi\xi t}] \sqrt{a} H^*(a\xi) d\xi,$$

Ainsi, dans une transformée en ondelettes, le signal est analysé par passage à travers un banc de filtres non-uniforme, dont chaque filtre se déduit d'un filtre passebande unique par homothétie. Dans un tel filtrage, qui est dit à *Q*-constant, c'est maintenant la bande passante relative (rapport bande passante absolue/fréquence centrale) de chacun de ces filtres qui est une constante, d'où une résolution fréquentielle absolue d'autant plus réduite que l'échelle d'analyse est associée à une fréquence plus élevée.

### 6.2 Approches bilinéaires

Par opposition aux représentations (atomiques) linéaires qui décomposent un signal en constituants élémentaires, l'objet d'une distribution conjointe d'énergie est de répartir cette dernière sur les deux variables de description.

Si l'on considère comme précédemment les deux grandes classes de distributions relatives l'une au temps-fréquence ( $\lambda = v$ ) et l'autre au temps-échelle ( $\lambda = a$ ), la question est alors de trouver une distribution  $\rho_x(t, \lambda)$  telle que l'on ait :

$$E_x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_x(t, \lambda) \, \mathrm{d}\mu(t, \lambda) .$$

L'énergie étant par nature une grandeur quadratique, il est naturel, même si cela n'est pas nécessaire, de chercher à cet effet des distributions également quadratiques (bilinéaires ou sesquilinéaires), généralisant en cela la situation bien connue des densités d'énergie temporelle et fréquentielle :

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(v)|^2 dv.$$

Il s'avère que des sous-produits quadratiques de décompositions linéaires, comme le spectrogramme ou le scalogramme, peuvent jouer un tel rôle de distributions d'énergie puisque, pour des normalisations adéquates, on a :

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}|F_x(t,\,v)|^2\;\mathrm{d}t\;\mathrm{d}v\;=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}|T_x(t,\,a)|^2\,\frac{\mathrm{d}t\;\mathrm{d}a}{a^2}\;=E_x\;.$$

Cependant, le module carré d'une transformée linéaire n'est qu'un cas très particulier de transformée bilinéaire. Il importe donc de trouver des solutions plus générales au problème posé. On conviendra d'écrire celles-ci sous la forme générique

$$\rho_{x}(t, \lambda) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} K(u, u'; t, \lambda) x(u) x^{*}(u') du du',$$

forme paramétrée par un noyau qui, pour satisfaire à la contrainte de distribution d'énergie, devra être tel que

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} K(u, u'; t, \lambda) d\mu(t, \lambda) = \delta(u - u').$$

Classe de Cohen et classe affine. Le cadre général étant fixé, le choix effectif du noyau de paramétrisation nécessite d'imposer des contraintes supplémentaires devant être satisfaites par la distribution cherchée. Une façon de procéder est d'imposer des contraintes "naturelles" pour la distribution  $\rho_x(t,\lambda)$  et de les traduire en propriétés d'admissibilité pour son noyau  $K(u,u';t,\lambda)$ . Parmi ces contraintes figurent en tout premier lieu celles relatives à des principes de covariance, tels que l'effet d'une transformation puisse indifféremment s'obtenir sur la représentation ou sur le signal dont elle est issue.

Imposer la covariance par rapport à l'opérateur de translations en temps et en fréquence conduit à une classe générale de distributions temps-fréquence, appelée *classe de Cohen*, dont l'expression générale peut s'écrire

$$C_{x}(t, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(s - t, \xi - v) W_{x}(s, \xi) ds d\xi,$$

où  $\Pi(t,v)$  est une fonction temps-fréquence arbitraire agissant de façon corrélatoire sur la distribution "centrale"

$$W_x(t, v) = \int_{-\infty}^{+\infty} x \left(t + \frac{\tau}{2}\right) x^* \left(t - \frac{\tau}{2}\right) e^{-i2\pi v \tau} d\tau,$$

