

---

# Traitemen Numérique du Signal

---

ELE102

---

Chaouki Diab

---

ISSAE -  
Cnam Liban

---

# Traitemen Numérique du Signal

---

<b>Liste des Figures</b>	<b>5</b>
<b>Introduction</b>	<b>8</b>
Définitions et problématique	8
Organisation du cours : Aperçu global	9
<b>Chapitre 1 – Numérisation des Signaux</b>	<b>13</b>
<b>1.1 Analyse de Fourier</b>	<b>13</b>
Développement en série de Fourier d'une fonction périodique	13
Transformation de Fourier d'une fonction	15
Relations entre coefficients de Fourier et spectre de Fourier	16
Conservation de l'énergie du signal dans le spectre de Fourier	16
Quelques propriétés de la transformée de Fourier (TF)	16
Impulsion de Dirac	17
Relation entre Impulsion de Dirac et fonction de Heavyside	17
<b>1.2 Signaux Usuels</b>	<b>18</b>
<b>1.3 Étapes de la Numérisation</b>	<b>19</b>
Échantillonnage des signaux:	21
Limitations pratiques	22
Quantification	23
Paramètres d'un quantificateur:	24
Types de quantificateurs:	24
<b>Chapitre 2 - La Transformée de Fourier Discrète</b>	<b>28</b>
<b>2.1 Échantillonnage du spectre:</b>	<b>28</b>
Différence entre un signal apériodique et un signal périodique	30
Cas des signaux réels	30
<b>2.2 Transformée de Fourier Rapide (TFR)</b>	<b>31</b>
<b>Chapitre 3 - Généralisation aux Signaux Bidimensionnels – Cas d'images</b>	<b>37</b>
Échantillonnage Bidimensionnel:	37
Notion de fréquence spatiale:	39
Représentations fréquentielles	39
La Transformation de Fourier Discrète Bidimensionnelle:	41
La Transformation de Cosinus Discrète 1D:	43
Transformée Directe :	43
Transformée Inverse:	43
Propriétés de la TCD:	43
Relation entre TCD et TFD	44
Représentation d'image par TC2D:	45
La TCD : Meilleure transformation sous optimale	45

<b>Chapitre 4 – Signaux et Systèmes à temps discret</b>	<b>47</b>
<b>4.1 Systèmes à temps discret :</b>	<b>47</b>
Exemple 1: Système à retard idéal :	47
Exemple 2: Système Moyenneur à fenêtre glissante:	48
Système sans mémoire	48
Système linéaire	48
Système invariant dans le temps :	48
Système Causal:	49
Système Stable:	49
<b>4.2 Systèmes récursifs – Systèmes non récursifs</b>	<b>49</b>
Réponse impulsionnelle	51
Réponse forcée à une entrée quelconque – Convolution numérique	51
<b>4.3 Transformée en Z</b>	<b>52</b>
<b>4.3.1 Propriétés de la transformée en Z</b>	<b>54</b>
Domaine de Convergence	54
Linéarité	56
Dérivation	56
Inter-corrélation	56
Théorème de la valeur initiale	57
Théorème de la valeur finale	57
Table des transformées en Z usuelles:	57
Systèmes en cascade	57
<b>4.3.2 Réponse en fréquence d'un SLI numérique</b>	<b>58</b>
Réponse de régime à une exponentielle imaginaire numérique	58
Trois méthodes pour déterminer la réponse en fréquence d'un système SLI numérique	58
<b>4.3.3 Pôles et Zéros de <math>H(z)</math></b>	<b>59</b>
<b>4.3.4 L'inversion de la transformée en Z</b>	<b>61</b>
Méthode des Résidus:	61
Décomposition en fractions rationnelles	62
Division polynomiale suivant les puissances croissantes de $z^{-1}$	64
Par l'équation aux différences (équation récurrente pour une impulsion unité)	65
<b>Chapitre 5 – Les Filtres Numériques</b>	<b>67</b>
<b>5.1 Classification des filtres:</b>	<b>67</b>
<b>5.2 Filtres à Réponse Impulsionnelle de durée Finie (RIF):</b>	<b>69</b>
<b>5.2.1 Réalisations</b>	<b>69</b>
Différence entre les deux réalisations :	70
Cas des signaux de longueur finie	70
Cas des signaux de longueur très grande (voire infinie)	70
<b>5.2.2 Filtres à phase linéaire</b>	<b>71</b>
<b>5.2.3 Exemple des Filtres en peigne</b>	<b>73</b>

<b>5.3 Filtres à Réponse Impulsionnelle de durée Infinie (RII):</b>	<b>75</b>
Structure en cascade	77
Structure en parallèle	78
<b>5.4 Le principe de transposition</b>	<b>80</b>
<b>5.5 Le principe du filtrage adaptatif</b>	<b>80</b>
<b>5.6 Synthèse des Filtres Numériques</b>	<b>81</b>
Synthèse des Filtres RIF	81
Cas où l'intégrale peut être évaluée analytiquement:	81
Cas où l'intégrale ne peut pas être évaluée analytiquement:	83
Conception à ondulations constantes (equiripple characteristic)	84
Synthèse des filtres RII:	84
Par équivalence de la dérivation :	85
Par équivalence de l'intégration	86
Par invariance de la réponse impulsionnelle	87
<b>5.7 Comparaison entre les filtres FIR et IIR</b>	<b>88</b>
<b>Chapitre 6 - Synthèse des filtres analogiques</b>	<b>90</b>
<b>6.1 Rappel sur les filtres analogiques</b>	<b>90</b>
<b>6.2 Synthèse des filtres analogiques</b>	<b>92</b>
La normalisation du gabarit	92
L'approximation de la fonction de transfert normalisée	93
Approximation de Butterworth	95
Approximation de Chebychev	98
Approximation de Cauer (ou elliptique)	105
La dé-normalisation	107
<b>Chapitre 7 - Décomposition en sous-bandes et Transformée en Ondelettes</b>	<b>108</b>
<b>7.1 Principe de la décomposition en sous-bandes</b>	<b>109</b>
<b>7.2 Solution d'Esteban - Galand: les QMF</b>	<b>111</b>
<b>7.3 Solution de Smith-Barnwell: les CQF</b>	<b>112</b>
<b>7.4 Relation avec la représentation temps-fréquence</b>	<b>113</b>
<b>7.5 Représentation par Transformée en Ondelettes</b>	<b>114</b>
Lien avec les sous-bandes:	116
<b>7.6 Cas d'images</b>	<b>117</b>
Application de la DWT	120
Décodage progressif	121
Problème de bords:	122
<b>Chapitre 8 - Applications au Traitement d'images</b>	<b>124</b>
<b>8.1 Introduction</b>	<b>124</b>
<b>8.2 Exemples de traitement d'images</b>	<b>124</b>

Zoom: Réduction et Agrandissement	124
Filtrage Linéaire	124
Réduction du bruit ou Lissage spatial:	125
Le Filtre moyenneur	127
Filtres non-linéaires:	127
Le filtre sigma:	127
Le V-filtre:	128
Le filtre de Nagao:	128
Le filtre médian:	128
<b>8.3 Les Filtres d'Ordre</b>	<b>128</b>
Notions sur la statistique d'ordre:	129
Définition d'un filtre d'ordre:	129
Filtres d'ordre usuels:	129
Filtre médian:	129
Filtre moyenneur:	130
<b>8.4 Détection des contours:</b>	<b>130</b>
Approche dérivative	131
Opérateurs linéaires (de convolution):	132
Opérateurs non-linéaires (Filtres d'ordre):	135

## Liste des Figures

Figure 1-1. Suite d'impulsions périodique et de période $T$	15
Figure 1-2 Impulsion isolée $i(t)$ de largeur $\tau$	16
Figure 1-3 Spectre $I(f)$ de l'impulsion isolée	16
Figure 1-4 : Représentations graphiques des fonctions $I(t)$ , $g(t)$ et $i(t)$ .	18
Figure 1-5 Types de signaux selon la continuité de la variable et de l'amplitude	20
Figure 1-6 : illustration de l'opération d'échantillonnage	21
Figure 1-7 : Effet de l'échantillonnage sur le spectre : a) Chevauchement spectral. b) Respect de la condition de Shannon	22
Figure 1-8. Illustration d'une quantification uniforme linéaire	25
Figure 1-9. Algorithme de calcul itératif d'un quantificateur Lloyd-Max.	27
Figure 2-1. Schéma-bloc de décomposition d'une TFD sur $N$ points en 2 TFD de $N/2$ points chacune.	33
Figure 2-2. Graphe de fluence d'une TFD sur 2 points	34
Figure 2-3. Graphe de fluence liant $F_k^N(x)$ , $F_{k+N/2}^N(x)$ , $F_k^{N/2}(x_p)$ et $F_k^{N/2}(x_i)$	34
Figure 2-4. Le Graphe de fluence de la Figure 2-3 réduit à une multiplication	34
Figure 2-5. Le Graphe de fluence de la FFT sur 8 points	35
Figure 3-1 : Formes possibles de cellules élémentaires	38
Figure 3-2- Images échantillonées avec différents pas	38
Figure 3-3- Supports du spectre fréquentiel d'une image selon différents pas d'échantillonnage	40
Figure 3-4- Image monochrome dont l'amplitude varie sinusoïdalement dans la direction horizontale	41
Figure 3-5- Images dont l'amplitude varie sinusoïdalement selon l'une des deux directions horizontale ou verticale	41
figure 3-6 - Implantation de la TF2D en utilisant la TFD	42
Figure 3-7. Schéma-bloc illustrant le calcul des TCD de 2 signaux réels en utilisant une seule FFT sur $N$ points.	44

<i>Figure 3-8. Fonctions de base de la TCD-8x8 :</i>	46
<i>Figure 3-9. TC2D-8x8 d'une image de 32x32 pixels</i>	46
<i>Figure 4-1. Représentation d'un système à temps discret</i>	47
<i>Figure 4-2. Exemple de structure d'un filtre numérique récursif</i>	50
<i>Figure 4-3. Structure du SLI correspondant au calcul des intérêts composés</i>	51
<i>Figure 4-4. Réponse impulsionale</i>	51
<i>Figure 4-5. Zone de stabilité en z et lien avec la zone de stabilité en p.</i>	54
<i>Figure 4-6. Couronne de convergence de la transformée en Z</i>	55
<i>Figure 4-7. Systèmes en cascade</i>	57
<i>Figure 4-8. Système SLI simple</i>	59
<i>Figure 4-9. Zéros (o) et pôles (x) d'une <math>H(z)</math> dans le plan complexe avec le cercle unité</i>	59
<i>Figure 5-1. Réponses fréquentielles des filtres idéaux</i>	67
<i>Figure 5-2. Gabarit d'un filtre</i>	68
<i>Figure 5-3. Réalisation transversale (<math>n_0 = 0</math> ici).</i>	69
<i>Figure 5-4. Réalisation par TFD</i>	70
<i>Figure 5-5. Réponse fréquentielle et Répartition des pôles et des zéros de <math>H(z)=1-z^{-1}</math>.</i>	74
<i>Figure 5-6. Réponse fréquentielle et Répartition des pôles et des zéros d'un filtre en peigne <math>G(z)=1-z^{-2}</math>.</i>	74
<i>Figure 5-7. Réponse fréquentielle et Répartition des pôles et des zéros d'un filtre en peigne <math>G(z)=1-z^{-10}</math>.</i>	74
<i>Figure 5-8. Réponse fréquentielle et Répartition des pôles et des zéros d'un filtre en peigne <math>G(z)=1-z^{-20}</math>.</i>	75
<i>Figure 5-9. Structure transversale d'un filtre en peigne de la forme <math>G(z)=1-z^{-N}</math>.</i>	75
<i>Figure 5-10. Réalisation récursive de forme directe 1.</i>	76
<i>Figure 5-11. Réalisation récursive obtenue en inter-changeant la partie récursive et non-récursive de forme directe 1.</i>	77
<i>Figure 5-12. Réalisation récursive et canonique de forme directe 2.</i>	78
<i>Figure 5-13. Mise en parallèle d'un système discret</i>	79
<i>Figure 5-14. Réalisation canonique d'un système en cascade formé d'un sous-système d'ordre 1 et d'un autre d'ordre 2.</i>	79
<i>Figure 5-15. Réalisation canonique d'un système en parallèle formé d'un sous-système d'ordre 1 et d'un autre d'ordre 2.</i>	79
<i>Figure 5-16. a) Filtre transversal simple b) Version transposée de ce filtre. c) Version transposée avec l'entrée à gauche.</i>	80
<i>Figure 5-17. Réponses temporelles des principales fenêtres</i>	82
<i>Figure 5-18. Réponses fréquentielles des principales fenêtres</i>	83
<i>Figure 5-19. Application de <math>z=1/(1-pT)</math> du plan des p dans le plan des z</i>	86
<i>Figure 6-1. Réponse en amplitude et en phase d'un filtre n'introduisant pas de distorsion</i>	90
<i>Figure 6-2. Gabarit fréquentiel d'un filtre passe-bas</i>	91
<i>Figure 6-3. Gabarit fréquentiel d'un filtre passe-bande</i>	92
<i>Figure 6-4. Gabarit prototype passe-bas (Normalisé)</i>	93
<i>Figure 6-5. Gabarit prototype passe-bas normalisé et modifié selon la fonction caractéristique <math>K(j\Omega)</math></i>	94
<i>Figure 6-6. Allure de <math> K(j\Omega) </math> et du gain <math>10\log_{10} H(j\Omega) ^2</math>.</i>	95
<i>Figure 6-7. Allure des fonctions caractéristiques du filtre passe-bas de Butterworth</i>	96
<i>Figure 6-8. Localisation des racines du filtre passe-bas de Butterworth pour différentes valeurs de n.</i>	96
<i>Figure 6-9. Gabarit d'un filtre passe-bas simple.</i>	97
<i>Figure 6-10. Répartition des pôles de l'approximation de Butterworth du passe-bas normalisé.</i>	98
<i>Figure 6-11. Réponse en fréquence de l'approximation de Butterworth du passe-bas normalisé.</i>	98
<i>Figure 6-12. Gain et délai de groupe de l'approximation de Butterworth d'un passe-bas normalisé.</i>	98
<i>Figure 6-13. Allure des fonctions caractéristiques du filtre passe-bas de Chebychev (type I)</i>	99
<i>Figure 6-14. Allure des carrés des polynômes de Chebychev (pour <math>n \leq 4</math>)</i>	99
<i>Figure 6-15. Répartition des pôles de l'approximation de Chebychev type 1 du passe-bas normalisé.</i>	101

<i>Figure 6-16. Gain et délai de groupe de l'approximation de Chebychev 1 d'un passe-bas normalisé.</i>	101
<i>Figure 6-17. Réponse en fréquence de l'approximation de Chebychev 1 du passe-bas normalisé.</i>	102
<i>Figure 6-18. Allure des fonctions caractéristiques du filtre passe-bas de Chebychev type II.</i>	102
<i>Figure 6-19. Répartition des pôles de l'approximation de Chebychev type 2 du passe-bas normalisé.</i>	103
<i>Figure 6-20. Réponse en fréquence de l'approximation de Chebychev 2 du passe-bas normalisé.</i>	104
<i>Figure 6-21. Gain et délai de groupe de l'approximation de Chebychev 2 d'un passe-bas normalisé.</i>	104
<i>Figure 6-22. Allure des fonctions caractéristiques elliptiques de Cauer d'un filtre passe-bas.</i>	105
<i>Figure 6-23. Répartition des pôles de l'approximation elliptique du passe-bas normalisé.</i>	106
<i>Figure 6-24. Réponse en fréquence de l'approximation elliptique du passe-bas normalisé.</i>	106
<i>Figure 6-25. Gain et délai de groupe de l'approximation elliptique d'un passe-bas normalisé.</i>	107
<i>Figure 7-1. Système de codage/décodage en sous-bandes</i>	108
<i>Figure 7-2. Système de décomposition/reconstruction à 2 bandes</i>	109
<i>Figure 7-3. Représentation schématique des QMF en fréquence</i>	112
<i>Figure 7-4. Exemples de formes d'ondelettes et de fonctions d'échelle correspondantes: a) orthogonales b) bi-orthogonales</i>	114
<i>Figure 7-5. Partition du plan fréquentiel pour une décomposition 2D en 4 sous-bandes</i>	117
<i>Figure 7-6. Schéma-bloc de décomposition 2D en 4 sous-bandes</i>	118
<i>Figure 7-7. Reconstruction 2D à partir de 4 sous-bandes</i>	119
<i>Figure 7-8. Décomposition 2D pyramidale à la résolution <math>2^2</math> (7 sous-bandes)</i>	119
<i>Figure 7-9. Exemples d'images décomposées pyramidalement jusqu'à la résolution <math>2^3</math></i>	120
<i>Figure 7-10 Formes d'ondelettes biorthogonales de Daubechies</i>	121
<i>Figure 7-11. Modes de Décodage progressif</i>	122
<i>Figure 7-12. Illustration des erreurs de reconstruction sur les bords</i>	123
<i>Figure 7-13. Exemples d'extensions proposées et erreurs de reconstruction associées</i>	123
<i>Figure 8-1. Illustration des effets des zooms numériques</i>	125
<i>Figure 8-2. Images bruitées et filtrées</i>	126
<i>Figure 8-3. Masques du Filtre DE NAGAO</i>	128
<i>Figure 8-4. Schéma-bloc d'un filtre d'ordre</i>	129
<i>Figure 8-5. modèles théoriques de contour</i>	130
<i>Figure 8-6. Illustration des méthodes dérivatives sur un signal monodimensionnel <math>I(x)</math></i>	131
<i>Figure 8-7. Exemples de détection de contours</i>	133

# Introduction

## Définitions et problématique

Le signal est le support physique de l'information émise par une source et destinée à un récepteur; c'est le véhicule de l'intelligence dans les systèmes.

Il transporte les ordres dans les équipements de contrôle et de télécommande, il achemine sur les réseaux l'information, la parole ou l'image. Il est particulièrement fragile et doit être manipulé avec beaucoup de soins.

Les signaux intervenant dans les échanges d'informations sont de nature complexe et peuvent être masqués par des perturbations indésirables (bruits, distorsions, ... etc.).

On appelle bruit tout phénomène perturbateur (interférence, bruit de fond, etc...) gênant la perception ou l'interprétation d'un signal, ceci par analogie avec les nuisances acoustiques du même nom.

Le rapport signal sur bruit (RSB ou SNR: Signal-to-Noise Ratio en anglais) est une mesure du degré de contamination du signal par du bruit. Il s'exprime sous la forme du rapport des puissances respectives du signal Ps et du bruit Pn:  $RSB = Ps/Pn$ .

Il est souvent exprimé selon une échelle logarithmique mesurée en *décibels* :

$$\text{Éq 0-1} \quad RSB (\text{dB}) = 10 \log_{10} \left( \frac{Ps}{Pn} \right)$$

Mathématiquement, les signaux sont représentés par une fonction à une ou plusieurs variables. Une grande majorité des signaux ont comme variable commune le **temps**. Toutefois, le temps n'est pas la seule variable dont un signal peut dépendre; les variations de la pression en fonction de l'altitude ou les variations de la température en fonction du lieu (ou de la position) sont également des exemples de signaux.

Il existe aussi des fonctions de plusieurs variables comme par exemple une **image** photographique fixe qui est caractérisée par une intensité lumineuse dépendante de deux variables représentant les **coordonnées** dans le plan de l'image.

Une image 3D (Tridimensionnelle) est un signal à trois variables qui sont les coordonnées de l'espace.

Une séquence d'images animées (vidéo) est une représentation d'un signal à trois variables dont le troisième est le temps et les deux premières sont celles d'une image fixe puisque une séquence d'images animées est une succession de plusieurs images fixes.

Le traitement que subit un signal a pour but d'extraire des informations, de modifier le message qu'il transporte ou de l'adapter aux moyens de transmission; c'est là qu'interviennent les techniques numériques.

En effet, si l'on imagine de substituer au signal un ensemble de nombres qui représentent sa grandeur ou amplitude à des instants convenablement choisis, le traitement, même dans sa forme la plus élaborée, se ramène à une séquence d'opérations logiques et arithmétiques sur cet ensemble

de nombres, associées à des mises en mémoire.

La conversion du signal continu analogique en un signal numérique est réalisée par des capteurs qui opèrent sur des enregistrements ou directement dans les équipements qui produisent ou reçoivent le signal. Les opérations qui suivent cette conversion sont réalisées par des calculateurs numériques agencés ou programmés pour effectuer l'enchaînement des opérations définissant le traitement.

Le traitement numérique du signal désigne l'ensemble des opérations, calculs arithmétiques et manipulations de nombres, qui sont effectués sur un signal à traiter, représenté par une suite ou un ensemble de nombres, en vue de fournir une autre suite ou un autre ensemble de nombres, qui représentent le signal traité.

Les fonctions les plus variées sont réalisables de cette manière, comme l'analyse spectrale, le filtrage linéaire ou non linéaire, le transcodage, la modulation, la détection, l'estimation et l'extraction de paramètres. Les machines utilisées sont des calculateurs numériques.

Les systèmes correspondant à ce traitement obéissent aux lois des systèmes discrets. Les nombres sur lesquels il porte peuvent dans certains cas être issus d'un processus discret. Cependant, ils représentent souvent l'amplitude des échantillons d'un signal continu et dans ce cas, le calculateur prend place derrière un dispositif convertisseur analogique-numérique et éventuellement devant un convertisseur numérique-analogique.