appelée distribution de Wigner-Ville. On peut noter que la définition de la classe de Cohen comme corrélation de la distribution de Wigner-Ville avec une fonction de paramétrisation arbitraire se transpose aisément en une relation de produit dans le domaine conjugué de Fourier. On a ainsi :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C_x(t, v) e^{i2\pi(\xi t + v\tau)} dt dv = f(\xi, \tau) A_x(\xi, \tau) ,$$

en notant:

$$A_x(\xi, \tau) = \int_{-\infty}^{+\infty} x \left( s + \frac{\tau}{2} \right) x * \left( s - \frac{\tau}{2} \right) e^{i2\pi \xi s} ds.$$

On reconnaît dans cette dernière quantité, qui n'est autre que la transformée de Fourier de la distribution de Wigner-Ville, la *fonction d'ambiguïté* (symétrisée) du signal, qui généralise au temps *et* à la fréquence la notion de fonction de corrélation.

D'une façon toute parallèle à la démarche suivie pour construire la classe de Cohen, imposer la covariance par rapport à l'opérateur de translation en temps et de dilatation conduit à une classe générale de distributions temps-échelle, appelée *classe affine*, dont l'expression générale peut s'écrire

$$\Omega_{\mathbf{x}}(t,a) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(\frac{s-t}{a}, a\xi) W_{\mathbf{x}}(s, \xi) \, \mathrm{d}s \, \mathrm{d}\xi ,$$

ou encore

$$\Omega_{x}(t, a) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f\left(a\xi, \frac{\tau}{a}\right) A_{x}(\xi, \tau) e^{-i2\pi\xi t} d\xi d\tau.$$

*Interprétations*. Les formulations unifiées qu'offrent les classes bilinéaires (Cohen et affine) présentent trois intérêts principaux:

- par la spécification de la fonction de paramétrisation arbitraire, il est possible de retrouver la plupart des définitions de distributions énergétiques connues ;
- il est facile de traduire une contrainte que l'on souhaite voir satisfaite par une représentation conjointe en une propriété d'admissibilité correspondante pour la fonction de paramétrisation ;
- il est enfin possible, en utilisant de tels arguments d'admissibilité, soit de vérifier *a priori* les propriétés d'une définition particulière, soit de construire une classe de solutions en fonction d'un cahier des charges spécifique.

Les tableaux qui suivent donnent quelques exemples de définitions de distributions temps-fréquence qui peuvent être retrouvées comme cas particuliers de la classe de Cohen, moyennant une spécification correcte de la fonction de paramétrisation arbitraire. Les exemples donnés n'épuisent pas les définitions envisageables ou même existantes. Ils sont cependant représentatifs de l'essentiel des différentes classes de solutions, les autres définitions s'en déduisant par des variations plus ou moins mineures.

$$f(\xi,\tau) = C_x(t,v;f)$$
 nom
$$\frac{f(\xi,\tau)}{\int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)e^{-i2\pi\nu\tau} d\tau$$
 Wigner-Ville
$$e^{i2\pi s\xi\tau} = \int_{-\infty}^{+\infty} x\left(t-\left(s-\frac{1}{2}\right)\tau\right)x^*\left(t-\left(s+\frac{1}{2}\right)\tau\right)e^{-i2\pi\nu\tau} d\tau$$
 s-Wigner
$$e^{i\pi\xi\tau} = x(t)X^*(\nu)e^{-i2\pi\nu t}$$
 Rihaczek
$$e^{\pm i\pi\xi\tau} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} x(s) \ U(\pm(t-s)) \ e^{-i2\pi\nu s} \ ds$$
 Page-Levin
$$\frac{\sin\pi\xi\tau}{\pi\xi\tau} = \int_{-\infty}^{+\infty} \left[\frac{1}{|\tau|}\int_{t-|\tau|/2}^{t+|\tau|/2} x\left(s+\frac{\tau}{2}\right)x^*\left(s-\frac{\tau}{2}\right) \ ds \right] e^{-i2\pi\nu\tau} \ d\tau$$
 Born-Jordan
$$e^{-(\pi\xi\tau/\sigma)^2/2} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\sigma}{|\tau|} e^{-2\sigma^2(s-t)^2/\tau^2}x\left(s+\frac{\tau}{2}\right)x^*\left(s-\frac{\tau}{2}\right) e^{-i2\pi\nu\tau} \ ds \ d\tau$$
 Choī-Williams
$$A_h^*(\xi,\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(\tau) \ g(s-t) \ x\left(s+\frac{\tau}{2}\right)x^*\left(s-\frac{\tau}{2}\right) e^{-i2\pi\nu\tau} \ ds \ d\tau$$
 séparable