Dans la conception de tels systèmes et l'étude de leur fonctionnement, la numérisation du signal revêt une importance fondamentale et les opérations d'échantillonnage et de codage doivent être analysées dans leur principe et leurs conséquences. La théorie des distributions constitue une approche concise, simple et efficace pour cette analyse.

## **Organisation du cours : Aperçu global**

Après un certain nombre de rappels sur l'analyse de Fourier, les distributions et la représentation des signaux, le chapitre 1 rassemble les résultats les plus importants et les plus utiles sur l'échantillonnage et le codage d'un signal.

L'essor du traitement numérique date de la découverte d'algorithmes de calcul rapide de la Transformée de Fourier Discrète. En effet, cette transformation est à la base de l'étude des systèmes discrets et elle constitue dans ce domaine numérique l'équivalent de la Transformation de Fourier dans le domaine analogique, c'est le moyen de passage de l'espace des temps discret à l'espace des fréquences discret. Elle s'introduit naturellement dans une analyse spectrale avec un pas de fréquence diviseur de la fréquence d'échantillonnage des signaux à analyser.

Les algorithmes de calcul rapide apportent des gains tels qu'ils permettent de faire les opérations en temps réel dans de nombreuses applications pourvu que certaines conditions élémentaires soient remplies. Ainsi, la Transformation de Fourier Discrète constitue non seulement un outil de base dans la détermination des caractéristiques du traitement et dans l'étude de ses incidences sur le signal, mais de plus, elle donne lieu à la réalisation d'équipements toutes les fois qu'une analyse de spectre intervient, par exemple, dans les systèmes comportant des bancs de filtres ou quand, par la puissance de ses algorithmes, elle conduit à une approche avantageuse pour un

circuit de filtrage. Le chapitre 2 lui est consacré; il donne d'une part une présentation des propriétés élémentaires et du mécanisme des algorithmes de calcul rapide et de leurs applications, et d'autre part, un ensemble de variantes associées aux situations pratiques. En tant que système, le calculateur de Transformée de Fourier Discrète est un système linéaire discret, invariant dans le temps.

Dans le chapitre 3, on introduit l'image numérique comme étant un signal à deux dimensions. On y trouve une généralisation de la numérisation des signaux, de la transformée de Fourier et de la TFD au cas d'images. On y introduit la TCD (Transformée Cosinus Discrète) qui est un outil très utilisé dans les applications d'images, en particulier celles liées à la compression.

Une grande partie de ce document est consacrée à l'étude des systèmes linéaires discrets invariants dans le temps à une dimension, qui sont facilement accessibles et très utiles. Les systèmes à plusieurs dimensions et en particulier à deux et trois dimensions connaissent un grand développement; ils sont appliqués par exemple aux images; cependant, leurs propriétés se déduisent en général de celles des systèmes à une dimension dont ils ne sont souvent que des extensions simplifiées.

La linéarité et l'invariance temporelle entraînent l'existence d'une relation de convolution qui régit le fonctionnement du système, ou filtre, ayant ces propriétés. Cette relation de convolution est définie à partir de la réponse du système au signal élémentaire que représente une impulsion, la réponse impulsionnelle, par une intégrale dans le cas des signaux analogiques.

Ainsi, si  $x(t)$  désigne le signal à filtrer,  $h(t)$  la réponse impulsionnelle du filtre, le signal filtré  $y(t)$  est donné par l'équation :

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau$$

Dans ces conditions, une telle relation qui pourtant traduit directement le fonctionnement réel du filtre, offre un intérêt pratique limité. En effet, d'une part il n'est pas très aisés de déterminer la réponse impulsionnelle à partir des critères qui définissent l'opération de filtrage envisagée et d'autre part une équation comportant une intégrale ne permet pas facilement de reconnaître et vérifier le comportement du filtre.

La conception est beaucoup plus facile à aborder dans le domaine des fréquences car la transformation de Laplace ou la transformation de Fourier permettent d'accéder à un plan transformé où les relations de convolution du plan amplitude-temps deviennent de simples produits de fonctions.

À la réponse impulsionnelle, la transformation de Fourier fait correspondre la réponse en fréquence du système, et le filtrage se ramène au produit de cette réponse en fréquence par la transformée de Fourier, ou spectre, du signal à filtrer.

Dans les systèmes numériques, qui sont du type discret, la convolution se traduit par une sommation. Le filtre est défini par une suite de nombres qui constitue sa réponse impulsionnelle. Ainsi, si la suite à filtrer s'écrit  $x(n)$ , la suite filtrée  $y(n)$  s'exprime par la sommation suivante, où  $n$  et  $m$  sont des entiers :

$$y(n) = \sum_m h(m)x(n-m)$$

Deux cas se présentent alors :

- soit la sommation porte sur un nombre fini de termes, c'est-à-dire que les  $h(m)$  sont nuls sauf pour un nombre fini de valeurs de la variable entière  $m$ . Le filtre est dit à réponse impulsionnelle finie; en faisant allusion à sa réalisation, on le désigne encore par non récursif car il ne nécessite pas de boucle de réaction de la sortie sur l'entrée dans sa mise en œuvre. Il est à mémoire finie, puisqu'il ne garde le souvenir d'un signal élémentaire, une impulsion par exemple, que pendant une durée limitée. Les nombres  $h(m)$  sont appelés les coefficients du filtre, qu'ils définissent complètement. Ils peuvent se calculer d'une manière directe très simple, par exemple en faisant le développement en série de Fourier de la réponse en fréquence à réaliser. Ce type de filtre présente des caractéristiques originales très intéressantes ; par exemple, la possibilité d'une réponse rigoureusement linéaire en phase, c'est-à-dire d'un temps de propagation de groupe constant ; les signaux dont les composantes se trouvent dans la bande passante du filtre ne sont pas déformés à la traversée de ce filtre. Cette possibilité est exploitée dans les systèmes de transmission de données ou en analyse spectrale par exemple.
- soit la sommation porte sur un nombre infini de termes, les  $h(m)$  ont une infinité de valeurs non nulles ; le filtre est dit à réponse impulsionnelle infinie ou encore de type récursif, car il faut réaliser sa mémoire par une boucle de réaction de la sortie sur l'entrée. Son fonctionnement est régi par une équation selon laquelle un élément de la suite de sortie  $y(n)$  est calculée par la sommation pondérée d'un certain nombre d'éléments de la suite d'entrée  $x(n)$  et d'un certain nombre d'éléments de la suite de sortie précédents. Par exemple, si  $L$  et  $K$  sont des entiers, le fonctionnement du filtre peut être défini par l'équation suivante :

$$y(n) = \sum_{l=0}^L a_l x(n-l) - \sum_{k=1}^K b_k y(n-k)$$

Les  $a_l$  ( $l = 0, 1, \dots, L$ ) et  $b_k$  ( $k = 1, 2, \dots, K$ ) sont les coefficients. Comme pour les filtres analogiques, l'étude de ce type de filtre ne se fait pas en général simplement de manière directe; il est nécessaire de passer par un plan transformé. La transformation de Laplace ou la transformation de Fourier pourraient être utilisées. Cependant, il existe une transformation beaucoup mieux adaptée, la transformation en  $Z$ , qui est l'équivalent pour les systèmes discrets. Un filtre est caractérisé par sa fonction de transfert en  $Z$ , désignée généralement par  $H(Z)$ , et qui fait intervenir les coefficients par l'équation suivante :

$$H(Z) = \frac{\sum_{l=0}^L a_l Z^{-l}}{1 + \sum_{k=1}^K b_k Z^{-k}}$$

Une bonne partie de ce cours est consacrée à l'étude des caractéristiques de ces filtres numériques.

Le chapitre 4 présente les propriétés des systèmes linéaires discrets invariants dans le temps, rappelle les propriétés principales de la transformation en Z et donne les éléments nécessaires à l'étude des filtres.

Le chapitre 5 traite les filtres numériques:

- On y distingue ceux à réponse impulsionnelle finie et ceux à réponse impulsionnelle infinie: leurs propriétés sont étudiées, les techniques de calcul des coefficients sont décrites ainsi que les structures de réalisation.
- Les filtres à réponse impulsionnelle infinie ayant des propriétés comparables à celles des filtres analogiques continus, il est naturel d'envisager pour leur réalisation des structures du même type que celles qui sont couramment employées en filtrage analogique. Ils sont généralement réalisés par une mise en cascade de cellules élémentaires du premier et second ordre, le chapitre 5 décrit ces cellules et leurs propriétés, ce qui d'une part facilite considérablement l'approche de ce type de système et d'autre part fournit un ensemble de résultats très utiles dans la pratique.

Le chapitre 6 est un rappel des méthodes de synthèse des filtres analogiques où l'on décrit les étapes à suivre pour approximer, à partir du gabarit, la fonction de transfert en utilisant les méthodes d'approximation les plus utilisées (Butterworth, Chebyshev I et II, elliptique).

Les bancs de filtres pour la décomposition et la reconstruction des signaux sont devenus un outil de base pour la compression. Leur fonctionnement est décrit au chapitre 7 avec les méthodes de calcul et les structures de réalisation. Les filtres peuvent être déterminés à partir de spécifications dans le temps.

Pour terminer, le chapitre 8 décrit brièvement quelques applications de traitement d'images, en montrant comment les méthodes et techniques de base sont exploitées.

# Chapitre 1 – Numérisation des Signaux

La conversion d'un signal analogique sous forme numérique implique une double approximation. D'une part, dans l'espace des temps, le signal fonction du temps  $s(t)$  est remplacé par ses valeurs  $s(nT)$  à des instants multiples entiers d'une durée  $T$ ; c'est l'opération d'échantillonnage.

D'autre part, dans l'espace des amplitudes, chaque valeur  $s(nT)$  est approchée par un multiple entier d'une quantité élémentaire  $\Delta$ ; c'est l'opération de quantification.

La valeur approchée ainsi obtenue est ensuite associée à un nombre binaire; c'est le codage, ce terme étant souvent utilisé pour désigner l'ensemble, c'est-à-dire le passage de la valeur  $s(nT)$  au code binaire qui la représente.

L'objet du présent chapitre est d'analyser l'incidence sur le signal de ces deux approximations. Pour mener à bien cette tâche, on utilise deux outils de base qui sont l'analyse de Fourier et la théorie des distributions.

## 1.1 Analyse de Fourier

L'analyse de Fourier est un moyen de décomposer un signal en une somme de signaux élémentaires particuliers, qui ont la propriété d'être faciles à mettre en œuvre et à observer. L'intérêt de cette décomposition réside dans le fait que la réponse au signal d'un système obéissant au principe de superposition peut être déduite de la réponse aux signaux élémentaires.

Ces signaux élémentaires sont périodiques et complexes. Afin de permettre une étude en amplitude et en phase des systèmes, on les exprime par la fonction  $s_e(t)$  telle que:

$$\text{Eq 1-1} \quad s_e(t) = e^{j2\pi ft} = \cos(2\pi ft) + j \sin(2\pi ft)$$

où  $f$  représente l'inverse de la période, c'est la fréquence du signal élémentaire.

Dans la mesure où les signaux élémentaires sont périodiques, il est clair que l'analyse se simplifie dans le cas où le signal est lui-même périodique. Ce cas va être examiné d'abord, bien qu'il ne corresponde pas aux signaux les plus intéressants, puisqu'un signal périodique est parfaitement déterminé et ne porte pratiquement pas d'information.

### Développement en série de Fourier d'une fonction périodique

Soit  $x(t)$ , une fonction de la variable  $t$  périodique et de période  $T$ , c'est-à-dire satisfaisant la relation:

$$\text{Eq 1-2} \quad x(t + T) = x(t)$$

Sous certaines conditions, on démontre que cette fonction est développable en série de Fourier, c'est-à-dire que l'égalité suivante est vérifiée :

$$\text{Eq 1-3} \quad x(t) = \sum_{n=-\infty}^{+\infty} X_n \cdot e^{j \frac{2\pi n t}{T}}$$

L'indice  $n$  est un entier et les  $X_n$  sont appelés les coefficients de Fourier.  $x(t)$  peut être représenté par une combinaison linéaire des fonctions exponentielles complexes orthogonales  $\varphi_n(t) = e^{j\frac{2\pi nt}{T}}$ ,  $\varphi_n(t)$  périodique de fréquence  $n/T$ . Les  $X_n$  sont définis par l'expression:

$$\text{Éq 1-4} \quad X_n = \frac{1}{T} \int_0^T x(t) \cdot e^{-j\frac{2\pi nt}{T}} dt \quad n = -\infty, \dots, +\infty$$

En fait, ces coefficients minimisent l'écart quadratique entre la fonction  $x(t)$  et le développement (Éq 1-3): la valeur  $X_n$  est obtenue en dérivant par rapport au coefficient d'indice  $n$  l'expression:

$$\text{Éq 1-5}$$

$$\int_0^T \left( x(t) - \sum_{m=-\infty}^{+\infty} X_m e^{-j\frac{2\pi mt}{T}} \right)^2 dt$$

et en annulant cette dérivée.

Ces coefficients de la série de Fourier définissent le spectre de raies de  $x(t)$ . La distance entre deux raies adjacentes est  $f_1 = \frac{1}{T}$  (fréquence du signal périodique, qu'on appelle la fréquence du fondamental).

Ainsi, les signaux élémentaires qui résultent de la décomposition d'un signal **périodique** ont :

- des amplitudes  $X_n$ , en général, complexes ( $X_n = |X_n|e^{-j\theta_n}$ ) et,
- des fréquences  $\frac{n}{T}$  (multiples entiers de la fréquence du fondamental) qu'on appelle les harmoniques; ils couvrent un ensemble discret de l'espace des fréquences.

La représentation de  $|X_n|$  en fonction de la fréquence est appelée "Spectre d'amplitude" du signal alors que celle de  $\theta_n$  en fonction de la fréquence est appelée "Courbe de phase" du signal. Ensemble, elles constituent le spectre fréquentiel du signal. Puisque  $n$  est de type entier, le spectre en fréquence d'un signal périodique n'existe qu'à des fréquences discrètes  $\frac{n}{T}$ . Celles-ci se rapportent à un spectre discontinu en fréquence ou "Spectre de raies".

Une propriété importante est exprimée par l'égalité de Bessel-Parseval qui traduit le fait que dans la décomposition du signal il y a conservation de la puissance:

$$\text{Éq 1-6}$$

$$\sum_{n=-\infty}^{+\infty} |X_n|^2 = \frac{1}{T} \int_0^T |x(t)|^2 dt$$

**Exemple:** Développement en série de Fourier de la fonction  $i_p(t)$  constituée par une suite d'impulsions, séparées par la durée  $T$ , de largeur  $\tau$  et d'amplitude  $a$ , dont l'une est centrée sur l'origine des temps (Figure 1-1). En appliquant (Éq 1-3), ses coefficients de Fourier sont alors :

$$\text{Éq 1-7}$$

$$X_n = \frac{1}{T} \int_0^T i_p(t) \cdot e^{-j\frac{2\pi nt}{T}} dt = \frac{1}{T} \int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} a \cdot e^{-j\frac{2\pi nt}{T}} dt = \frac{a\tau}{T} \frac{\sin\left(\frac{n\pi\tau}{T}\right)}{\frac{n\pi\tau}{T}}$$

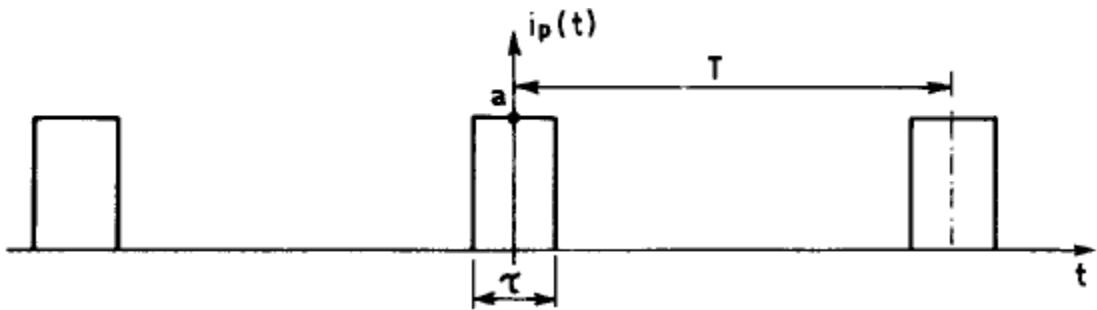


Figure 1-1. Suite d'impulsions périodique et de période T

et donc le développement en série de Fourier de  $i_p(t)$  est alors :

Éq 1-8

$$i_p(t) = \frac{a\tau}{T} \sum_{n=-\infty}^{+\infty} \frac{\sin\left(\frac{n\pi\tau}{T}\right)}{\frac{n\pi\tau}{T}} e^{j\frac{2\pi nt}{T}} = \frac{a\tau}{T} \sum_{n=-\infty}^{+\infty} \text{sinc}\left(\frac{n\pi\tau}{T}\right) \cdot e^{j\frac{2\pi nt}{T}}$$

Par contre, **si le signal n'est pas périodique** et de durée très grande, voire infinie, les signaux élémentaires résultant de la décomposition couvrent un domaine continu de l'espace des fréquences. On parle alors de la transformée de Fourier et non plus de la série de Fourier.

### Transformation de Fourier d'une fonction

Lorsque T augmente, la densité des raies du spectre augmente (les raies se rapprochent les unes des autres; la distance entre deux raies adjacentes devient infinitiment petite). Les raies à variable discrète  $\frac{n}{T}$  est alors remplacée par une **densité spectrale** de raies de variable continue f:

Éq 1-9

$$T \cdot X_n \xrightarrow{T \rightarrow +\infty} X(f)$$

La somme sur la variable discrète n devient une intégrale sur la variable continue f. Ainsi, l'expression (Éq 1-3) devient :

Éq 1-10

$$x(t) = \int_{-\infty}^{+\infty} X(f) \cdot e^{j2\pi ft} df$$

avec

Éq 1-11

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j2\pi ft} dt$$

La fonction  $X(f)$  est la **transformée de Fourier** de  $x(t)$ . Plus communément,  $X(f)$  est appelée spectre du signal  $s(t)$ . La relation (Éq 1-10) est la **transformée de Fourier inverse** qui donne le signal temporel à partir de son spectre.

$X(f)$  existe pour tout signal absolument intégrable sur  $\Re : \left| \int_{-\infty}^{+\infty} x(t) dt \right| < +\infty$ . C'est le cas des

signaux physiques (Amplitude bornée et support borné) et en particulier des **signaux à énergie finie**.

**Exemple:** soit à calculer la transformée de Fourier  $I(f)$  d'une impulsion isolée  $i(t)$  (Figure 1-2), appelée aussi fonction "Porte" ou "Rectangle", de largeur  $\tau$ , d'amplitude a et centrée sur l'origine du temps.

$$\text{Éq 1-12 } I(f) = \int_{-\infty}^{+\infty} i(t) \cdot e^{-j2\pi ft} dt = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} a \cdot e^{-j2\pi ft} dt = a\tau \cdot \frac{\sin(\pi f\tau)}{\pi f\tau} = a\tau \cdot \text{sinc}(\pi f\tau)$$

La Figure 1-3 représente la fonction  $I(f)$ , qui sera très fréquemment utilisée par la suite. Il est important de remarquer qu'elle s'annule aux fréquences multiples entiers non nuls de l'inverse de la durée de l'impulsion.

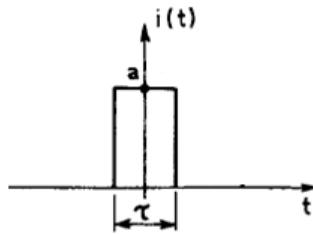


Figure 1-2 Impulsion isolée  $i(t)$  de largeur  $\tau$

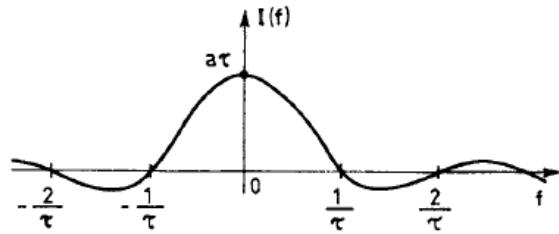


Figure 1-3 Spectre  $I(f)$  de l'impulsion isolée

### *Relations entre coefficients de Fourier et spectre de Fourier*

La correspondance entre coefficients de Fourier et spectre apparaît nettement sur cet exemple. En effet, en rapprochant les relations (Éq 1-7) et (Éq 1-12) on vérifie que, au facteur près, les coefficients du développement en série de Fourier de la suite d'impulsions correspondent aux valeurs que prend le spectre de l'impulsion isolée aux fréquences multiples entiers de l'inverse de la période des impulsions. En fait, on a la relation:

Éq 1-13

$$T \cdot X_n = X\left(\frac{n}{T}\right)$$

### *Conservation de l'énergie du signal dans le spectre de Fourier*

De même, une relation comparable à l'égalité de Bessel-Parseval existe pour une fonction non périodique. Dans ce cas, c'est non plus la puissance mais l'énergie du signal qui se trouve conservée :

Éq 1-14

$$\int_{-\infty}^{+\infty} |X(f)|^2 df = \int_{-\infty}^{+\infty} |x(t)|^2 dt$$

### *Quelques propriétés de la transformée de Fourier (TF)*

**Signal réel:** La TF d'un signal  $x(t)$  réel est une grandeur complexe dont la partie réelle est paire et la partie imaginaire est impaire:  $X(-f) = X^*(f)$ .

*Si  $x(t)$  est réel et pair,  $X(f)$  est également réelle et paire.*

**Linéarité :**

$$a x(t) + b y(t) \Leftrightarrow a X(f) + b Y(f)$$

**Conjugaison Complexé :**

$$x^*(t) \Leftrightarrow X^*(-f)$$

**Convolution / Multiplication :**

$$x(t) * y(t) \Leftrightarrow X(f) \cdot Y(f)$$

$$x(t) \cdot y(t) \Leftrightarrow X(f) * Y(f)$$

**Retard / Translation :**

$$x(t-t_0) \Leftrightarrow X(f) \cdot e^{-j2\pi f t_0}$$

$$x(t) \cdot e^{j2\pi f_0 t} \Leftrightarrow X(f-f_0)$$

**Modulation:**

$$x(t) \cdot \cos(2\pi f_0 t) \Leftrightarrow \frac{1}{2} [X(f + f_0) + X(f - f_0)]$$

**Dérivée:**

$$d^n x(t)/dt^n \Leftrightarrow (j2\pi f)^n X(f)$$

**Facteur d'échelle:**

$$x(at) \Leftrightarrow \frac{1}{|a|} X\left(\frac{f}{a}\right)$$

$$\begin{array}{lll} \textbf{Impulsion de Dirac:} & \delta(t) & \leftrightarrow 1 \\ & 1 & \leftrightarrow \delta(f) \end{array}$$

Une propriété essentielle de la transformation de Fourier, qui est en fait la principale raison de son utilisation, est qu'elle transforme un produit de convolution en un produit simple. En effet, le produit de convolution  $y(t)$  des deux fonctions du temps  $x(t)$  et  $h(t)$  [dont les transformées de Fourier sont respectivement  $X(f)$  et  $H(f)$ ], est défini par :

$$\text{Eq 1-15} \quad y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(t-\tau).h(\tau) d\tau = \int_{-\infty}^{+\infty} x(\tau).h(t-\tau) d\tau$$

La transformée de Fourier de ce produit s'écrit :

$$\text{Eq 1-16} \quad Y(f) = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} x(t-\tau).h(\tau) d\tau \right) e^{-j2\pi f t} dt = \int_{-\infty}^{+\infty} x(u) e^{-j2\pi f u} du \cdot \int_{-\infty}^{+\infty} h(\tau) e^{-j2\pi f \tau} d\tau = X(f).H(f)$$

Réiproquement, on montre que la transformée de Fourier d'un produit simple est un produit de convolution.

### *Impulsion de Dirac*

L'impulsion de Dirac  $\delta(t)$ , dite aussi fonction impulsion unité, n'est pas une fonction ordinaire. On l'appelle souvent fonction généralisée ou distribution. On peut la définir comme étant la limite de l'impulsion isolée  $i(t)$  de largeur  $\tau$ , d'amplitude  $a = \frac{1}{\tau}$  et définie sur l'intervalle  $[-\frac{\tau}{2}, +\frac{\tau}{2}]$ , lorsque  $\tau$  tend vers 0.

$$\text{Eq 1-17} \quad \delta(t) = \lim_{\tau \rightarrow 0} i(t)$$

Comme  $\int_{-\infty}^{+\infty} i(t) dt = 1$ ,  $\int_{-\infty}^{+\infty} \delta(t) dt = 1$ .  $\delta(t)$  est nulle partout sauf à  $t=0$ : c'est une impulsion centrée à  $t=0$ , d'amplitude infiniment grande et de largeur infiniment petite mais dont la surface est finie et vaut 1.

On montre que :

$$\text{Eq 1-18} \quad \int_{-\infty}^{+\infty} x(t) \cdot \delta(t-t_0) dt = x(t_0)$$

Ceci n'est autre que le produit de convolution de  $x(t)$  avec  $\delta(t)$ , calculé au point  $t_0$ :

$$\text{Eq 1-19}$$

$$y(t_0) = x(t) * \delta(t)|_{t=t_0} = \int_{-\infty}^{+\infty} x(\tau) \cdot \delta(t_0 - \tau) d\tau = x(t_0)$$

Cela signifie que  $\delta(t)$  est l'élément neutre du produit de convolution.

### *Relation entre Impulsion de Dirac et fonction de Heavyside*

La fonction de Heavyside  $\Gamma(t)$  ou fonction "échelon" est définie par:

$$\text{Eq 1-20} \quad \Gamma(t) = \begin{cases} 1 & \text{pour } t \geq 0 \\ 0 & \text{pour } t < 0 \end{cases}$$

Cette fonction peut être présentée comme étant la limite de la fonction  $g(t)$ , donnée par la Figure 1-4, lorsque  $\tau$  tend vers 0 :

$$\text{Eq 1-21} \quad \Gamma(t) = \lim_{\tau \rightarrow 0} g(t)$$

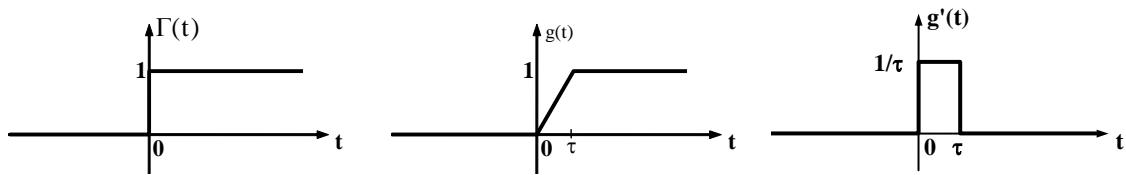


Figure 1-4 : Représentations graphiques des fonctions  $\Gamma(t)$ ,  $g(t)$  et  $i(t)$ .

Or, la dérivée de  $g(t)$  par rapport à  $t$  n'est autre que l'impulsion de l'Éq 1-17 (voir Figure 1-4) dont la limite lorsque  $\tau$  tend vers 0 est  $\delta(t)$ . D'où on peut écrire les relations suivantes :

Éq 1-22

$$\Gamma(t) = \lim_{\tau \rightarrow 0} g(t) \quad \text{et} \quad \delta(t) = \lim_{\tau \rightarrow 0} g'(t) \quad \Rightarrow \quad \frac{d\Gamma(t)}{dt} = \delta(t) \quad \text{ou} \quad \Gamma(t) = \int_{-\infty}^t \delta(u) \cdot du$$

## 1.2 Signaux Usuels

Les signaux temporels sont définis par une fonction du temps  $s(t)$ . Cette fonction peut être une expression analytique ou la solution d'une équation différentielle, auquel cas le signal est appelé déterministe puisque son évolution, en fonction de la variable indépendante  $t$ , peut être prédictée parfaitement par la représentation mathématique appropriée.

Les signaux déterministes n'ont pas beaucoup d'intérêt dans le sens où ils ne portent pas d'information. Par contre, ils ont un intérêt certain dans l'élaboration des différentes techniques et méthodes de traitement des signaux.

Les signaux de ce type les plus utilisés sont les signaux sinusoïdaux tels que  $s(t) = A \cos(\omega t + \alpha)$  où  $A$  est l'amplitude,  $\omega = 2\pi f$  la pulsation (ou fréquence angulaire) et  $\alpha$  la phase du signal.

Ils sont faciles à reproduire, à reconnaître aux différents points d'un système et offrent une possibilité de visualisation simple des caractéristiques. De plus, ils servent de base à la décomposition d'un signal déterministe quelconque, par l'intermédiaire de la Transformation de Fourier.

À ce niveau, ils revêtent une importance particulière dans l'étude des systèmes linéaires et invariants dans le temps (SLIT) qui obéissent au principe de superposition. Ainsi, ces systèmes peuvent être caractérisés uniquement par leurs réponses en fréquence  $H(\omega)$ .

Pour chaque valeur de  $\omega$ , donc de la fréquence,  $H(\omega)$  est un nombre complexe dont le module est l'amplitude de la réponse. Par convention, on désigne par phase de la réponse du système la fonction  $\varphi(\omega)$  telle que:

Éq 1-23

$$H(\omega) = |H(\omega)| \cdot e^{-j\varphi(\omega)}$$

Cette convention permet d'exprimer le temps de propagation de groupe  $\tau(\omega)$ , fonction positive dans les systèmes réels, par :

Éq 1-24

$$\tau(\omega) = \frac{d\varphi(\omega)}{d\omega}$$

Le temps de propagation de groupe fait référence aux lignes de transmission, sur lesquelles les différentes fréquences d'un signal se propagent à des vitesses différentes, ce qui entraîne une dispersion dans le temps de l'énergie du signal.

Le temps de propagation de groupe caractérise donc la dispersion apportée à un signal par une ligne de transmission ou un système équivalent.

En appliquant au système le signal sinusoïdal  $s(t)$ , on obtient en sortie le signal résultant  $s_r(t)$  tel que :

$$\text{Éq 1-25} \quad s_r(t) = A \cdot |H(\omega)| \cdot \cos [\omega t + \alpha - \varphi(\omega)]$$

C'est encore un signal sinusoïdal et la comparaison avec le signal appliqué permet une visualisation de la réponse du système. On imagine aisément l'importance de cette procédure pour les opérations de test par exemple.

Les signaux déterministes ne représentent pas très bien les signaux réels, car ils ne portent pas d'information, si ce n'est pas leur présence même. Les signaux réels sont généralement aléatoires dont le comportement est imprévisible. Ils sont souvent caractérisés par leurs propriétés statistiques et fréquentielles.

En effet, un signal aléatoire est défini à chaque instant  $t$  par la loi de probabilité de son amplitude  $s(t)$ . Cette loi peut s'exprimer par une densité de probabilité  $p(x, t)$  définie comme suit :

$$\text{Éq 1-26} \quad p(x, t) = \lim_{\Delta x \rightarrow 0} \frac{\text{Proba } [x \leq s(t) \leq x + \Delta x]}{\Delta x}$$

Il est stationnaire si ces propriétés statistiques sont indépendantes du temps, c'est-à-dire que sa densité de probabilité est indépendante du temps :  $p(x, t) = p(x)$ .

### 1.3 Étapes de la Numérisation

La variable indépendante de la représentation mathématique d'un signal peut être continue ou discrète. Dans le premier cas, le signal correspondant est appelé **signal analogique**, alors que, dans le second cas, il est appelé **signal discret** ou **signal échantillonné**.

De même, l'amplitude d'un signal peut également être continue ou discrète :

- Elle est continue lorsqu'elle peut avoir une infinité de valeurs.
- Si le nombre de valeurs possibles est limité (ensemble fini), on dit que le signal est à amplitude discrète.

Un signal analogique, dont l'amplitude est discrète, est appelé **signal quantifié**. Un signal discret, dont l'amplitude est discrète aussi, est appelé **signal numérique** (Tableau 1).

Tableau 1 : Types des signaux selon la continuité de la variable et de l'amplitude

Variable Amplitude	Continue	Discrète
Continu	Analogique	Discret
Discrète	Quantifié	Numérique

Lorsqu'on représente les valeurs d'amplitude des échantillons d'un signal numérique par des valeurs binaires (0 et 1), on obtient un **signal binaire**.

Ainsi, l'**échantillonnage** d'un signal analogique consiste à prélever, périodiquement avec une période (ou un pas) d'échantillonnage ( $T_e$ ), un échantillon qui représente l'état du signal à l'endroit ou au moment du prélèvement. Le nombre d'échantillons prélevés par une unité (de temps ou

d'espace) est égale à la fréquence d'échantillonnage:  $F_e = \frac{1}{T_e}$ .

Le signal obtenu est un signal discret (suite d'échantillons) à amplitude continue. La **quantification**, quant à elle, consiste à réduire le nombre des valeurs que peut prendre l'amplitude des échantillons dans un ensemble fini des valeurs appelés **niveaux de quantification**. Le signal obtenu est un signal numérique (suite de valeurs numériques discrètes).

Le **codage** consiste, enfin, à associer un mot-code à chaque niveau de quantification et à remplacer ensuite l'amplitude quantifiée de chaque échantillon par le mot-code du niveau de quantification correspondant. Le signal obtenu est un signal binaire (suite de bits: 0 et 1).

L'échantillonnage suivi de la quantification constituent ce qu'on appelle **numérisation** du signal analogique (Figure 1-5). Suivie du codage, cette chaîne devient une **binarisation** (conversion analogique-numérique) du signal analogique.

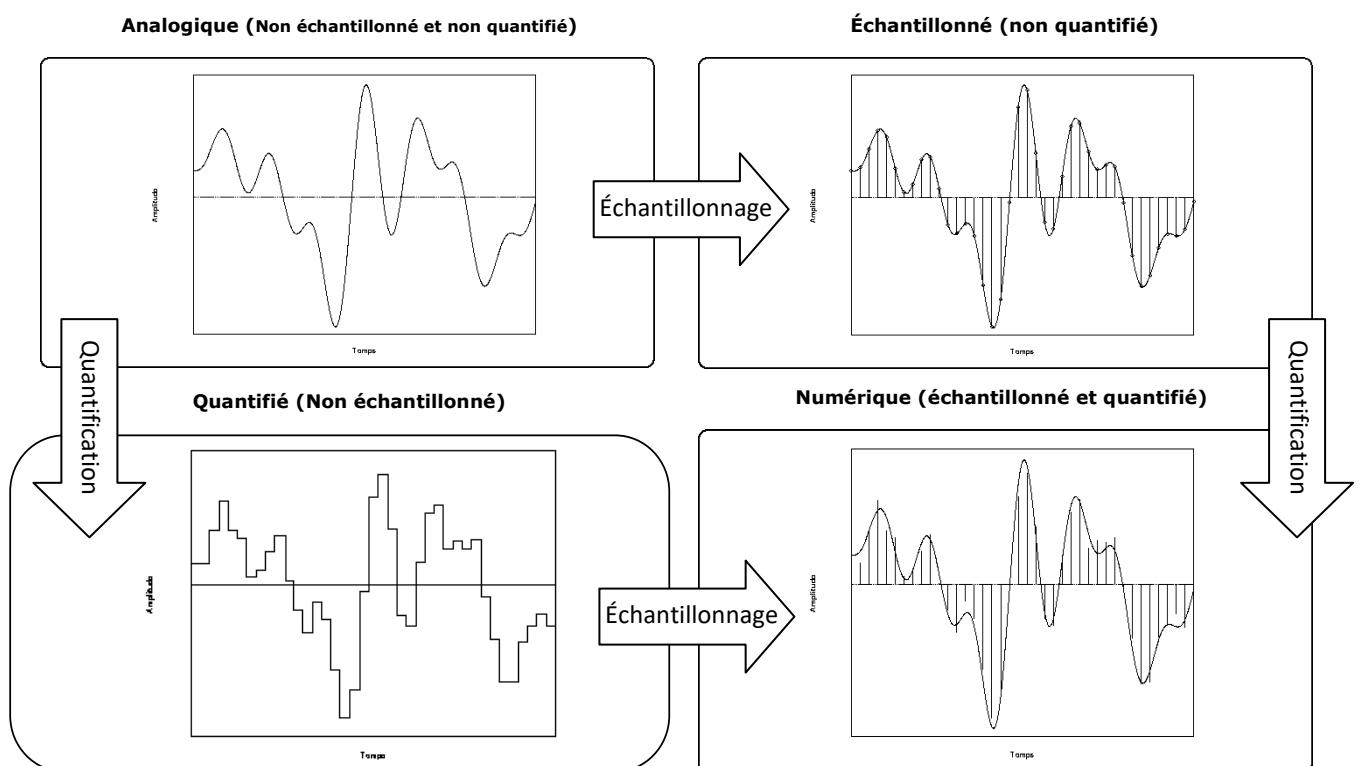


Figure 1-5 Types de signaux selon la continuité de la variable et de l'amplitude

Un système numérique de traitement de l'information agit sur un signal numérique à l'entrée pour produire, après traitement, un signal numérique à la sortie.

Ceci nécessite une opération préliminaire de numérisation de tout signal analogique avant d'être traité par un tel système.

Lorsque l'information traitée doit être restituée sous forme analogique, on procède à une conversion numérique-analogique du signal de sortie.

Une reconstitution fidèle du signal analogique à partir de sa représentation numérique implique une contrainte sur le choix de la fréquence d'échantillonnage utilisée. Cette contrainte découle du théorème d'échantillonnage de Shannon.

La reconstitution est donc l'opération inverse de l'échantillonnage.

### Échantillonnage des signaux:

L'échantillonnage d'un signal  $x(t)$ , avec une fréquence  $f = 1/T$ , peut être décrit mathématiquement par le produit du signal avec une distribution périodique de Dirac qu'on appelle fonction d'échantillonnage (Voir l'illustration de la figure 5):

Eq 1-27

$$x_e(t) = x(t) \cdot \delta_T(t) \text{ avec } \delta_T(t) = \sum_{k=-\infty}^{+\infty} \delta(t - kT)$$

$\delta_T(t)$  est une suite d'impulsions appelée peigne de Dirac, de période  $T$ .

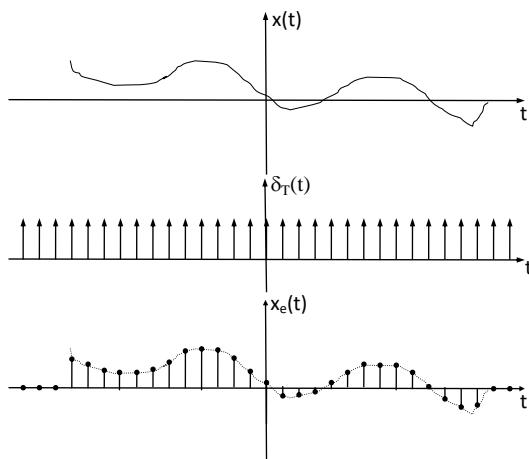


Figure 1-6 : illustration de l'opération d'échantillonnage

L'analyse de cette relation dans le domaine fréquentiel permet de définir les "bonnes conditions" d'échantillonnage pour une classe de signaux dont le spectre est limité.

Étant périodique et de période  $T$ , le peigne de Dirac peut être développé en série de Fourier comme suit:

Eq 1-28

$$\delta_T(t) = \sum_{n=-\infty}^{+\infty} \Delta_n e^{j \frac{2\pi n t}{T}}$$

avec

$$\Delta_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \delta(t) \cdot e^{-j \frac{2\pi n t}{T}} dt = \frac{1}{T}$$

En remplaçant  $\Delta_n$  par sa valeur et en utilisant la propriété de la translation, la TF de  $\delta_T(t)$  est donnée par:

Eq 1-29

$$F\{\delta_T(t)\} = \frac{1}{T} \sum_{n=-\infty}^{+\infty} \delta(f - n/T) = \frac{1}{T} \delta_{1/T}(f)$$

C'est aussi un peigne de Dirac, de poids  $1/T$  et de période  $1/T$  sur l'axe de fréquence. Il en résulte que la TF du signal échantillonné  $x_e(t)$  est liée à la TF du signal continu  $x(t)$  par un produit de convolution:

Eq 1-30

$$X_e(f) = X(f) * F\{\delta_T(t)\} = \frac{1}{T} \sum_{n=-\infty}^{+\infty} X(f - n/T)$$

Cette relation traduit le fait que le spectre du signal original  $X(f)$  se trouve répliqué sur l'axe de fréquence toutes les  $1/T$ .

L'échantillonnage temporel ou spatial se traduit donc dans le domaine fréquentiel par une périodisation du spectre:  $X_e(f)$  est constitué de la somme du spectre original  $X(f)$  et de tous les spectres secondaires qui sont des répliques du spectre original translaté de  $n/T$ .

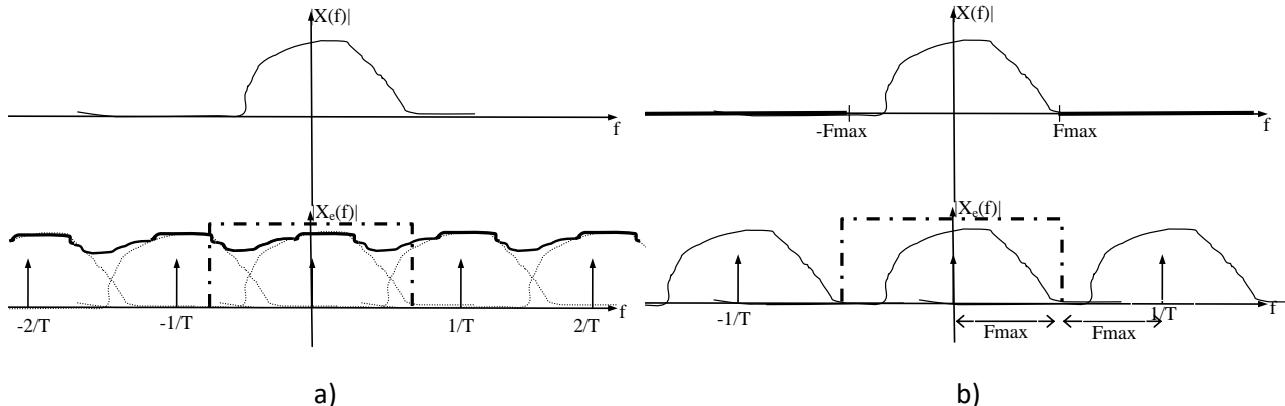


Figure 1-7 : Effet de l'échantillonnage sur le spectre : a) Chevauchement spectral. b) Respect de la condition de Shannon

On ne peut obtenir à nouveau le signal original que si l'on peut éliminer, par filtrage idéal, les spectres secondaires. Ceci est en principe réalisé au moyen d'un filtre passe-bas analogique.