Tableau 2-1
Classe de Cohen: quelques exemples de fonctions de paramétrisation et les représentations temps-fréquence associées

$$f(\xi, \tau)$$
  $\Omega_{\chi}(t, a; f)$  nom

$$e^{-i2\pi\nu_0\tau}$$
 
$$\int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi(\nu_0/a)\tau} d\tau$$
 Wigner-Ville

$$A_h^*(\xi, \tau)$$
  $\frac{1}{a} \left| \int_{-\infty}^{+\infty} x(s) h^*\left(\frac{s-t}{a}\right) ds \right|^2$  scalogramme

$$G(\xi)h(\tau) = \frac{1}{a} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h\left(\frac{\tau}{a}\right) g\left(\frac{s-t}{a}\right) x \left(s+\frac{\tau}{2}\right) x^* \left(s-\frac{\tau}{2}\right) \mathrm{d}s \,\mathrm{d}\tau \text{ séparable}$$

$$\varphi(\xi\tau) e^{-\mathrm{i}2\pi v_0 \tau}$$

Cohen "à noyau produit"

$$C_x\left(t, \frac{v_0}{a}, \varphi\right)$$

$$G(\xi) e^{-i2\pi H(\xi)\tau}$$

Wigner affine

$$\int_{-\infty}^{+\infty} G(a\xi) X\left(\frac{H(a\xi)}{a} - \frac{\xi}{2}\right) X^*\left(\frac{H(a\xi)}{a} + \frac{\xi}{2}\right) e^{-i2\pi i \xi} d\xi$$

#### Tableau 2-2

Classe affine : quelques exemples de fonctions de paramétrisation et les représentations temps-échelle associées

иош

 $\tilde{C}_{x}(t, v; f)$ 

$$v_0(\xi/2v_0) \coth(\xi/2v_0)$$

$$\frac{(\xi/2v_0)}{\sinh(\xi/2v_0)}$$

Bertrand

$$v \int_{-\infty}^{+\infty} \frac{(\gamma/2)}{\sinh(\gamma/2)} X \left( v \frac{(\gamma/2) e^{-(\gamma/2)}}{\sinh(\gamma/2)} \right) X^* \left( v \frac{(\gamma/2) e^{+(\gamma/2)}}{\sinh(\gamma/2)} \right) e^{-i2\pi v \gamma t} d\gamma$$

$$v_0\Big(1+(\xi/2v_0)^2\Big)^{1/2}$$

Unterberger (forme "active")

$$v \int_{0}^{+\infty} \left(1 + \frac{1}{\gamma^{2}}\right) X(\nu \gamma) X^{*}\left(\frac{\nu}{\gamma}\right) e^{i2\pi\nu(\gamma - 1/\gamma)t} d\gamma$$

$$v_0 (1 + (\xi/2v_0)^2)^{1/2}$$

$$(1 + (\xi/2v_0)^2)^{-1/2}$$
 Unterberger (forme "passive")

$$v \int_{0}^{+\infty} \frac{2}{\gamma} X(\nu \gamma) X^{*} \left(\frac{\nu}{\gamma}\right) e^{i2\pi \nu (\gamma - 1/\gamma)t} d\gamma$$