La distance qui sépare les centres de deux spectres adjacents est  $1/T$ ; si le spectre principal possède des fréquences supérieures à  $F_{max} = 1/(2T)$ , alors les spectres secondaires se chevauchent, et un **recouvrement spectral** (aliasing effect, en anglais) se produit, perdant ainsi l'information portée par les fréquences concernées (Figure 1-7).

Dans ce cas, même avec un filtre idéal, il serait impossible de restituer le spectre du signal analogique  $x(t)$ . D'où, le théorème d'échantillonnage de Shannon qui annonce:

*Un signal analogique  $x(t)$  ayant une largeur de bande finie limitée à  $F_{max}$  ne peut être reconstitué exactement à partir de ses échantillons  $x(kT)$  que si ceux-ci ont été prélevés avec une période d'échantillonnage  $T$  inférieure ou égale à  $1/(2F_{max})$ .*

En d'autres termes, la condition de réversibilité est assurée si la fréquence d'échantillonnage  $F_e \geq 2F_{max}$ .

### Limitations pratiques

Spectre fini :

Tout signal à énergie finie, donc physiquement réalisable, ne peut être simultanément à bande limitée. Ce résultat est un corollaire du théorème de Paley-Wiener.

L'échantillonnage d'un signal physiquement réalisable entraîne donc toujours un certain recouvrement qui exclut toute possibilité de réversibilité parfaite. Or, la condition d'énergie finie impose, toutefois, que le spectre tende vers zéro lorsque  $|f|$  tend vers l'infini. Il existe, par

conséquent, une fréquence au-delà de laquelle le spectre est quasiment nul, d'où la possibilité de choisir une cadence d'échantillonnage ne provoquant qu'un recouvrement négligeable et garantissant une réversibilité acceptable.

#### Lissage dû à l'échantillonnage pratique:

De plus, un système pratique d'échantillonnage ne prélève l'intensité en un seul point mais va intégrer le signal sur une petite zone. Cette intégration se traduit par une perte de détails qui entraîne une suppression des hautes fréquences et donc un filtrage passe-bas dans le domaine fréquentiel. En contre partie, ce filtrage passe-bas permet de réduire les erreurs liées aux phénomènes de recouvrement spectral.

#### Effet du bruit de fond:

En effet, le spectre du signal à échantillonner contient souvent une composante à large bande due à la présence additionnelle de bruit de fond généré dans le milieu de mesure, le capteur, les circuits d'amplification, etc.

Il est alors indispensable d'introduire un pré-filtrage du signal analogique avant de procéder à l'échantillonnage afin de supprimer tout risque de recouvrement spectral sans devoir imposer une fréquence d'échantillonnage abusive. Ce filtre d'anti-repliement serait un filtre passe-bas de bande passante  $B < Fe/2$ , en tenant compte de la sélectivité du filtre qui possède généralement une bande de transition non nulle.

Ceci montre la nécessité de bien connaître les caractéristiques spectrales du signal à échantillonner afin de pouvoir effectuer un choix judicieux de la fréquence d'échantillonnage sans risquer de perdre les informations utiles de hautes-fréquences que porte le signal à traiter.

#### Filtre d'interpolation:

Enfin, le filtre passe-bas idéal de restitution (appelé aussi filtre d'interpolation) du signal analogique à partir des échantillons n'est pas réalisable. Une approximation de ce filtre est utilisée pour réaliser l'interpolation.

Pour toutes ces raisons, il est souvent nécessaire d'échantillonner les signaux à des fréquences supérieures à la fréquence théorique de Shannon.

## Quantification

L'opération qui suit l'échantillonnage dans la numérisation des signaux est la quantification de l'amplitude.

Pour un signal analogique, l'amplitude varie d'une manière continue entre deux limites Min et Max qui déterminent sa **gamme dynamique** (ou dynamique, tout court).

La quantification consiste à diviser la gamme dynamique de l'amplitude en un nombre fini d'intervalles juxtaposés, appelés classes de quantification, et en attribuant à toutes les valeurs d'un intervalle une seule valeur dite quantifiée.

Mathématiquement, la quantification est une application qui fait correspondre à chaque valeur possible de  $x(t)$ , une valeur et une seule  $x_q(t)$  choisie parmi un ensemble fini de niveaux de

quantification.

Cette application est généralement une fonction d'échelle. Le problème consiste à déterminer, pour un quantificateur donné, les classes et les niveaux de quantification correspondants.

#### **Paramètres d'un quantificateur:**

Un quantificateur quelconque est défini par les paramètres suivants:

- Le nombre de niveaux de quantification Q ou nombre de classes.
- Les  $(Q + 1)$  **seuils de décisions**  $s_n$  ( $n = 0, \dots, Q$ ) qui constituent les bornes des classes de quantification.
- Les Q valeurs quantifiées ou **niveaux de quantification**  $q_n$  ( $n = 1, \dots, Q$ ).

On appelle "Pas de quantification"  $\Delta_n$  la distance qui sépare les deux seuils de décision successifs  $s_n$  et  $s_{n-1}$ .

Lorsque  $\Delta_n$  est constant (indépendant de n), le quantificateur correspondant est dit uniforme; il est non uniforme dans le cas contraire.

#### **Types de quantificateurs:**

En fonction du choix des seuils de décision et des niveaux de quantification, on peut distinguer les principaux quantificateurs suivants:

- Quantificateur uniforme Linéaire
- Quantificateur uniforme optimal
- Quantificateur non uniforme optimal.

#### **Quantificateur uniforme:**

Le quantificateur uniforme est caractérisé par une répartition uniforme des seuils de décision  $s_n$  sur toute la gamme dynamique: la même importance est accordée à toutes les régions de la dynamique.

Le pas de quantification  $\Delta$  qui sépare deux seuils de décision successifs est constant:

$$\text{Éq 1-31} \quad \Delta = s_n - s_{n-1} = \frac{\text{Max} - \text{Min}}{Q} \quad n = 1, \dots, Q$$

Un seuil de décision  $s_n$  est alors donné par:

$$\text{Éq 1-32} \quad s_n = n \cdot \Delta + \text{Min} \quad n = 0, \dots, Q$$

Le quantificateur associe à chaque classe, une valeur représentative qui est son niveau de quantification. Ainsi, à la  $n^{\text{ème}}$  classe, formée des valeurs appartenant à l'intervalle  $[s_{n-1}, s_n]$ , on associe le niveau de quantification  $q_n$  tel que  $q_n \in [s_{n-1}, s_n]$ .

On distingue deux types de quantificateur uniforme:

#### **Le quantificateur uniforme linéaire:**

Il consiste à choisir le niveau de quantification  $q_n$  de la classe n au milieu du segment  $[s_{n-1}, s_n]$ :

Éq 1-33

$$q_n = \frac{s_{n-1} + s_n}{2} \quad n = 1, \dots, Q$$

Il est dit linéaire puisque les niveaux de quantification sont donnés par une relation linéaire:

Éq 1-34

$$q_n = n \cdot \Delta + (\text{Min} - \Delta/2) \quad n = 1, \dots, Q$$

$\Delta$  représente aussi, dans ce cas, la distance entre deux niveaux de quantification successifs.

Ainsi, les  $s_n$  et les  $q_n$ , peuvent être déterminés sans tenir compte de la loi de probabilité du signal à quantifier parce qu'ils sont définis à partir de Min, Max et Q seulement, ce qui n'est pas le cas pour d'autres types de quantificateur.

La quantification est **une opération non réversible**: l'approximation faite sur la valeur exacte de l'amplitude introduit donc une distorsion qui dépend autant de la nature du signal que de la loi de quantification adoptée. Le **bruit de quantification** est la différence entre le signal original et le signal quantifié:

Éq 1-35

$$e(t) = x(t) - x_q(t)$$

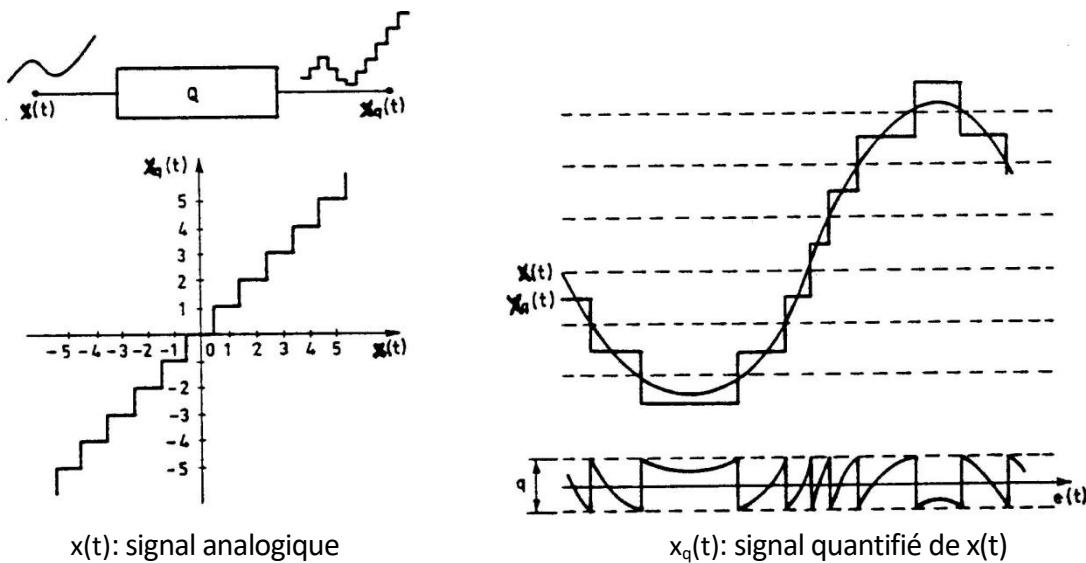


Figure 1-8. Illustration d'une quantification uniforme linéaire

Pour un quantificateur quelconque, la distorsion de quantification ou la puissance de ce bruit est donnée par:

Éq 1-36

$$D = \sum_{n=1}^Q \int_{s_{n-1}}^{s_n} (x - q_n)^2 p(x) dx$$

où  $p(x)$  est la densité (ou loi) de probabilité du signal  $x$ .  $D$  représente la somme des erreurs quadratiques moyennes commises dans chaque classe de quantification.

Remarquons que, pour un quantificateur linéaire uniforme, l'amplitude du bruit de quantification est comprise entre  $-\Delta/2$  et  $+\Delta/2$ . Une illustration de la quantification uniforme linéaire et du bruit de quantification résultant est donnée par la Figure 1-8.

Le quantificateur uniforme optimal:

C'est un quantificateur uniforme du fait que les seuils de décision sont équi-répartis sur la gamme dynamique du signal à quantifier : le pas de quantification  $\Delta$ , qui sépare deux seuils de décision successifs, est une constante.

Il est optimal du fait que, par le choix de  $q_n$  à l'intérieur de chaque classe  $n$ , il minimise la distorsion de quantification dans cette classe.

La distorsion de quantification commise dans une classe  $n$  est donnée par:

Éq 1-37

$$D_n = \int_{s_{n-1}}^{s_n} (x - q_n)^2 p(x) dx$$

Cette distorsion est minimale lorsque sa dérivée partielle par rapport à  $q_n$  est nulle:

Éq 1-38

$$\frac{\partial D_n}{\partial q_n} = -2 \int_{s_{n-1}}^{s_n} (x - q_n) p(x) dx = 0 \Rightarrow q_n = \frac{\int_{s_{n-1}}^{s_n} x p(x) dx}{\int_{s_{n-1}}^{s_n} p(x) dx}$$

pour  $n = 1, \dots, Q$

Pour chaque classe, le niveau représentatif de la classe est obtenu en calculant son barycentre où chaque valeur  $x$  est pondérée par sa densité de probabilité  $p(x)$ .

Ce quantificateur n'est pas linéaire du fait que le niveau de quantification  $q_n$  n'est pas une fonction linéaire de  $n$ , et par conséquent, la distance qui sépare les niveaux de quantification successifs n'est pas constante.

Quantificateur non uniforme optimal:

Certaines applications exigent une quantification qui réduit au mieux la distorsion globale  $D$  pour un nombre de niveaux  $Q$  donné.

Le quantificateur qui vérifie cette contrainte est dit optimal. La distorsion minimale est obtenue en différentiant l'expression Éq 1-36 de  $D$  par rapport à  $q_n$  et  $s_n$  et en annulant simultanément les deux dérivées obtenues, ce qui donne:

Éq 1-39

$$\frac{\partial D}{\partial q_n} = -2 \int_{s_{n-1}}^{s_n} (x - q_n) p(x) dx = 0 \Rightarrow q_n = \frac{\int_{s_{n-1}}^{s_n} x p(x) dx}{\int_{s_{n-1}}^{s_n} p(x) dx}$$

pour  $n = 1, \dots, Q$

Éq 1-40

$$\frac{\partial D}{\partial s_n} = (s_n - q_n)^2 p(s_n) - (s_n - q_{n+1})^2 p(s_n) = 0$$

$$\Rightarrow \text{ si } p(s_n) \neq 0, \quad s_n = \frac{q_n + q_{n+1}}{2} \quad \text{pour } n = 1, \dots, Q-1$$

Le quantificateur obtenu à partir de ces deux systèmes d'équations est connu sous le nom du "quantificateur optimal de Lloyd-Max". Les deux relations Éq 1-39 et Éq 1-40 définissent  $2Q-1$  équations non linéaires à  $2Q-1$  inconnus qui sont fortement interdépendantes. La solution ne peut

être obtenue que par un calcul numérique itératif.

Un exemple d'algorithme de calcul de cette solution est donnée par la Figure 1-9.

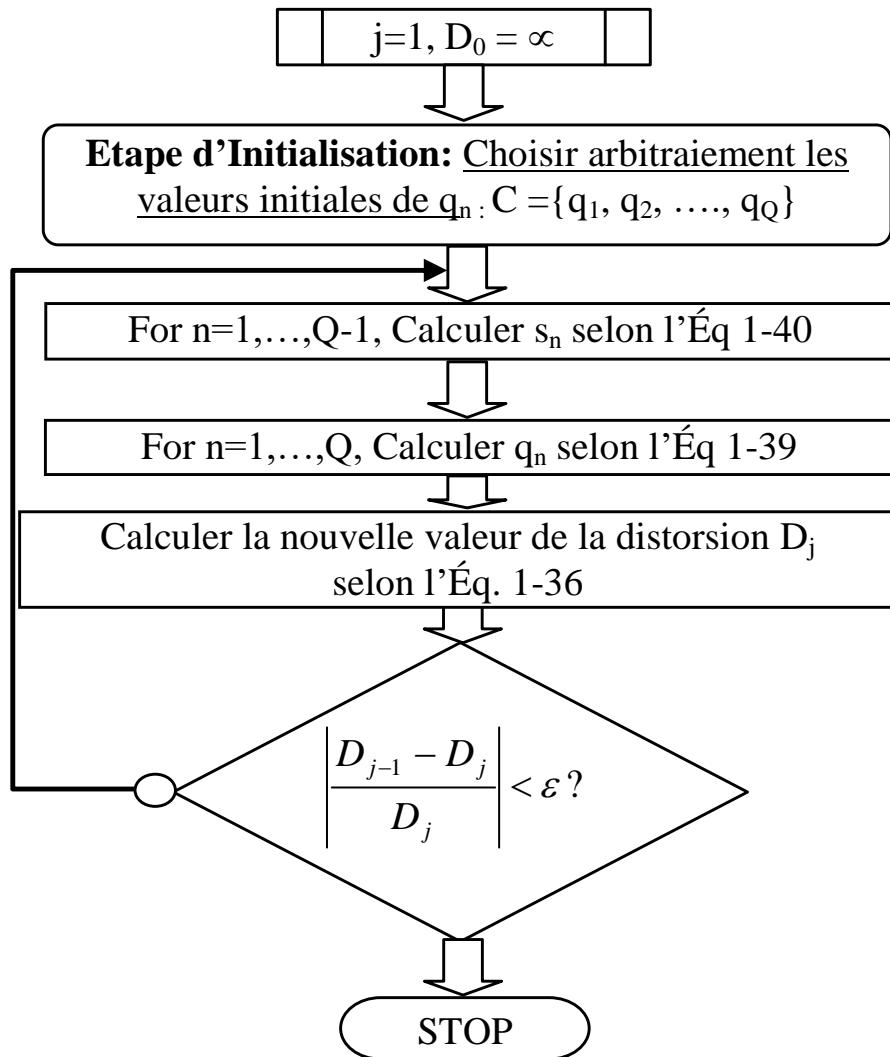


Figure 1-9. Algorithme de calcul itératif d'un quantificateur Lloyd-Max.

# Chapitre 2 - La Transformée de Fourier Discrète

La transformation de Fourier d'un signal échantillonné est une fonction analogique d'une variable continue (la fréquence), et de ce fait, elle n'est pas sous une forme appropriée pour un traitement numérique.

Vu son importance considérable en traitement des signaux, il est nécessaire de la mettre sous une forme pratiquement utilisable. Cette forme est appelée **transformation de Fourier discrète** et est dénotée par son abréviation TFD.

## 2.1 Échantillonnage du spectre:

Le signal échantillonné idéalisé  $x_e(t) = x(t) \cdot \delta_T(t)$  s'écrit aussi sous la forme:

Éq 2-1

$$x_e(t) = \sum_{n=-\infty}^{+\infty} x(nT) \cdot \delta(t - nT)$$

qui, par transformation de Fourier des deux membres, donne une autre forme de  $X_e(f)$ :

Éq 2-2

$$X_e(f) = \sum_{n=-\infty}^{+\infty} x(nT) e^{-j2\pi f nT}$$

qui, en normalisant l'axe du temps par rapport à T (i.e. T = 1), devient:

Éq 2-3

$$X_e(f) = \sum_{n=-\infty}^{+\infty} x(n) e^{-j2\pi f n}$$

$X_e(f)$  est périodique et de période 1. Sous cette dernière forme,  $X_e(f)$  est exprimée par sa représentation en série de Fourier, où  $x(n)$  est alors donné par (attention la variable continue est  $f$  et non pas  $t$ ):

Éq 2-4

$$x(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} X_e(f) e^{j2\pi f n} df$$

Cette relation donne les valeurs des échantillons  $x(n)$  à partir de leur spectre  $X_e(f)$ .

Pour enlever la variable continue, on la discrétise, de manière à conserver la totalité de l'information contenue dans le spectre, comme on l'a déjà fait pour le signal analogique.

Ceci est obtenu en remplaçant  $f$  par  $k\Delta f$ .  $\Delta f$  est le pas d'échantillonnage utilisé sur l'axe des fréquences,  $k$  est un entier. Comme  $X_e(f)$  est périodique et de période 1, on peut exprimer  $\Delta f$  en fonction du nombre  $N$  de valeurs à prélever sur une période du spectre:  $\Delta f = \frac{1}{N}$ .

Dans ce cas, si  $f$  varie entre  $\frac{1}{2}$  et  $+\frac{1}{2}$ ,  $k$  varie de  $-\frac{N}{2}$  à  $\frac{N}{2}-1$ ; en utilisant la méthode des rectangles pour l'intégration numérique, l'intégrale de l'Éq 2-4 est approximée par :

Éq 2-5

$$x(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X_e(k \cdot \Delta f) \cdot e^{j \frac{2\pi k n}{N}} = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X(k \cdot \Delta f) \cdot e^{j \frac{2\pi k n}{N}}$$

Notons que  $X_e(k \cdot \Delta f)$  est égale à  $X(k \cdot \Delta f)$  car, dans le respect de la condition de Shannon, la période principale de  $X_e(f)$  est identique au spectre initial  $X(f)$ . Nous pouvons noter  $X(k \cdot \Delta f)$  par  $X(k)$  tout en gardant à l'esprit que  $X(k)$  est la valeur du spectre du signal à la fréquence  $\frac{k \cdot f_e}{N}$ .

En remplaçant  $f$  par  $\frac{k}{N}$  ( $f_e$  étant normalisée à 1) dans l'Éq 2-3, on obtient  $X(k)$  comme suit :

Éq 2-6

$$X(k) = \sum_{n=-\infty}^{+\infty} x(n) e^{-j \frac{2\pi k n}{N}} \quad k = -\frac{N}{2}, \dots, \frac{N}{2}$$

$X(k)$  est la transformation de Fourier discrète (TFD) de  $x(n)$ . La transformée de Fourier discrète inverse (TFDI) donnant  $x(n)$  à partir de  $X(k)$  est exprimée à partir de l'Éq 2-5 par :

Éq 2-7

$$x(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X(k) \cdot e^{j \frac{2\pi k n}{N}}$$

Il reste à voir comment choisir la valeur de  $N$ :

En examinant l'Éq 2-5, on peut vérifier que le second membre est périodique sur  $n$  et de période  $N$ . Les échantillons  $x(n)$  de l'Éq 2-5 constituent donc une période et une seule de la TFD inverse.

Ainsi, l'expression  $\frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X(k) \cdot e^{j \frac{2\pi k n}{N}}$  représente la TFDI, non pas de  $x(n)$ , mais d'un signal

$x_p(n)$  lié à  $x(n)$  par la relation:

$$\text{Éq 2-8} \quad x_p(n) = \sum_{i=-\infty}^{+\infty} x(iN + n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X(k) \cdot e^{j \frac{2\pi k n}{N}}$$

C'est une répétition périodique et de période  $N$  du signal  $x(n)$ . Ceci veut dire que la TFDI, sur  $N$  points, ne peut exprimer qu'un nombre fini  $N$  d'échantillons du signal de départ.  $x(n)$  doit donc avoir un nombre fini d'échantillons pour qu'on puisse calculer sa TFD.

Si  $M$  est le nombre d'échantillons de  $x$ , alors on doit choisir  $N$  tel que  $N \geq M$ . C'est la condition à respecter sur  $N$  pour avoir une représentation équivalente du signal  $x(n)$  dans le domaine de fréquences discrètes.

Si cette condition n'est pas respectée ( $N < M$ ), un recouvrement, similaire au recouvrement spectral, a lieu entre les échantillons et par conséquent, on ne peut pas extraire  $x(n)$  exactement à partir d'une période de  $x_p(n)$ .

D'autre part, il est évident que les signaux de durée  $M$  inférieure à  $N$  peuvent être considérés comme des signaux de période  $N$ , en prolongeant le signal de  $N-M$  échantillons nuls.

## Dualité temps-fréquence:

Un échantillonnage en fréquence introduit une périodisation dans le temps, exactement comme l'échantillonnage dans le temps a introduit une périodisation dans les fréquences; ce qui confirme la dualité temps-fréquence.

## Différence entre un signal apériodique et un signal périodique

Considérons le cas d'un signal discret  $x(n)$  apériodique de durée finie  $N$ , défini entre  $n_0$  et  $n_0+N-1$ ; sa TFD est une fonction périodique et de période  $N$ ; elle est donnée, sur une période, par:

$$\text{Éq 2-9} \quad X(k) = \sum_{n=n_0}^{n_0+N-1} x(n) e^{-j\frac{2\pi kn}{N}} \quad \text{pour } k = 0, \dots, N-1$$

La transformation inverse est donnée par:

$$\text{Éq 2-10} \quad x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi kn}{N}} \quad \text{pour } n = n_0, \dots, n_0 + N - 1$$

### Remarques:

- La période de  $X(k)$  a été choisie entre 0 et  $N-1$ , mais elle peut être choisie entre n'importe quelles deux limites pourvu que le nombre des points considérés soit égal à  $N$ . Son choix entre  $-N/2$  et  $N/2-1$  assure la représentation bilatérale usuelle, mais une précaution est à prendre dans le cas où  $N$  est impair; les limites dans ce cas seront  $-(N-1)/2$  et  $(N-1)/2$ .
- Les coefficients  $X(k)$  ne comportent aucune information concernant le paramètre  $n_0$ . Ils peuvent être aussi bien la TFD d'un signal périodique que d'un signal apériodique. D'où l'importance de connaître à priori le paramètre  $n_0$  dans le cas d'un signal apériodique, car sinon, l'Éq 2-10 conduit à une période quelconque de  $x_p(n)$  selon la valeur utilisée de  $n_0$ .
- D'après la remarque précédente, les relations de  $X(k)$  et  $x(n)$  restent également valables pour un signal  $x(m)$  périodique et de période  $N$  où il suffit de prendre  $n_0 = 0$ , pour reconstituer une période de  $x(n)$ .

## Cas des signaux réels

On a vu que la partie réelle de la transformée de Fourier d'un signal réel est une fonction paire et que sa partie imaginaire est une fonction impaire:  $X(-f) = X^*(f)$  (ou modules identiques et arguments opposés)

Cette relation est également valable pour les échantillons  $X(k)$  prélevés sur  $X(f)$  à condition de conserver la symétrie de ces échantillons par rapport à  $f = 0$ .

Ceci est garanti dans la relation de  $X(k)$  puisque, pour  $k = 0$ ,  $X(k)$  est la valeur à l'origine de  $X(f)$ . Dans ce cas, on peut mettre la relation de  $X(k)$  sous la forme suivante:

$$\text{Éq 2-11} \quad X(k) = \sum_{n=n_0}^{n_0+N-1} x(n) e^{-j\frac{2\pi kn}{N}} \quad \text{pour } k = 0, \dots, \frac{N}{2}$$

$$\text{et} \quad X(N-k) = X^*(k) \quad \text{pour } k = 1, \dots, \frac{N}{2}-1$$

Ainsi, on peut éviter la moitié des calculs de  $X(k)$  si  $x(n)$  est réel. De plus, il est important de remarquer que dans le cas des signaux réels, les coefficients  $X(0)$  et  $X(\frac{N}{2})$  sont réels:  $\text{Im}\{X(0)\} = \text{Im}\{X(\frac{N}{2})\} = 0$ .

La linéarité de la transformation de Fourier discrète permet de calculer les TFD de deux signaux réels de même taille, en réduisant de moitié les calculs nécessaires si on les calculait séparément, ce qui représente un avantage considérable puisque le temps de calcul dans un système de traitement numérique est proportionnel au nombre d'opérations à effectuer, d'où l'intérêt de réduire le plus possible les calculs redondants. En effet, si  $x_1(n)$  et  $x_2(n)$  sont deux signaux réels de  $N$  échantillons chacun (ou périodiques et de période  $N$ ), on peut former un signal complexe  $x_3(n)$  avec:

$$\text{Éq 2-12} \quad x_3(n) = x_1(n) + j x_2(n) \quad \text{pour } n = 0, \dots, N-1$$

En vertu de la linéarité de la TFD, on peut écrire:

$$\text{Éq 2-13} \quad X_3(k) = X_1(k) + j X_2(k) \quad \text{pour } k = 0, \dots, N-1$$

où  $X_1$ ,  $X_2$ , et  $X_3$  sont des séries complexes. D'où:

$$\text{Éq 2-14} \quad \text{Re}\{X_3(k)\} = \text{Re}\{X_1(k)\} - \text{Im}\{X_2(k)\}$$

$$\text{et} \quad \text{Im}\{X_3(k)\} = \text{Im}\{X_1(k)\} + \text{Re}\{X_2(k)\} \quad \text{pour } k = 0, \dots, N-1$$

Comme  $X_1(k)$  et  $X_2(k)$  sont les TFD de deux signaux réels, alors on peut écrire:

$$\begin{aligned} \text{Éq 2-15} \quad & \text{Re}\{X_1(k)\} = \text{Re}\{X_1(N-k)\} \\ & \text{Im}\{X_1(k)\} = -\text{Im}\{X_1(N-k)\} \quad \text{pour } k = 0, \dots, N-1 \\ \text{et,} \quad & \text{Re}\{X_2(k)\} = \text{Re}\{X_2(N-k)\} \\ & \text{Im}\{X_2(k)\} = -\text{Im}\{X_2(N-k)\} \quad \text{pour } k = 0, \dots, N-1 \end{aligned}$$

Ceci implique:

$$\begin{aligned} \text{Éq 2-16} \quad & \text{Re}\{X_3(N-k)\} = \text{Re}\{X_1(k)\} + \text{Im}\{X_2(k)\} \\ \text{et} \quad & \text{Im}\{X_3(N-k)\} = -\text{Im}\{X_1(k)\} + \text{Re}\{X_2(k)\} \quad \text{pour } k = 0, \dots, N-1 \end{aligned}$$

En additionnant deux à deux les relations Éq 2-14 et Éq 2-16, on obtient:

$$\begin{aligned} \text{Éq 2-17} \quad & \text{Re}\{X_1(k)\} = \frac{1}{2} [\text{Re}\{X_3(k)\} + \text{Re}\{X_3(N-k)\}] \\ \text{et} \quad & \text{Re}\{X_2(k)\} = \frac{1}{2} [\text{Im}\{X_3(k)\} + \text{Im}\{X_3(N-k)\}] \quad \text{pour } k = 0, \dots, \frac{N}{2} \end{aligned}$$

et en les soustrayant deux à deux, on obtient:

$$\begin{aligned} \text{Éq 2-18} \quad & \text{Im}\{X_1(k)\} = \frac{1}{2} [\text{Im}\{X_3(k)\} - \text{Im}\{X_3(N-k)\}] \\ \text{et} \quad & \text{Im}\{X_2(k)\} = \frac{1}{2} [\text{Re}\{X_3(k)\} - \text{Re}\{X_3(N-k)\}] \quad \text{pour } k = 0, \dots, \frac{N}{2} \end{aligned}$$

On a obtenu ainsi, les TFDs des deux signaux réels à partir d'une seule TFD d'un signal complexe. De la même manière et grâce à la linéarité de la TFDI, on peut calculer les transformations de Fourier inverses de deux signaux réels à partir d'une seule TFDI.

## 2.2 Transformée de Fourier Rapide (TFR)

Le calcul direct de la TFD d'ordre N à partir de la relation Éq 2-9 nécessite  $N^2$  opérations de multiplications complexes, ce qui devient trop grand lorsque N augmente. L'importance du rôle joué par la TFD en traitement numérique des signaux est renforcée énormément par l'existence d'un algorithme de calcul rapide. Comme le temps de calcul dans un système de traitement numérique est proportionnel au nombre d'opérations à effectuer, on a intérêt à réduire le plus possible les calculs redondants.

La TFD peut être considérée comme le produit d'une matrice, appelée matrice de transformation, par un vecteur formé par les échantillons d'un signal.

$$\text{Éq 2-19} \quad \begin{pmatrix} X(0) \\ X(1) \\ \vdots \\ \vdots \\ X(N-1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & \cdots & \cdots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & \cdots & \cdots & \cdots & e^{-j\frac{2\pi(N-1)}{N}} \\ 1 & e^{-j\frac{2\pi \cdot 2}{N}} & \ddots & & & e^{-j\frac{2\pi \cdot 2(N-1)}{N}} \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ 1 & e^{-j\frac{2\pi(N-1)}{N}} & & \ddots & & e^{-j\frac{2\pi(N-1)(N-1)}{N}} \end{pmatrix} \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ \vdots \\ x(N-1) \end{pmatrix}$$

L'idée de base d'un algorithme de calcul rapide de la TFD provient de la redondance des éléments de la matrice de transformation. Cette redondance n'est pas quelconque: elle possède une structure bien déterminée dont on peut tirer profit en décomposant cette matrice en d'autres matrices, de taille inférieure, possédant des redondances similaires, ce qui permet de généraliser récursivement ce processus.

Il existe plusieurs algorithmes de calcul rapide de la TFD, appelés ***Transformations de Fourier Rapides*** (TFR ou FFT de leur nom anglais ***Fast Fourier Transforms***), dont le plus connu est celui de Tukey-Cooley du nom de ses inventeurs.

La TFR consiste à décomposer une TFD d'ordre N en m TFD d'ordre  $N_i$  avec  $N = \prod_{i=1}^m N_i$

où  $N_i$  sont des nombres premiers. Le cas le plus simple, et que l'on considère ici, est celui où tous les  $N_i$  sont égaux à 2:  $N = 2^m$ .

Le calcul de la TFR se fait en plusieurs étapes récursives dont la première consiste à décomposer la TFD d'ordre N en 2 TFD d'ordre  $\frac{N}{2}$  chacune. Comme N est un entier pair, on peut partager la suite  $x(n)$  en deux suites de  $\frac{N}{2}$  valeurs, la première étant formée par les valeurs d'indices pairs et la seconde formée par les valeurs d'indices impairs.

Mathématiquement, on peut représenter ce partage en substituant  $n = 2i$  pour les indices pairs et  $n = 2i + 1$  pour les indices impairs. Notons par  $F_k^N(x)$  la k<sup>jème</sup> composante de la TFD calculée sur N points du signal x, et par  $W_N$  la valeur exponentielle  $e^{\frac{j2\pi}{N}}$ .

La relation Éq 2-9 devient, en considérant  $n_0 = 0$  (cas du signal périodique):

$$\begin{aligned} \text{Eq 2-20} \quad F_k^N(x) &= \sum_{n=0}^{N-1} x(n) (\mathbf{W}_N)^{-kn} \\ &= \sum_{i=0}^{\frac{N}{2}-1} x(2i) (\mathbf{W}_N)^{-2ik} + \sum_{i=0}^{\frac{N}{2}-1} x(2i+1) (\mathbf{W}_N)^{-(2i+1)k} \quad k = 0, \dots, N-1 \end{aligned}$$

$$\text{or, } (\mathbf{W}_N)^{-2ik} = e^{-\frac{j2\pi(2i)k}{N}} = e^{-\frac{j2\pi ik}{N/2}} = (\mathbf{W}_{N/2})^{-ik} \text{ ou, en général, } (\mathbf{W}_N)^2 = \mathbf{W}_{N/2}$$

ce qui donne:

$$\text{Eq 2-21} \quad F_k^N(x) = \sum_{i=0}^{\frac{N}{2}-1} x(2i) (\mathbf{W}_{N/2})^{-ik} + (\mathbf{W}_N)^{-k} \cdot \sum_{i=0}^{\frac{N}{2}-1} x(2i+1) (\mathbf{W}_{N/2})^{-ik} \quad k = 0, \dots, N-1$$

Chacune des sommes dans cette expression représente une TFD d'ordre  $\frac{N}{2}$ . La première somme est celle des valeurs d'indices pairs et la seconde est celle des valeurs d'indices impairs du signal original  $x(n)$ . On peut donc écrire:

$$\text{Eq 2-22} \quad F_k^N(x) = F_k^{N/2}(x_p) + (\mathbf{W}_N)^{-k} F_k^{N/2}(x_i) \quad k = 0, \dots, N-1$$

La Figure 2-1 illustre cette relation sachant que  $F_k^{N/2}(x_p)$  et  $F_k^{N/2}(x_i)$  sont périodiques et de période  $\frac{N}{2}$ , c.à.d.:

$$\text{Eq 2-23} \quad F_{\frac{k}{2}+\frac{N}{2}}^N(x_p) = F_{\frac{k}{2}}^N(x_p) \quad \text{et} \quad F_{\frac{k}{2}+\frac{N}{2}}^N(x_i) = F_{\frac{k}{2}}^N(x_i) \quad k = 0, \dots, \frac{N}{2}-1$$

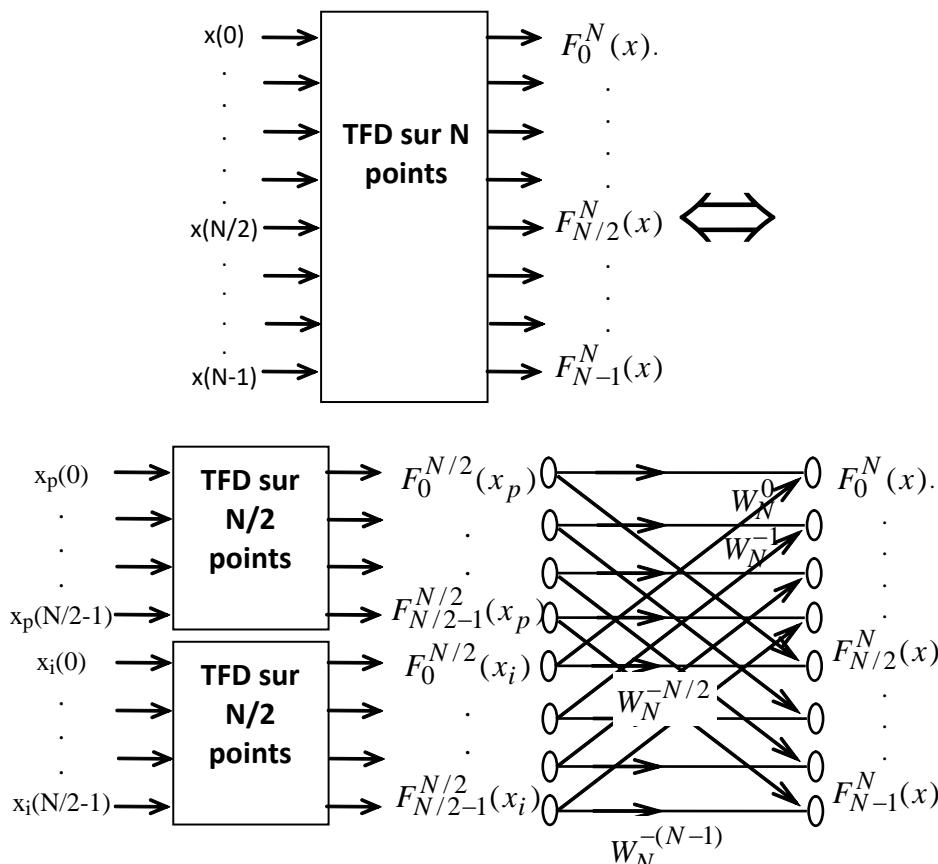


Figure 2-1. Schéma-bloc de décomposition d'une TFD sur N points en 2 TFD de N/2 points chacune.

Comme  $\frac{N}{2}$  est pair, on peut de nouveau partager chacune des TFD d'ordre  $\frac{N}{2}$  en deux TFD d'ordre  $\frac{N}{4}$ , en groupant de nouveau les valeurs d'indices pairs et les valeurs d'indices impairs dans chaque TFD d'ordre  $\frac{N}{2}$ , et on recommence avec les nouvelles TFD jusqu'à l'obtention de  $\frac{N}{2}$  TFD d'ordre 2, ce qui nécessite en total m étages. Ainsi, le calcul est réduit à des TFD d'ordre 2 dont le graphe de fluence est donné par la Figure 2-2. On appelle ce type de graphe "papillon".

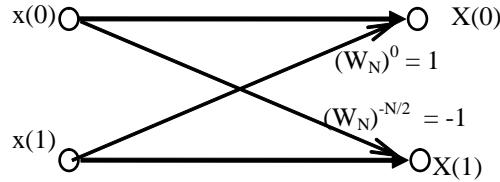


Figure 2-2. Graphe de fluence d'une TFD sur 2 points

Reprendons la relation Éq 2-22 où l'on remarque que  $F_k^N(x)$  et  $F_{k+N/2}^N(x)$  s'expriment toutes les 2 en fonction de  $F_k^{N/2}(x_p)$  et  $F_k^{N/2}(x_i)$  (si on tient compte de la périodicité exprimée par Éq 2-23):

$$\begin{aligned} \text{Éq 2-24} \quad F_{k+N/2}^N(x) &= F_k^{N/2}(x_p) + (W_N)^{-(k+\frac{N}{2})} \cdot F_k^{N/2}(x_i) \\ &= F_k^{N/2}(x_p) - (W_N)^{-k} F_k^{N/2}(x_i) \quad \text{pour } k = 0, \dots, \frac{N}{2}-1 \end{aligned}$$

puisque  $(W_N)^{-(k+\frac{N}{2})} = e^{-\frac{j2\pi(k+\frac{N}{2})}{N}} = e^{-j\pi} \cdot e^{-\frac{j2\pi k}{N}} = -(W_N)^{-k}$

Ainsi le papillon reliant  $F_k^N(x)$ ,  $F_{k+N/2}^N(x)$ ,  $F_k^{N/2}(x_p)$  et  $F_k^{N/2}(x_i)$  est celui de la Figure 2-3.

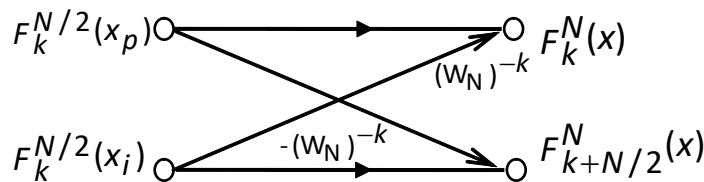


Figure 2-3. Graphe de fluence liant  $F_k^N(x)$ ,  $F_{k+N/2}^N(x)$ ,  $F_k^{N/2}(x_p)$  et  $F_k^{N/2}(x_i)$

On remarque le produit  $(W_N)^{-k} F_k^{N/2}(x_i)$  est calculé deux fois. On peut éviter ceci en calculant une fois ce produit. Ce papillon devient alors celui de la Figure 2-4.

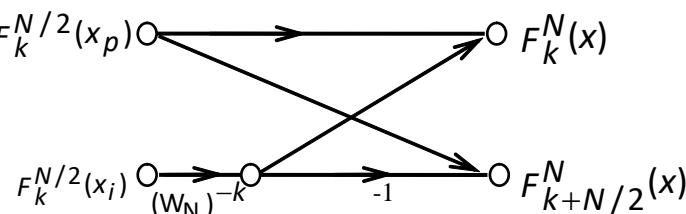


Figure 2-4. Le Graphe de fluence de la Figure 2-3 réduit à une multiplication

Un papillon consiste donc en une multiplication complexe et deux additions complexes aussi. Chaque étage de calcul de la TFR fait apparaître  $\frac{N}{2}$  papillons et donc nécessite  $\frac{N}{2}$  multiplications et  $N$  additions complexes. Ainsi pour  $m = \log_2 N$  étages, il faut  $\frac{N}{2} \cdot \log_2 N$  multiplications et  $N \cdot \log_2 N$  additions complexes. Si on veut aller plus loin en réduction, on constate que tous les papillons du premier étage (TFD d'ordre 2), ne comportent aucune multiplication et que dans les autres étages, les deux coefficients,  $(W_N)^0 = 1$  et  $(W_N)^{-N/4} = -j$ , ne correspondent pas à des multiplications, ce qui permet de réduire le nombre de multiplications dans l'étage  $i$  ( $i = 2, \dots, m$ ) de  $2x\frac{N}{2^i}$ , d'où une réduction de  $\frac{N}{2}$  dans le premier étage et  $2 \sum_{i=2}^m \frac{N}{2^i} = N - 2$  pour les autres étages (voir l'exemple de la Figure 2-5 pour  $N=8$ ).