$$v_0 [1 + (\xi/4v_0)^2]$$

$$1 - (\xi/4v_0)^2$$

distribution D

$$v \int_{-\infty}^{+\infty} \left[ 1 - \left( \frac{\gamma}{4} \right)^2 \right] X \left( v \left( 1 - \frac{\gamma}{4} \right)^2 \right) X^* \left( v \left( 1 + \frac{\gamma}{4} \right)^2 \right) e^{-i2\pi v \gamma t} d\gamma$$

Tableau 2-3

Distributions de Wigner affines (cf. Tableau 2.2) : quelques exemples de fonctions de paramétrisation et les représentations temps-fréquence associées

Contrainte	Admissibilité
Energie	f(0, 0) = 1
Marginale en temps Marginale en fréquence	$f(\xi, 0) = 1$ $f(0, \tau) = 1$
Réalité	$f(\xi, \tau) = f^*(-\xi, -\tau)$
Positivité	$f(\xi, \tau) = f^*(-\xi, -\tau)$ $f(\xi, \tau) = \sum_{k=1}^{+\infty} c_k A_{h_k}^*(\xi, \tau), c_k > 0$
Causalité	$f(\xi, \tau) = \frac{i}{\pi} \operatorname{vp} \int_{-\infty}^{+\infty} \frac{f(\zeta, \tau)}{\xi - \zeta} e^{i\pi \zeta  \tau } d\zeta$
Invertibilité	$f(\xi, \tau) \neq 0$
Dilatations	$f(\xi, \tau) = f\left(\frac{\xi}{k}, k\tau\right)$
Filtrages Modulations	$f(\xi, \tau) f(\xi, \tau') = f(\xi, \tau + \tau')$ $f(\xi, \tau) f(\xi', \tau) = f(\xi + \xi', \tau)$
Support (large) en temps	$ s  > \frac{ \tau }{2} \implies \int_{-\infty}^{+\infty} f(\xi, \tau) e^{i2\pi\xi s} d\xi = 0$
Support (large) en fréquence	$ v  > \frac{ \xi }{2} \implies \int_{-\infty}^{+\infty} f(-\xi, \tau) e^{-i2\pi v \tau} d\tau = 0$
Unitarité	$ f(\xi,t) =1$
Fréquence instantanée	$f(\xi, 0) = 1$ , $\frac{\partial f}{\partial \tau}(\xi, 0) = 0$
Retard de groupe	$f(0,\tau) = 1$ , $\frac{\partial f}{\partial \xi}(0,\tau) = 0$

Tableau 2-4 Classe de Cohen: quelques exemples de contraintes sur  $C_{\chi}(t, v)$  et les conditions d'admissibilité associées sur la fonction de paramétrisation  $f(\xi, \tau)$ .

				Distributions		
Contraintes	S	WV	R	P	BJ	CW
Energie	V	4	<b>V</b>	√	1	√
Marginale en temps		√	1	$\checkmark$	4	4
Marginale en fréquence		4	1	4	√	4
Réalité	$\sqrt{}$	<b>√</b>		√	<b>√</b>	4
Positivité	1					
Causalité				√		
Invertibilité		√	1	√		4
Dilatations		4	4	4	4	√
Filtrages		<b>√</b>	4			
Modulations	4	√	4			
Support (large) en temps		√	<b>√</b>	<b>V</b>	<b>√</b>	
Support (large) en fréquence		4	4	√	<b>√</b>	
Unitarité		√	<b>√</b>			
Fréquence instantanée		4	<b>√</b>	√	<b>√</b>	4
Retard de groupe		4	4		√	4

Tableau 2-5

Classe de Cohen: quelques exemples de contraintes et de représentations les vérifiant ou non. (S : spectrogramme ; WV : Wigner-Ville ; R : Rihaczek ; P : Page ; BJ : Born-Jordan ; CW : Choï-Williams)