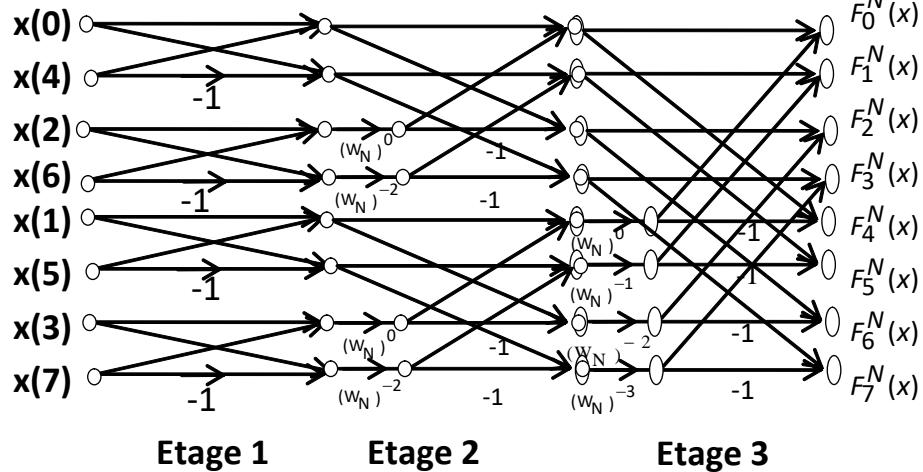


Figure 2-5. Le Graphe de fluence de la FFT sur 8 points

Donc, le nombre total d'opérations nécessaires pour calculer la TFR d'ordre  $N$  est:

$$\text{Eq 2-25} \quad P_N = \frac{N}{2} \log_2 N - \left(\frac{3N}{2} - 2\right) = \frac{N}{2} \log_2 N - \frac{3N}{2} + 2 \text{ multiplications complexes}$$

$$S_N = N \log_2 N \text{ additions complexes}$$

sachant qu'une multiplication complexe nécessite 4 multiplications réelles et qu'une addition complexe nécessite 2 additions réelles, on obtient:

$$\text{Eq 2-26} \quad P_N = 4\left(\frac{N}{2} \log_2 N - \frac{3N}{2} + 2\right) = 2(N \log_2 \frac{N}{8} + 4) \text{ multiplications réelles}$$

$$S_N = 2 N \log_2 N \text{ additions réelles}$$

ce qui représente un gain important par rapport au calcul direct de la TFD d'ordre  $N$  qui nécessite  $4N^2$  multiplications réelles et  $4N^2 - 2N$  additions réelles.

Par exemple, pour  $N = 1024$  ( $m=10$ ):

$$P_N = 14344 \text{ multiplications réelles et } S_N = 20480 \text{ additions réelles,}$$

alors que le calcul direct nécessite 4194304 multiplications réelles et 2095104 additions réelles

ce qui représente un gain de 292,4 en nombre de multiplications et 102,3 en nombre d'additions.

Cet algorithme de Tukey-Cooley est dit de base 2 ("**“radix-2”**) car chaque étape de décomposition divise la série en 2 (N est une puissance de 2), alors qu'une division en 4 sera dite de base 4 ("**“radix-4”**"). Il est possible d'envisager des décompositions à base variable à chaque étape de la décomposition; par exemple, pour  $N = 100 = 2 \times 5 \times 2 \times 5 = 5 \times 4 \times 5$ , etc. L'algorithme correspondant sera dit à base variable ("**“mixed-radix”**").

# Chapitre 3 - Généralisation aux Signaux Bidimensionnels – Cas d'images

Avec les sons, les images constituent l'un des moyens les plus importants qu'utilise l'homme pour communiquer avec ses semblables. La notion de signal englobe aussi bien les signaux monodimensionnels que les signaux multidimensionnels.

La représentation de la luminance  $L$  d'une photographie en fonction des coordonnées  $x$  et  $y$  du plan se fait à l'aide d'un signal bidimensionnel  $L(x,y)$ .

Une image peut être définie comme un ensemble d'informations provenant d'un objet à l'aide d'un dispositif (ou système) de formation de cette image. Quel que soit ce système de formation, une image analogique est une distribution continue d'une grandeur physique  $f$  dans un plan à support continu et borné dont les points sont repérés par les coordonnées  $(x,y)$ .

Ainsi  $f(x,y)$  est la valeur de cette grandeur au point  $(x,y)$ . Cette grandeur, suivant le cas, peut avoir diverses origines physiques: intensité lumineuse visible ou non (rayons X, infrarouge), la mesure de la pression acoustique dans le cas d'une échographie, ... etc.

Un échantillonnage approprié de l'image analogique dans les deux directions spatiales suivi d'une quantification de l'amplitude des échantillons avec un nombre fini de niveaux (appelés niveaux de gris) produit une image numérique. Notons que certains systèmes génèrent directement des images numériques ce qui permet d'éviter la procédure de numérisation. Par exemple, les images de synthèse sont numériques par nature; en médecine les tomographes X, les scanners, les appareils d'Imagerie par Résonance Magnétique (IRM) fournissent des images numériques. Par contre, la radiographie conventionnelle, par exemple, génère des images analogiques qu'il faut numériser pour pouvoir les traiter par des systèmes numériques.

## Échantillonnage Bidimensionnel:

L'échantillonnage spatial consiste à représenter l'image par un nombre fini de points.

Généralement, les échantillons sont prélevés périodiquement avec une période d'échantillonnage bidirectionnelle  $\Delta x$  dans la direction horizontale et  $\Delta y$  dans la direction verticale.

Un tel échantillonnage peut être vu de la façon suivante: on superpose à l'image une grille régulière formée de cellules élémentaires (de forme carrée, hexagonale ou circulaire) dont les centres sont distants de  $\Delta x$  et  $\Delta y$  dans les deux directions (Figure 3-1).

Chaque cellule est représentée par un point-image dit "pixel" (Picture Element). Le nombre de pixels, et par suite sa résolution, dépend du choix de  $\Delta x$  et  $\Delta y$ . Ce choix découle du théorème d'échantillonnage de Shannon qui donne les conditions que doivent satisfaire  $\Delta x$  et  $\Delta y$  pour que l'image échantillonnée représente parfaitement à l'image originale:

*"Un signal analogique  $f(x,y)$ , dont le spectre spatial est borné par les fréquences spatiales horizontale et verticale  $U_{max}$  et  $V_{max}$  (cycles/unité de distance), ne peut être reconstitué exactement à partir de ses échantillons  $f(k\Delta x, l\Delta y)$  que si ceux-ci ont été prélevés avec des*

“ périodes  $\Delta x$  et  $\Delta y$  inférieures ou égales respectivement à  $1/(2U_{max})$  et à  $1/(2V_{max})$ . ”

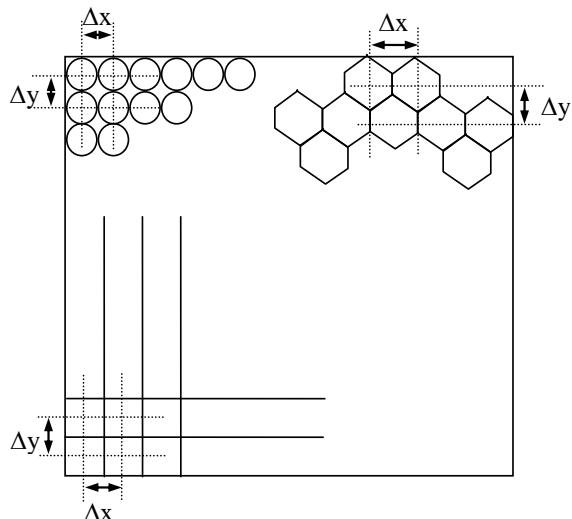


Figure 3-1 : Formes possibles de cellules élémentaires

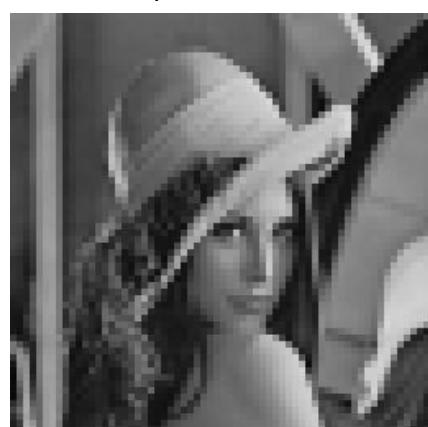
A titre d'exemple, la Figure 3-2 présente 4 images échantillonnées avec différents pas d'échantillonnage et donc avec un nombre variable de pixels par ligne et par colonne. On remarque, sur ces images, que les détails disparaissent au fur et à mesure que les tailles des pas d'échantillonnage augmentent. La résolution peut être définie comme étant la taille du plus petit détail qu'on peut discerner dans une image.



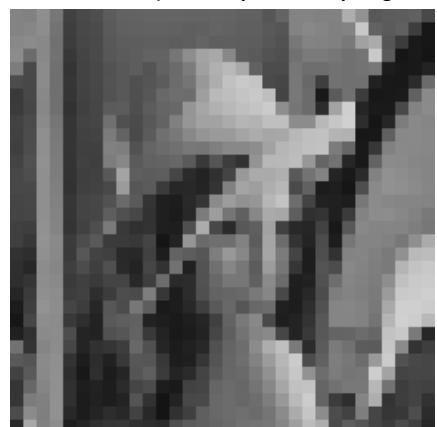
Lena - 256 x 256 pixels



Lena - 128 x 128 (Taille du pixel 4 fois plus grande)



Lena - 64x64 pixels (Taille 16 fois plus grande)



Lena - 32x32 Pixels (Taille 64 fois plus grande)

Figure 3-2- Images échantillonnées avec différents pas

## Notion de fréquence spatiale:

Les fréquences spatiales  $u$  et  $v$  sont les variables de la transformée de Fourier bidimensionnelle  $F(u,v)$  de la fonction  $f(x,y)$ .

Elles sont les “variables réciproques” des variables d’espaces  $x$  et  $y$ .

Dans le cas des signaux temporels, si le temps est exprimé en seconde, la fréquence est exprimée en Hertz ou  $\text{sec}^{-1}$ . De la même façon, et pour les signaux spatiaux, si la variable spatiale est exprimée en mm, sa fréquence spatiale est alors exprimée en  $\text{mm}^{-1}$  ou nombre de cycles/mm.

Pour illustrer intuitivement la notion de fréquence spatiale et son interprétation, considérons le cas d’une ligne de l’image dont l’intensité lumineuse  $L(x)$  varie sinusoïdalement autour de la valeur moyenne  $L_0$ . Si cette ligne contient quatre cycles (ou quatre segments identiques) par unité de longueur, on dit que ce signal spatial est périodique et de période  $1/4$ . Sa représentation fréquentielle est constituée de 3 raies spectrales:

- Une raie à l’origine ( $u=0$ ) d’amplitude  $L_0$  qui correspond à la composante continue ou la moyenne,
- et deux raies aux points  $u=4$  et  $u=-4$  correspondants à la fréquence du signal ou le nombre de cycles par unité de longueur.

Lorsque ce signal contient 8 cycles par unité de longueur, ses niveaux de gris varient deux fois plus rapidement et la fréquence est doublée; les raies de  $\pm 4$  se déplacent alors à  $\pm 8$  alors que la raie de la composante continue n’est pas modifiée. Au contraire, lorsque le signal n’est constitué que de deux cycles par unité de longueur, les niveaux de gris varient deux fois plus lentement et la fréquence est divisée par 2. Les raies fréquentielles se trouvent alors à  $\pm 2$ .

Ceci montre que les basses fréquences correspondent aux variations lentes de l’amplitude du signal (ou de l’image) alors que les hautes fréquences correspondent aux variations brusques ou rapides, c.à.d. aux détails du signal (ou de l’image).

Comme dans le cas des signaux monodimensionnels, l’échantillonnage spatial implique une périodisation du spectre dans le domaine fréquentiel. La Figure 3-3 montre le support d’un spectre fréquentiel d’une image (espace des couples  $(u,v)$  pour lesquelles le spectre est non nul):

- a) avant échantillonnage
- b) après échantillonnage avec deux fréquences supérieures à  $2U_{\max}$  et à  $2V_{\max}$ .
- c) après échantillonnage avec deux fréquences inférieures à  $2U_{\max}$  et à  $2V_{\max}$ .
- d) après échantillonnage avec deux fréquences égales à  $2U_{\max}$  et à  $2V_{\max}$ .

## Représentations fréquentielles

Plusieurs outils mathématiques permettent la représentation des images dans le domaine fréquentiel. Les plus connus sont:

la TFD Bidimensionnelle (TF2D),

la TCD Bidimensionnelle (TC2D).

Une représentation de l'image dans un espace transformé peut être mieux adaptée à certaines applications, comme par exemple, la compression.

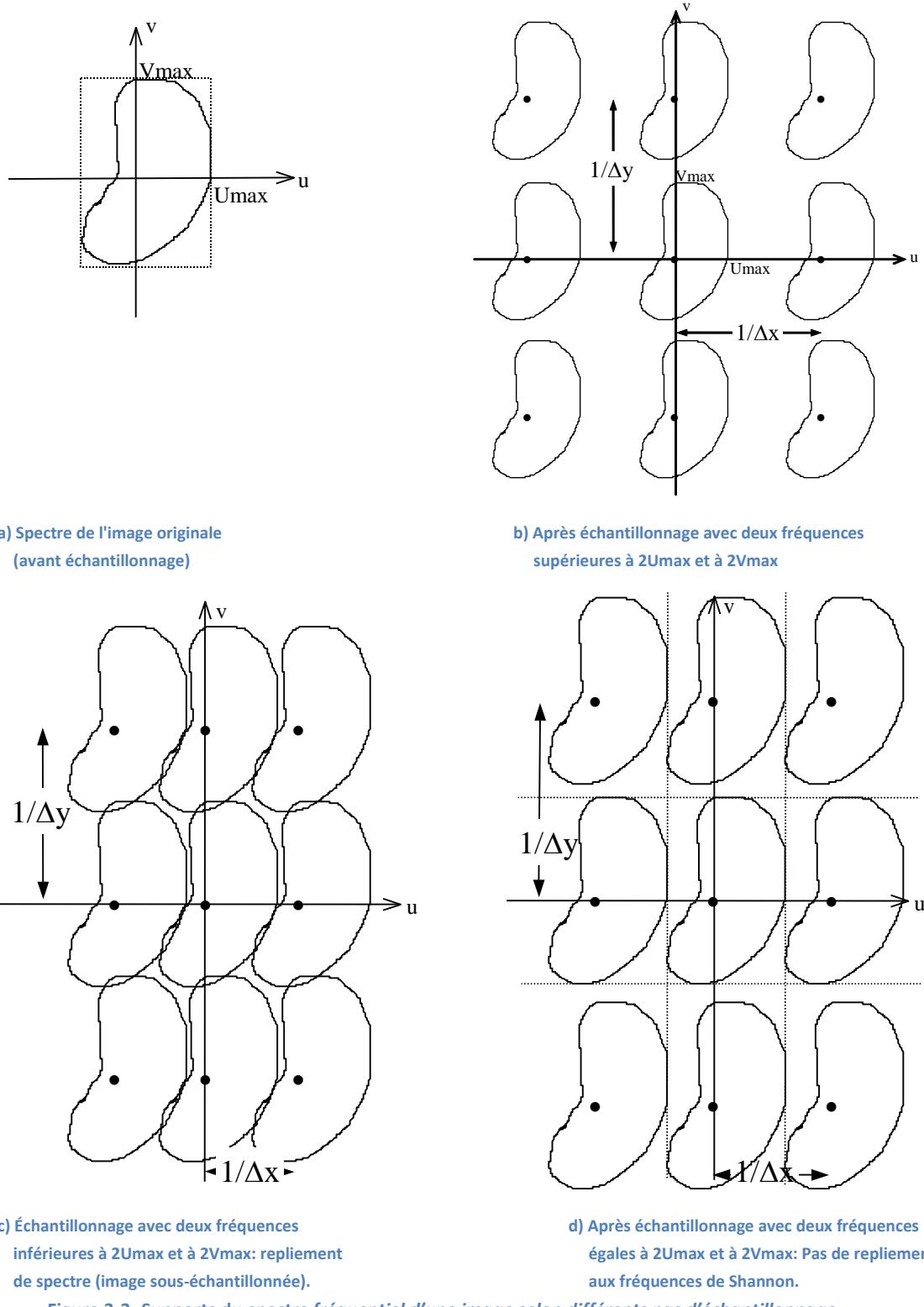


Figure 3-3- Supports du spectre fréquentiel d'une image selon différents pas d'échantillonnage

#### EXEMPLE:

Pour fixer les idées, considérons une ligne d'image dont l'amplitude  $L(x)$  varie sinusoïdalement autour de la valeur moyenne  $L_0$  (image de la Figure 3-4, par exemple).

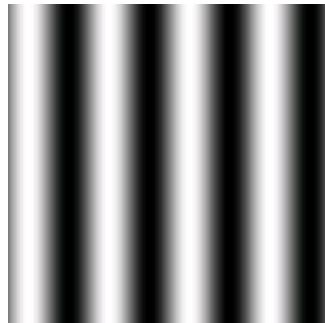
Cette ligne est entièrement caractérisée par deux points de l'espace transformé qui sont:

- l'origine (la composante continue) et

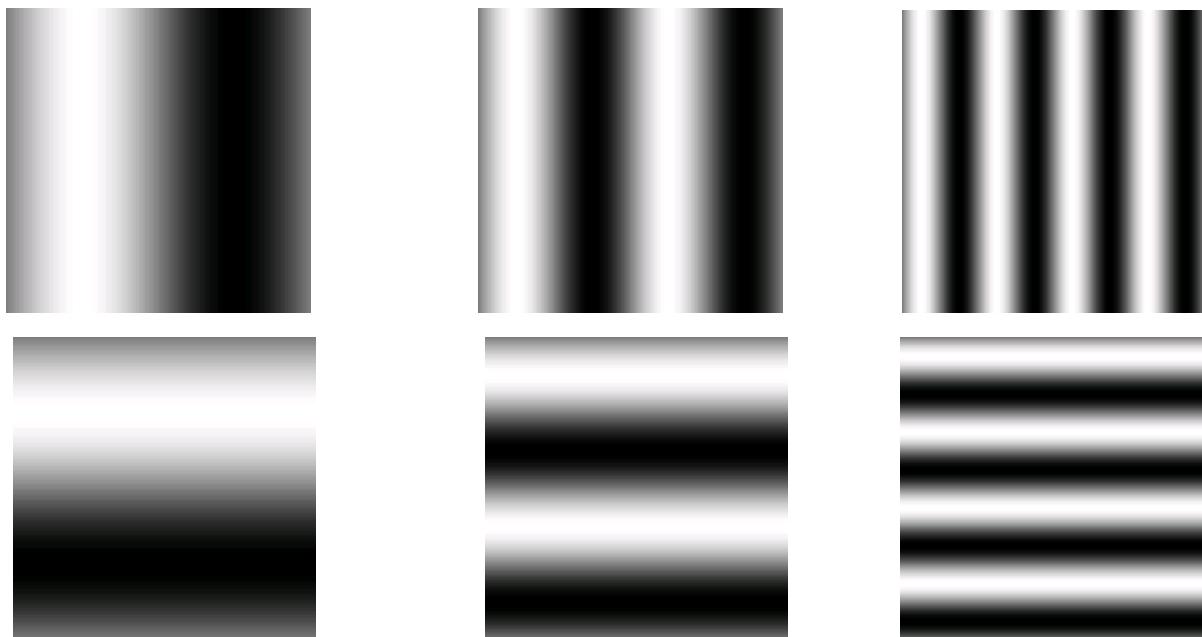
- le point dont l'abscisse correspond au nombre de cycles par unité de longueur.

L'information portée par le signal constitué par les N pixels de la ligne peut être représentée par les valeurs de ses deux composantes fréquentielles, ce qui consiste une compression.

La figure 3-6 présente plusieurs images dont le niveau de gris varie sinusoïdalement selon l'une des deux directions horizontale ou verticale et avec différentes fréquences spatiales.



**Figure 3-4- Image monochrome dont l'amplitude varie sinusoïdalement dans la direction horizontale**



**Figure 3-5- Images dont l'amplitude varie sinusoïdalement selon l'une des deux directions horizontale ou verticale**

Pour qu'une transformation soit efficacement utilisée, il faut qu'elle soit réversible. Parmi les transformations réversibles, on s'intéresse surtout aux transformations orthogonales telles que la TF et la TC.

### La Transformation de Fourier Discrète Bidimensionnelle:

On définit la TF2D (Transformée de Fourier Discrète bidimensionnelle) d'une image L de taille MxN (ou plus généralement d'un signal bidimensionnel), par extension de la TFD de (5):

$$\text{Éq 3-1} \quad F2_{kl}^{MN}(L) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} L_{mn} e^{-j2\pi(\frac{km}{M} + \frac{nl}{N})} \quad k = 0, \dots, M-1 \text{ et } l = 0, \dots, N-1$$

Les entiers  $k$  et  $l$  sont les variables des fréquences spatiales verticales et horizontales respectivement. Ces sont les variables réciproques des "variables d'espace"  $m$  et  $n$ .

La transformation inverse TF2D<sup>-1</sup> est alors donnée par:

$$\text{Éq 3-2} \quad IF2_{mn}^{MN}(L) = L_{mn} = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} F2_{kl}^{MN}(L) e^{j2\pi(\frac{km}{M} + \frac{nl}{N})} \quad m=0, \dots, M-1 \text{ et } n=0, \dots, N-1$$

La TF2D est une décomposition de l'image dans une base d'exponentielles complexes (donc de sinus/cosinus). Elle satisfait la propriété de la périodicité suivante:

$$\text{Éq 3-3} \quad F2_{(k+pM)(l+qN)}^{MN}(L) = F2_{kl}^{MN}(L) \quad \text{pour tout couple d'entiers (p,q).}$$

En appliquant la  $\text{TF2D}^{-1}$ , on aura de la même façon:

$$\text{Éq 3-4} \quad L_{(m+pM)(n+qN)} = L_{mn}$$

L'image  $L$  et sa TF2D sont périodiques et de période  $N$  pour les lignes et  $M$  pour les colonnes.

La TF2D d'une image quelconque (réelle ou complexe) est généralement complexe. Lorsque  $L$  est réelle, sa TF2D est une fonction complexe dont la partie réelle est paire et la partie imaginaire est impaire.

La TF2D est une transformation séparable puisque son calcul peut être décomposé en une séquence des TFD -1D sur les lignes suivies d'une séquence des TFD -1D sur les colonnes. En effet, la relation Éq 3-1 peut se mettre sous la forme:

$$\text{Éq 3-5} \quad F2_{kl}^{MN}(L) = \frac{1}{M} \sum_{m=0}^{M-1} \left[ \frac{1}{N} \sum_{n=0}^{N-1} L_{mn} e^{\frac{-j2\pi nl}{N}} \right] e^{\frac{-j2\pi km}{M}}$$

or,  $\left[ \frac{1}{N} \sum_{n=0}^{N-1} L_{mn} e^{\frac{-j2\pi nl}{N}} \right] = F_I^N(L_m) = \text{TFD d'ordre } N \text{ de la } m^{\text{ième}} \text{ ligne de l'image } L$

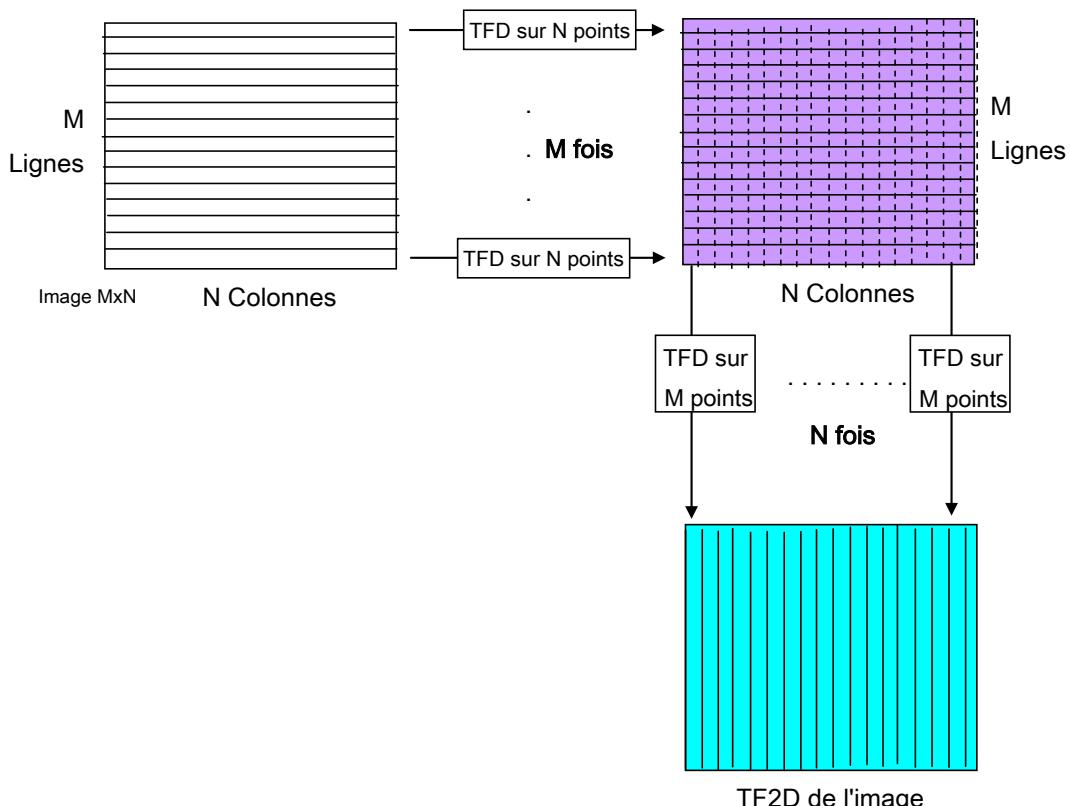


figure 3-6 - Implantation de la TF2D en utilisant la TFD

$$\text{Eq 3-6} \quad F_{kl}^{MN}(L) = \frac{1}{M} \sum_{m=0}^{M-1} F_l^N(L_m) e^{-j\frac{2\pi km}{M}} = \text{TFD d'ordre } M \text{ de la } l^{\text{ième}} \text{ colonne de } L$$

l'image formée des TFD des lignes de L

Donc, la TF2D d'une image peut être obtenue en remplaçant toutes les lignes de l'image par leurs TFD (d'ordre N) monodimensionnelles respectives, puis en calculant les TFD (d'ordre M) des colonnes de l'image ainsi formée. Ainsi le calcul de la TF2D d'ordre MxN revient à calculer M+N TFD monodimensionnelles. Ceci est illustré par la figure 3-6.

### La Transformation de Cosinus Discrète 1D:

La définition de la TCD, a été donnée pour la première fois par Nasser Ahmed en 1974. Depuis, plusieurs formes ont été proposées. Les équations suivantes, présentées sous la forme la plus utilisée dans la littérature traitant de ce sujet, donnent la transformée en cosinus discrète directe et inverse dans le cas mono-dimensionnel.

#### Transformée Directe :

Pour un signal de N échantillons:  $s_n$  [n=0,...,N-1], la TCD sur N points est définie par:

$$\text{Eq 3-7} \quad C_k^N(s) = c(k) \sum_{n=0}^{N-1} s_n \cos\left[\frac{(2n+1)k\pi}{2N}\right] \quad \text{Pour } k=0, \dots, N-1$$

#### Transformée Inverse:

La TCD<sup>-1</sup>, donnant le signal s à partir de sa TCD, est définie par :

$$\text{Eq 3-8} \quad s_n = \sum_{k=0}^{N-1} c(k).C_k^N(s) \cos\left[\frac{(2n+1)k\pi}{2N}\right] \quad \text{Pour } n=0, \dots, N-1$$

#### Propriétés de la TCD:

Parmi les propriétés les plus importantes de la TCD, on cite:

- 1- Les coefficients de la TCD sont réels.
- 2- La valeur à l'origine est directement proportionnelle à la valeur moyenne du signal s:

$$\text{Eq 3-9} \quad C_0^N(s) = \sqrt{\frac{1}{N}} \sum_{n=0}^{N-1} s_n = \sqrt{N} \cdot \mu \quad \text{où } \mu \text{ est la valeur moyenne de } s.$$

- 3- La TCD d'un signal discret de N échantillons est périodique et de période 4N:  $C_{4N+k}^N(s) = C_k^N(s)$
- 4- La TCD est une transformation linéaire: si  $s_n = A.s1_n + B.s2_n \Rightarrow C_k^N(s) = A.C_k^N(s1) + B.C_k^N(s2)$
- 5- Conservation de l'énergie: L'énergie des échantillons d'un signal est exactement celle portée par les composantes TCD de ce signal.
- 6- Les fonctions de base de la TCD forment une base orthonormée dont les vecteurs de base  $v(k)$  [ $k=0, \dots, N-1$ ] sont donnés par:

$$\text{Eq 3-10} \quad v(k) = \begin{bmatrix} c(k) \cos \frac{k\pi}{2N} & c(k) \cos \frac{3k\pi}{2N} & c(k) \cos \frac{5k\pi}{2N} & \dots & c(k) \cos \frac{(2N-1)k\pi}{2N} \end{bmatrix}$$

On montre que le produit scalaire  $\langle v(k), v(l) \rangle = \begin{cases} 0 & \text{si } k \neq l \\ \text{cte} & \text{si } k = l \end{cases}$

### Relation entre TCD et TFD

L'élaboration d'algorithmes efficaces pour le calcul de TCD a été initialement basée sur la transformée de Fourier rapide (FFT). Après l'algorithme de Haralick basé sur l'utilisation d'une FFT sur  $2N$  points pour calculer une TCD sur  $N$  points, Narashima et Peterson ont proposé un algorithme plus efficace utilisant une FFT sur  $N$  points pour calculer une TCD sur  $N$  points.

Cet algorithme qui a l'avantage d'être facile à mettre en œuvre consiste en un réarrangement de la séquence d'entrée  $x_n$  [ $n = 0, \dots, N-1$ ] en regroupant tous les échantillons d'indices pairs pris dans l'ordre croissant des indices, suivis par les échantillons d'indices impairs pris dans l'ordre décroissant des indices. Ce réarrangement produit une nouvelle séquence  $y_n$  donnée par:

$$\begin{aligned} \text{Eq 3-11} \quad & \begin{cases} y_n = x_{2n} & n = 0, 1, \dots, \frac{N}{2}-1 \\ y_{N-1-n} = x_{2n+1} & n = 0, 1, \dots, \frac{N}{2}-1 \end{cases} \end{aligned}$$

La TCD de  $x_n$ ,  $C_k^N(x)$  s'écrit alors:

Éq 3-12

$$\begin{aligned} C_k^N(x) &= 2c(k) \left[ \sum_{i=0}^{\frac{N}{2}-1} x_{2i} \cos \frac{\pi(4i+1)k}{2N} + \sum_{i=0}^{\frac{N}{2}-1} x_{2i+1} \cos \frac{\pi(4i+3)k}{2N} \right] = 2c(k) \left[ \sum_{i=0}^{\frac{N}{2}-1} y_i \cos \frac{\pi(4i+1)k}{2N} + \sum_{i=\frac{N}{2}}^{N-1} y_i \cos \frac{\pi(4i+1)k}{2N} \right] \\ &= 2c(k) \operatorname{Re} \left\{ \sum_{i=0}^{N-1} y_i e^{-j \frac{2i\pi k}{N}} e^{-j \frac{\pi k}{2N}} \right\} = 2c(k) \operatorname{Re} \left\{ e^{-j \frac{\pi k}{2N}} F_k^N(y) \right\} \end{aligned}$$

où  $F_k^N(y)$  [ $k = 0, \dots, N-1$ ] est la TFD sur  $N$  points de  $y_n$ .

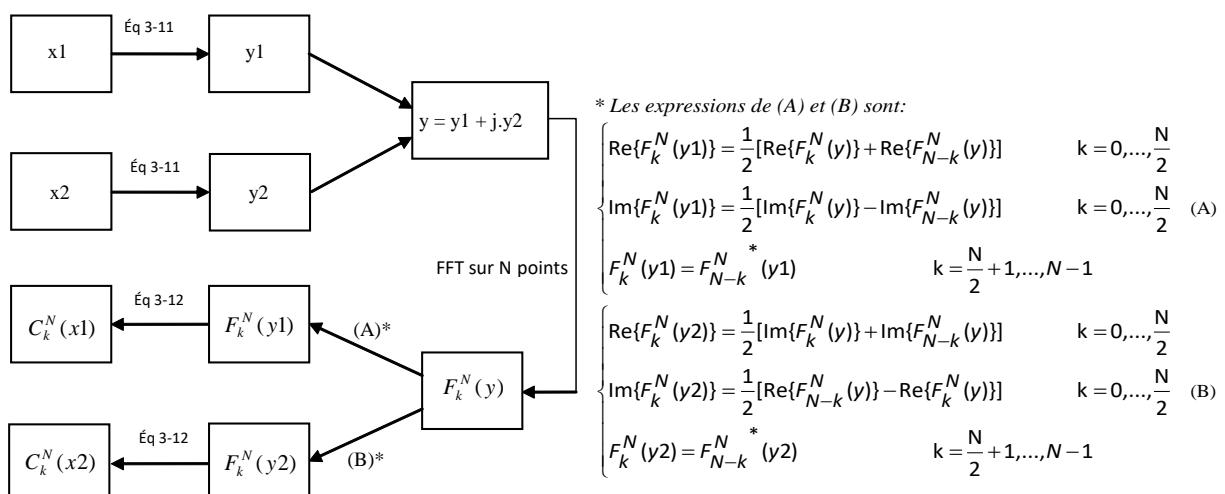


Figure 3-7. Schéma-bloc illustrant le calcul des TCD de 2 signaux réels en utilisant une seule FFT sur  $N$  points.

Étant donnée que  $y_n$  est une séquence réelle, la FFT sur  $N$  points peut servir à obtenir, simultanément, la TFD de deux séquences similaires (voir Éq 2-17 et Éq 2-18). Ainsi, le coût de calcul sera réduit de près de la moitié et, par conséquent, nous pouvons calculer deux TCD sur  $N$  points à

l'aide d'une FFT de N points, ce qui représente une réduction significative de la complexité de celle du calcul de la TCD.

La Figure 3-7 illustre le schéma de calcul de la TCD de deux séquences réelles.

Notons que le calcul de la TCD à partir de la FFT nécessite, en utilisant l'Éq 3-12,  $2(N-2)$  multiplications réelles supplémentaires.

### Représentation d'image par TC2D:

Pour une image L de M lignes et N colonnes ayant  $M \times N$  pixels d'amplitude  $L_{mn}$ , la TC2D sur  $M \times N$  points est donnée par :

$$\text{Éq 3-13} \quad C_{kl}^{MN}(L) = c(k).c(l) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} L_{mn} \cos\left[\frac{(2m+1)k\pi}{2M}\right] \cdot \cos\left[\frac{(2n+1)l\pi}{2N}\right]$$

*Pour  $k=0, \dots, M-1$  et  $l=0, \dots, N-1$*

avec :  $c(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{pour } k=0 \\ \sqrt{\frac{2}{N}} & \text{pour } k=1, \dots, N-1 \end{cases}$

La TC2D<sup>-1</sup>, donnant l'image L à partir de sa TC2D, est définie par :

$$\text{Éq 3-14} \quad L_{mn} = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} c(k).c(l).C_{kl}^{MN} \cos\left[\frac{(2m+1)k\pi}{2M}\right] \cdot \cos\left[\frac{(2n+1)l\pi}{2N}\right]$$

*Pour  $m=0, \dots, M-1$  et  $n=0, \dots, N-1$*

Comme la TF2D, la TC2D est séparable dans le sens où on peut la calculer en utilisant la TCD mono-dimensionnelle en l'appliquant sur les lignes suivie de son application sur les colonnes résultantes. En effet, la relation Éq 3-13 peut se mettre sous la forme:

$$\begin{aligned} \text{Éq 3-15} \quad C_{kl}^{MN}(L) &= c(k) \cdot \sum_{m=0}^{M-1} \left( c(l) \sum_{n=0}^{N-1} L_{mn} \cos\left[\frac{(2n+1)l\pi}{2N}\right] \right) \cdot \cos\left[\frac{(2m+1)k\pi}{2M}\right] \\ &= c(k) \cdot \sum_{m=0}^{M-1} C_k^M(L_m) \cdot \cos\left[\frac{(2m+1)k\pi}{2M}\right] \quad \text{Pour } k=0, \dots, M-1 \text{ et } l=0, \dots, N-1 \end{aligned}$$

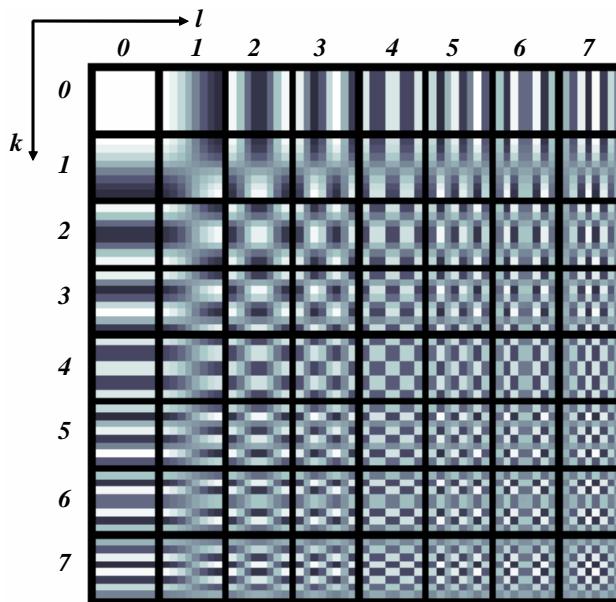
L'illustration de la figure 3-6 reste valable pour le calcul de la TC2D en remplaçant partout TFD par TCD.

#### *La TCD : Meilleure transformation sous optimale*

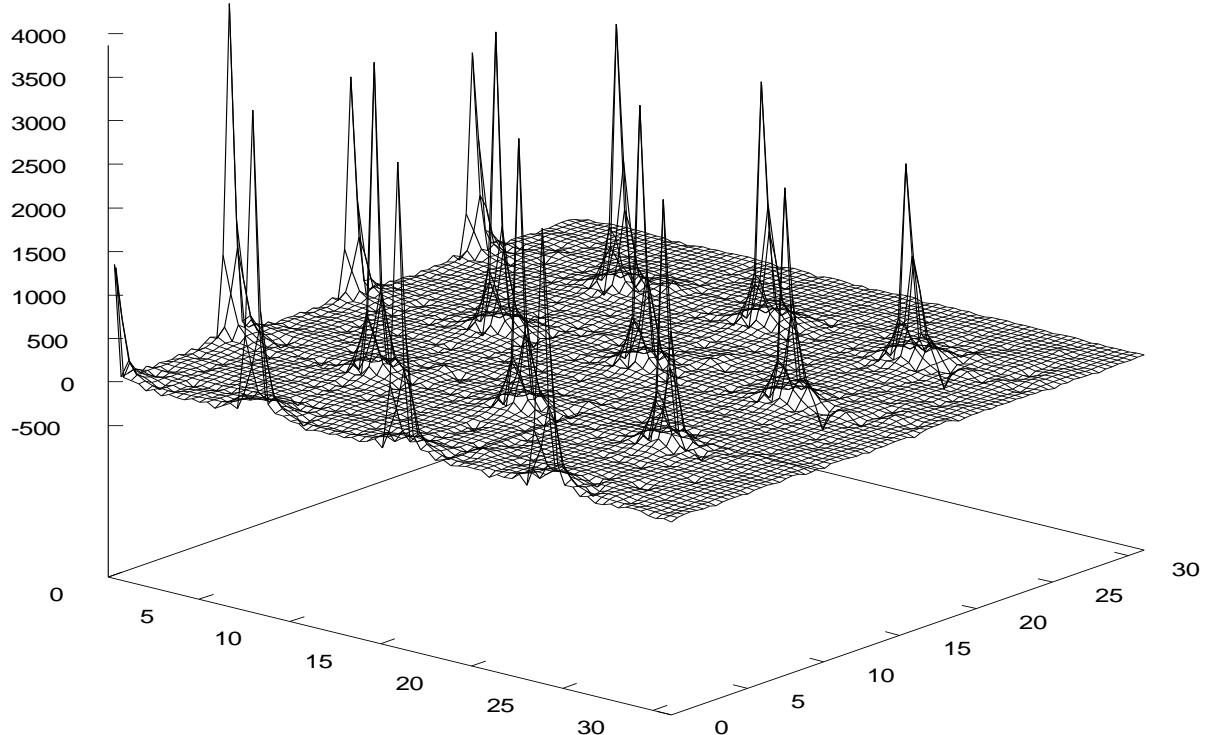
Le standard JPEG a adopté la TCD pour la plupart de ses modes de codage. Ce choix repose sur les critères suivants:

- La TCD assure une forte dé-corrélation avec une très bonne concentration d'énergie
- Elle possède des fonctions de base visuellement plaisantes (voir Figure 3-8, pour  $M = N = 8$ ).
- Elle offre une complexité réduite grâce à l'existence d'algorithmes de calcul rapides.

La Figure 3-9 montre la concentration d'énergie sur les coefficients basses fréquences réalisée par une TCD-8x8 sur une image de 32x32 pixels (comportant 16 blocs de 8x8 pixels chacun).