Contrainte	Admissibilité
Encrgie	$v_0 G(0) =  H(0) $
Marginale en temps	$v_0 G(\xi) = H(\xi) - \xi \frac{\mathrm{d}H}{\mathrm{d}\xi}(\xi)$
Marginale en fréquence	$G(0) = 1$ ; $H(0) = v_0$
Réalité	$G(\xi) = G(-\xi)$ ; $H(\xi) = H(-\xi)$
Localisation en temps	$v_0^2 G^2(\xi) = H^2(\xi) - \left(\frac{\xi}{2}\right)^2$
Unitarité	$v_0 G^2(\xi) = H(\xi) - \xi \frac{\mathrm{d}H}{\mathrm{d}\xi}(\xi)$
Retard de groupe	
Limite à bande étroîte	$G(0) = 1$ ; $H(0) = v_0$ ; $\frac{dG}{d\xi}(0) = \frac{dH}{d\xi}(0)$

Tableau 2-6 Distributions de Wigner affines : quelques exemples de contraintes et les conditions d'admissibilité associées sur les fonctions de paramétrisation  $H(\xi)$  et  $G(\xi)$ .

		Distributions			
Contraintes	В	Ua	Up	D	
Energie	√	√	<b>√</b>	<b>-</b> √	
Marginale en temps			4	4	
Marginale en fréquence	$\checkmark$	4	<b>√</b>	4	
Réalité	√	<b>√</b>	√	<b>√</b>	
Localisation en temps	<b>√</b>	4		<b>√</b>	
Unitarité	4				
Retard de groupe	<b>√</b>	<b>√</b>	√	4	
Limite à bande étroite	4	4	<b>√</b>	4	

#### Tableau 2-7

Distributions de Wigner affines : quelques exemples de contraintes et de représentations les vérifiant ou non. (B : Bertrand ; Ua : Unterberger, forme "active" ; Up : Unterberger, forme "passive" ; D : distribution-D)

.

### Pour en savoir plus

On trouvera un panorama récent et assez riche des applications du Traitement du Signal et des Images dans la référence suivante :

Le Courrier du CNRS, Numéro Spécial «Signaux et Images», No. 77, Juin 1991.

Les ouvrages généraux relatifs à la discipline abondent. Parmi les plus "généralistes", on peut citer tout particulièrement

- F. de Coulon, Théorie et Traitement des Signaux, Dunod, 1988.
- B. Picinbono, Théorie des Signaux et des Systèmes, Dunod, 1989.
- P. Duvaut, Traitement du Signal, Hermès, 1991.
- J.-P. Delmas, Eléments de Théorie du Signal : les Signaux Déterministes, Ellipses, 1991.
- M. Charbit, Eléments de Théorie du Signal : les Signaux Aléatoires, Ellipses, 1991.
- A. Papoulis, Signal Analysis, McGraw-Hill, 1984.

ainsi que les actes d'une Ecole d'Eté des Houches (Session XLV, 1985)

J.-L. Lacoume, T.S. Durrani and R. Stora (eds.), <u>Traitement du Signal/Signal Processing</u>, Elsevier, 1987,

ces derniers couvrant à la fois des questions de base et quelques thèmes avancés.

Parmi les ouvrages qui s'attachent plus particulièrement à l'aspect *numérique* des traitements, on retiendra entre autres

- M. Kunt, Traitement Numérique des Signaux, Dunod, 1984.
- M. Bellanger, <u>Traitement Numérique du Signal</u>, Masson, 1987.
- J.G. Proakis and D.G. Manolakis, <u>Introduction to Digital Signal Processing</u>, MacMillan, 1989.
- A.V. Oppenheim and A.S. Willsky, Signals and Systems, Prentice-Hall, 1983.
- L.B. Jackson, Digital Filters and Signal Processing, Kluwer, 1989.

La transformation de Fourier est détaillée (en particulier dans une perspective «signal») dans

R.N. Bracewell, The Fourier Transform and its Applications, McGraw-Hill, 1986.

G. Gasquet, P. Witomski, Analyse de Fourier et Applications, Masson, 1990.

On trouvera un exposé d'introduction aux processus aléatoires dans

A. Papoulis, <u>Probability, Random Variables and Stochastic Processes</u>, McGraw-Hill, 1984.

et une présentation plus approfondie dans

A. Blanc-Lapierre et B. Picinbono, Fonctions Aléatoires, Masson, 1981.

E. Wong and B. Hajek, <u>Stochastic Processes in Engineering Systems</u>, Springer-Verlag, 1985.

A.M. Yaglom, <u>Correlation Theory of Stationary and Related Random Functions</u>, Springer-Verlag, 1987.

Les notions de modélisation sont discutées par exemple dans

M. Najim, Modélisation et Identification en Traitement du Signal, Masson, 1988.

et leur application plus spécifique aux problèmes d'estimation spectrale paramétrique dans

- S.M. Kay, Modern Spectral Estimation, Prentice-Hall, 1988.
- S.L. Marple Jr, <u>Digital Spectral Analysis</u>, Prentice-Hall, 1987.

Une très bonne introduction aux questions de détection, d'estimation et de filtrage optimal se trouve dans

A.D. Whalen, <u>Detection of Signals in Noise</u>, Academic Press, 1971...

qui peut utilement servir de guide de lecture pour des ouvrages plus avancés comme

- H.L. Van Trees, <u>Detection</u>, <u>Estimation and Modulation Theory</u>, .
- H.V. Poor, An Introduction to Signal Detection and Estimation, Springer-Verlag, 1988.
- L.L. Scharf, Statistical Signal Processing, Addison-Wesley, 1991.

Parce qu'elles sont encore assez récentes, les méthodes temps-fréquence ont pour l'instant donné naissance à peu d'ouvrages de synthèse. On pourra cependant retenir

P. Flandrin, Temps-Fréquence, Hermès, 1993.

les ouvrages collectifs

- G. Longo and B. Picinbono (*eds.*), <u>Time and Frequency Representation of Signals and Systems</u>, CISM Courses and Lectures No. 309, Springer-Verlag, 1989.
- B. Boashash (ed.), Time-Frequency Signal Analysis, Longman-Cheshire, 1992.
- W. Mecklenbräuker (ed.), The Wigner Distribution and its Applications, Elsevier, 1993.

ainsi que les articles "tutorial"

- L. Cohen, "Time-Frequency Distribution A Review", *Proc. IEEE*, Vol. 77, No. 7, pp. 941-981, 1989.
- G.F. Boudreaux-Bartels and F. Hlawatsch, "Linear and Quadratic Time-Frequency Signal Representations", *IEEE Sig. Proc. Magazine*, 1992.

En ce qui concerne la théorie des ondelettes, on trouvera une introduction, doublée d'une perspective historique, dans

Y. Meyer, Ondelettes, Algorithmes et Applications, Armand Colin, 1992.

ainsi que dans l'article "tutorial"

O. Rioul and M. Vetterli, "Wavelets and Signal Processing", *IEEE Sig. Proc. Magazine*, Vol. 8, No. 4, pp. 14-38, 1991.

Pour une présentation plus approfondie, on pourra se reporter à

- Y. Meyer, Ondelettes et Opérateurs I. Ondelettes, Hermann, 1990.
- I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.

ainsi qu'à des ouvrages collectifs comme

- J.-M. Combes, A. Grossmann and Ph. Tchamitchian (eds.), Wavelets, Springer-Verlag, 1989.
- P.-G. Lemarić (ed.), <u>Les Ondelettes en 1989</u>, Lecture Notes in Math. No. 1438, Springer-Verlag, 1990.