**Figure 3-8. Fonctions de base de la TCD-8x8 :**  
les paramètres  $k$  et  $l$  correspondent aux fréquences spatiales dans les directions verticales et horizontales respectivement



**Figure 3-9. TC2D-8x8 d'une image de 32x32 pixels**

$$X = \begin{bmatrix} 140 & 132 & 132 & 155 & 176 & 168 & 168 & 157 \\ 127 & 132 & 153 & 160 & 155 & 161 & 161 & 168 \\ 132 & 155 & 160 & 160 & 153 & 157 & 160 & 155 \\ 145 & 157 & 153 & 168 & 168 & 161 & 155 & 160 \\ 153 & 151 & 158 & 155 & 172 & 160 & 157 & 153 \\ 157 & 147 & 143 & 140 & 160 & 163 & 163 & 176 \\ 157 & 157 & 160 & 160 & 153 & 155 & 176 & 176 \\ 147 & 147 & 160 & 153 & 143 & 155 & 162 & 180 \end{bmatrix} \quad X_{DCT} = \begin{bmatrix} 3747.4 & -162.5 & -27.4 & -20.5 & 13.1 & -3.8 & -0.4 & 7.4 \\ -48.0 & -37.4 & -99.4 & 33.0 & 3.7 & -1.9 & -18.2 & -6.7 \\ -18.9 & -66.0 & 30. & -12.6 & 5.7 & 30.6 & 9.7 & -3.2 \\ 6.1 & -27.3 & 15.9 & 86.0 & 35.4 & -10.6 & -4.8 & 30.6 \\ 3.4 & 16.4 & -33.8 & 33.7 & 23.6 & -10.5 & -4.3 & 4.5 \\ 36.2 & 13.2 & 1.9 & 15.7 & 4.4 & 3.1 & -37.4 & -10.0 \\ -25.0 & -5.1 & 19.5 & 25.1 & -6.8 & -19.8 & 2.4 & -0.2 \\ 9.1 & 31.6 & -17.7 & -1.9 & -24.2 & 13.2 & -15.5 & 17.3 \end{bmatrix}$$

**EXEMPLE:**

Soit la matrice  $X$  ci-dessous qui représente les valeurs d'éléments d'un bloc d'image de 8x8 pixels. Les valeurs des coefficients du bloc transformé par TCD sont données par  $X_{DCT}$ . On remarque la forte amplitude du coefficient relatif à la composante continue (pour  $k = 0$  et  $l = 0$ ) et des coefficients qui les entourent et qui sont de basses fréquences par rapport aux amplitudes des coefficients de moyennes et hautes fréquences.

# Chapitre 4 – Signaux et Systèmes à temps discret

Un signal à temps discret  $x[n]$ ,  $n=n_0, \dots, n_0+N-1$ , peut s'écrire mathématiquement comme suit :

Éq 4-1

$$x[n] = \sum_{l=n_0}^{n_0+N-1} x[l] \cdot \delta[n-l]$$

où  $\delta[n] = \begin{cases} 1 & \text{pour } n = 0 \\ 0 & \text{pour } n \neq 0 \end{cases}$  est l'impulsion unité qui est l'équivalente de l'impulsion de Dirac pour les signaux continus, avec une différence importante: elle est définie d'une manière simple et précise par son amplitude qui est égale à 1 (et non pas par une intégrale égale à 1).

Ainsi, le signal échelon  $u[n]$  défini par :  $u[n] = \begin{cases} 1 & \text{pour } n \geq 0 \\ 0 & \text{pour } n < 0 \end{cases}$  peut, selon l'Éq 4-1, s'écrire:

Éq 4-2

$$u[n] = \sum_{l=0}^{+\infty} \delta[n-l]$$

Il peut aussi s'exprimer par (relation équivalente à  $\Gamma(t) = \int_{-\infty}^t \delta(u) du$ ) :  $u[n] = \sum_{l=-\infty}^n \delta[l]$

## 4.1 Systèmes à temps discret :

Un système à temps discret est défini mathématiquement comme une transformation ou un opérateur  $T\{\bullet\}$  qui s'applique sur le signal discret de l'entrée  $x[n]$  pour produire en sortie un signal discret  $y[n]$ . Ceci se traduit par:

Éq 4-3

$$y[n] = T\{x[n]\}$$

Schématiquement, cette opération est représentée par la Figure 4-1.

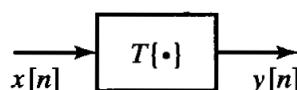


Figure 4-1. Représentation d'un système à temps discret

L'Éq 4-3 est une règle ou formule qui permet de calculer les  $y[n]$  en fonction des  $x[n]$  pour toute valeur de  $n$ . Il faut noter que la valeur de  $y$  pour tout  $n$  peut dépendre de toutes ou d'une partie des valeurs de la séquence  $x$ .

Dans ce qui suit, nous donnons quelques exemples simples et utiles de tels systèmes.

### Exemple 1: Système à retard idéal :

Ce système est défini par :  $y[n] = x[n-n_d]$ , où le retard du système  $n_d$  est un entier positif de valeur fixe pour le système.  $y[n]$  est une version retardée de  $x[n]$  de  $n_d$  positions (on décale  $x[n]$  de  $n_d$  échantillons vers la droite).

Dans le cas où  $n_d$  est négatif, le système est dit "à avance".  $x[n]$  est décalée vers la gauche.

### **Exemple 2: Système Moyenneur à fenêtre glissante:**

Dans le cas général, la sortie  $y[n]$  de ce système est la moyenne des  $M_1+M_2+1$  valeurs de la séquence  $x$  prises autour de  $x[n]$  :

$$y[n] = \frac{1}{M_1 + M_2 + 1} \sum_{k=-M_1}^{+M_2} x[n - k]$$

### **Système sans mémoire**

Le système est dit sans mémoire lorsque  $y[n]$  ne dépend que de la valeur  $x[n]$ .

**Exemple :**  $y[n] = (x[n])^2$ .

### **Système linéaire**

Les systèmes linéaires sont définis par le principe de superposition :

si  $y_1[n]$  et  $y_2[n]$  sont les réponses du système aux entrées  $x_1[n]$  et  $x_2[n]$ , alors le système est linéaire si et seulement si :

$$T\{x_1[n] + x_2[n]\} = T\{x_1[n]\} + T\{x_2[n]\} = y_1[n] + y_2[n] \quad \text{et} \quad T\{a.x_1[n]\} = a.T\{x_1[n]\}$$

où  $a$  est une constante quelconque.

Ceci peut être généralisé au cas où  $x[n]$  est une combinaison linéaire de plusieurs signaux  $x_k[n]$ . Dans ce cas,  $y[n]$  est donné par la même combinaison linéaire de  $y_k[n]$  où  $y_k[n]$  est la réponse du système à  $x_k[n]$  :

Éq 4-4

$$y[n] = T\{x[n]\} = T\left\{\sum_k a_k \cdot x_k[n]\right\} = \sum_k a_k \cdot T\{x_k[n]\} = \sum_k a_k \cdot y_k[n]$$

Basé sur ce principe, on peut montrer que les systèmes des exemples 1 et 2 donnés ci-dessus sont linéaires alors que celui de l'exemple du système sans mémoire n'est pas linéaire.

De même, le système donné par :  $w[n] = \log_{10}(|x[n]|)$  n'est pas linéaire.

### **Système invariant dans le temps :**

Un système est dit invariant dans le temps si tout décalage dans le temps, effectué sur le signal d'entrée, se traduit par le même décalage sur le signal de sortie :

Ayant  $y[n] = T\{x[n]\}$ , si pour tout  $n_0$ , on aura:  $T\{x[n - n_0]\} = y[n - n_0]$ , alors le système représenté par  $T\{\bullet\}$  est invariant dans le temps.

**Exemple:** Considérons le système dit "accumulateur" défini par :

$$y[n] = \sum_{k=-\infty}^n x[k]$$

Soit  $x_1[n] = x[n - n_0]$  et  $y_1[n]$  la réponse de l'accumulateur à  $x_1[n]$ . Pour que le système soit invariant dans le temps, il faut montrer que  $y_1[n] = y[n - n_0]$ .

En effet,

Éq 4-5

$$y_1[n] = \sum_{k=-\infty}^n x_1[k] = \sum_{k=-\infty}^n x[k - n_0] = \sum_{l=-\infty}^{n-n_0} x[l] = y[n - n_0]$$

### Système Causal:

Un système est dit causal lorsque, pour tout  $n_0$ ,  $y[n_0]$  ne dépend que des  $x[n]$  tels que  $n \leq n_0$ .

Exemples :

- Le système à retard est causal ( $n_d \geq 0$ ). Le système à avance est non causal ( $n_d < 0$ ).
- Le moyenneur à fenêtre glissante est causal lorsque  $M1 \leq 0$  et  $M2 \geq 0$ , il est non causal dans les autres cas.

### Système Stable:

Un système est dit stable, si pour toute entrée  $x[n]$  à amplitude finie ( $\exists B_x$  finie  $\geq 0 / |x[n]| < B_x \forall n$ ), la réponse  $y[n]$  est elle aussi à amplitude finie ( $\exists B_y$  finie  $\geq 0 / |y[n]| < B_y \forall n$ ).

Le système est instable même si il vérifie cette propriété pour certaines séquences d'entrée mais pas pour d'autres.

Exemples :

- Le système défini par :  $y[n] = K.x[n] + A$ , où  $K$  et  $A$  sont des constantes réelles, est non linéaire, invariant, et stable.
- L'élément « délai » défini par :  $y[n] = x[n-1]$  est linéaire, invariant, et stable.
- Le système défini par :  $y[n] = n.x[n]$  est linéaire, non invariant, et instable.
- Le système défini par :  $y[n] = |x[n]|^2$  est non linéaire, invariant, et stable.

## 4.2 Systèmes récursifs – Systèmes non récursifs

D'une manière générale, les systèmes numériques linéaires et invariants (SLI) peuvent être décrits par des équations aux différences finies, linéaires et à coefficients constants, de la forme :

Éq 4-6

$$\begin{aligned} y[n] + a_1.y[n-1] + a_2.y[n-2] + \dots + a_N.y[n-N] \\ = b_0.x[n] + b_1.x[n-1] + \dots + b_M.x[n-M] \end{aligned}$$

Une telle équation exprime le fait que la valeur de la sortie  $y[n]$  à l'instant courant est une combinaison linéaire des  $N$  sorties précédentes, de l'entrée courante, et des  $M$  entrées précédentes. On la note souvent de façon plus compacte:

Éq 4-7

$$y[n] = \sum_{i=0}^M b_i.x[n-i] - \sum_{i=1}^N a_i.y[n-i]$$

De tels systèmes sont dits récursifs étant donné le caractère récursif de l'équation correspondante: il faut avoir déjà calculé toutes les sorties précédentes pour pouvoir calculer la sortie courante.

Un cas particulier est celui des systèmes dont les valeurs de sortie ne dépendent que de l'entrée et de son passé :

Éq 4-8

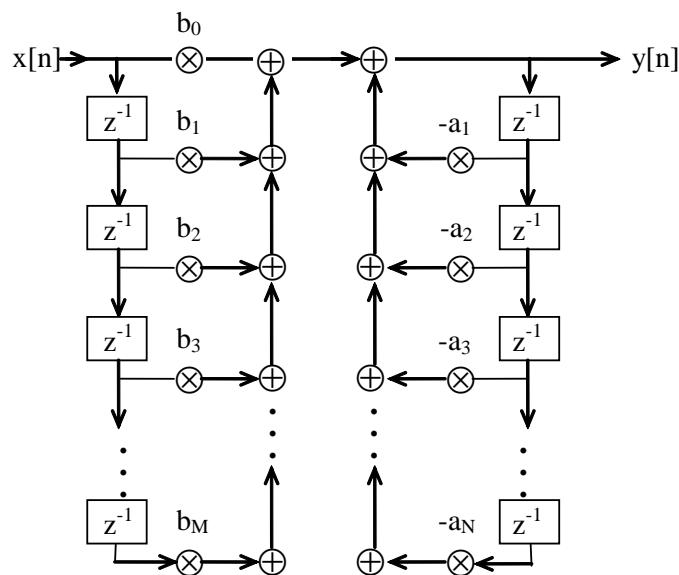
$$y[n] = \sum_{i=0}^M b_i \cdot x[n - i]$$

De tels systèmes sont dits *non récursifs*.

L'*ordre* d'un SLI numérique est donné par le degré de la récursivité de l'équation aux différences finies associée :  $N$ .

De la même façon qu'une équation différentielle entre l'entrée et la sortie (toutes deux analogiques) d'un système à temps continu définit un *filtre linéaire analogique*, on appelle *filtre numérique* un système numérique linéaire et invariant. On parlera donc dans la suite de *filtre récursif* et de *filtre non récursif*.

Il est possible de visualiser l'équation de récurrence associée à un filtre numérique, sous la forme d'une *structure* telle que celle de la Figure 4-2. L'élément "délai", symbolisé par  $z^{-1}$ , produit une sortie retardée d'un instant par rapport à son entrée.



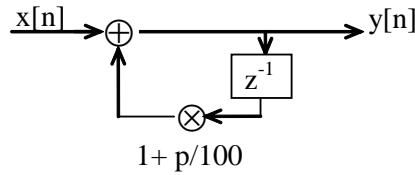


Figure 4-3. Structure du SLI correspondant au calcul des intérêts composés

### Réponse impulsionnelle

La réponse impulsionnelle, notée  $h(n)$ , d'un SLI numérique est alors définie comme sa réponse forcée  $y(n)$  à l'impulsion unité, en supposant que les conditions initiales sont nulles :  $y(n)=0$  pour  $n<0$ .

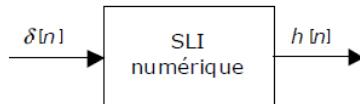


Figure 4-4. Réponse impulsionnelle

Au contraire des systèmes analogiques, le calcul de la réponse impulsionnelle d'un SLI numérique est immédiat : il suffit d'effectuer la récurrence numérique.

**Exemple :** La réponse impulsionnelle du système des intérêts composés est:

$$h(n) = \{1, 1+p/100, (1+p/100)^2, (1+p/100)^3, \dots\}$$

Elle est obtenue en utilisant la relation :  $h[n] = h[n - 1] \cdot (1 + \frac{p}{100}) + \delta[n]$  avec  $h[n] = 0$  pour  $n < 0$ .

On constate que la réponse impulsionnelle d'un filtre récursif est une séquence illimitée de valeurs non identiquement nulles (même si le filtre est stable), puisque chaque valeur de sortie dépend des valeurs de sorties précédentes. Par contre, la réponse impulsionnelle d'un filtre non récursif est donnée par la suite de ses coefficients  $b_i$ , et constitue donc une séquence qui s'annule après un nombre fini  $M+1$  de valeurs. Il s'ensuit que les filtres récursifs sont appelés *filtres à réponse impulsionnelle infinie (RII)* et que les filtres non récursifs sont dits *à réponse impulsionnelle finie (RIF)*.

### Réponse forcée à une entrée quelconque – Convolution numérique

Le calcul de la réponse à une entrée quelconque est tout aussi trivial que celui de la réponse impulsionnelle: il suffit d'effectuer la récurrence numérique. Il est cependant intéressant de constater que la réponse à une entrée quelconque peut être déterminée facilement si on connaît déjà la réponse impulsionnelle. On peut en effet toujours considérer qu'une séquence d'entrée  $\{x[n]\}=\{x[0], x[1], x[2], \dots\}$  est une somme d'impulsions numériques pondérées et décalées :

$$\text{Eq 4-9} \quad x[n] = x[0].\delta[n] + x[1].\delta[n-1] + x[2].\delta[n-2] + x[3].\delta[n-3] + \dots \quad (\text{pour tout } n)$$

Vu les propriétés de linéarité et d'invariance du système, la réponse à  $x[n]$  est donnée par :

$$\text{Eq 4-10} \quad y[n] = x[0].h[n] + x[1].h[n-1] + x[2].h[n-2] + x[3].h[n-3] + \dots \quad (\text{pour tout } n)$$

ce qui n'est rien d'autre qu'une combinaison linéaire des réponses impulsionnelles à chacun des nombres de la séquence d'entrée (et donc décalées d'autant).

Cette équation définit ce que l'on appelle une *convolution numérique*, notée  $*$ , et donnée dans le cas le plus général par:

$$\text{Eq 4-11}$$

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

Notons que la sommation est définie pour  $l$  allant de  $-\infty$  à  $+\infty$ , afin de prendre en compte les systèmes non causaux, dont la réponse impulsionnelle  $h[n]$  peut être définie pour  $n$  négatif.

Notons aussi que le produit de convolution est bien commutatif :

Éq 4-12

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

Le produit de convolution numérique possède une interprétation simple: pour obtenir une valeur particulière  $y[n_0]$  de la séquence  $y[n]$ , il suffit d'inverser  $h[n]$ , de positionner  $h[0]$  sur  $x[n_0]$ , et de calculer le produit scalaire entre les séquences  $x[n]$  et  $h[n]$  ainsi obtenues. Ainsi,  $y[n_0]$  est donnée par une combinaison linéaire des valeurs de  $x[n]$  autour de  $x[n_0]$  dont les coefficients sont les valeurs de  $h[n]$ .

Il est évident que, dans le cas des réponses impulsionnelles infinies,  $y[n]$  ne peut pas être calculée en utilisant le produit de convolution. La relation de récurrence de l'Éq 4-7 reste un moyen pratique pour calculer  $y[n]$  pour les différentes valeurs souhaitées de  $n$ .

### 4.3 Transformée en Z

Les opérations effectuées pour réaliser la convolution entre deux séquences rappellent celles qui permettent de calculer un produit de deux polynômes. En effet, le produit de deux polynômes se réduit à une somme de produits des coefficients du premier par les coefficients du second, décalés des puissances correspondantes.

On montre que ces deux opérations sont exactement identiques: la convolution entre deux séquences numériques est exactement équivalente au produit de deux polynômes dont les coefficients sont précisément les séquences à convoluer, comme l'illustre l'exemple suivant.

**Exemple:** Soient  $\{x[n]\}=\{1,2,3,0,0,\dots\}$  et  $\{h[n]\}=\{2,3,4,0,0,\dots\}$ . La convolution de  $\{x[n]\}$  avec  $\{h[n]\}$  donne  $\{2,7,16,17,12,0,0,\dots\}$ .

Construisons deux polynômes  $X(z)$  et  $H(z)$  à partir de  $\{x[n]\}$  et  $\{h[n]\}$ :

$$X(z)=1+2z^{-1}+3z^{-2}=z^{-2}(z^2+2z+3) \quad H(z)=2+3z^{-1}+4z^{-2}=z^{-2}(2z^2+3z+4)$$

Le produit de ces deux polynômes donne bien le polynôme  $Y(z)$

$$Y(z)=z^{-4}(2z^4+7z^3+16z^2+17z+12)=2+7z^{-1}+16z^{-2}+17z^{-3}+12z^{-4}$$

L'exemple précédent montre qu'il est intéressant d'associer à une séquence numérique  $x[n]$  un polynôme en  $z^{-1}$  (où  $z$  est une variable complexe), que l'on appelle *transformée en Z* de  $\{x[n]\}$  et que l'on note  $X(z)$ :

Éq 4-13

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

Si  $H(z)$  est la transformée en  $z$  de la réponse impulsionnelle  $h[n]$  d'un SLI (Système Linéaire Invariant) numérique :

Éq 4-14

$$h[n] \xrightarrow{Z} H(z) = \sum_{n=-\infty}^{+\infty} h[n] \cdot z^{-n}$$

La transformée en  $Z$  de la sortie du système est alors donnée par :

Éq 4-15  $y[n] \xrightarrow{Z} Y(z) = X(z) \cdot H(z)$

On voit donc que la transformée en  $Z$  (qui associe une fonction de la variable complexe  $z$  à un signal numérique) est l'homologue, dans le domaine numérique, de la transformée de Laplace (qui associe une fonction de la variable complexe  $p$  à un signal analogique):

Éq 4-16

$$x(t) \xrightarrow{L} X(p) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-pt} dt$$

Le passage de l'analogique au numérique correspond au passage de l'intégrale à la somme, et la substitution de  $z$  à  $e^p$  et à remplacer la variable continue  $t$  par la variable discrète  $n$ .

Comme la transformée de Laplace, la transformée en  $Z$  ne converge pas nécessairement pour toutes les valeurs de  $z$ . Elle jouit également de nombreuses propriétés, dont on donnera quelques-unes dans la section suivante. De même, on peut dresser une liste de  $X(z)$  pour certains signaux simples et usuels.

L'équation Éq 4-15 permet alors, si l'on connaît  $X(z)$  et  $H(z)$ , de calculer  $Y(z)$ , et d'en déduire  $y[n]$ . On peut d'ailleurs également utiliser pour ce faire la décomposition en fractions simples.

La transformée en  $Z$  est cependant largement moins indispensable en numérique que son homologue de Laplace ne l'est en analogique. La récurrence de l'Éq 4-7, qui se substitue à l'équation intégro-différentielle décrivant un SLI analogique, permet en effet le calcul direct de la sortie, sans devoir passer par le calcul de sa transformée en  $Z$ .

La seule propriété fondamentale de la transformée en  $Z$  que nous utiliserons ici est celle liée au retard de  $x[n]$  de  $n_0$  échantillons:

Éq 4-17

$$x[n - n_0] \xrightarrow{Z} X_d(z) = \sum_{n=-\infty}^{+\infty} x[n - n_0] \cdot z^{-n} = \sum_{k=-\infty}^{+\infty} x[k] \cdot z^{-(k+n_0)} = z^{-n_0} \sum_{k=-\infty}^{+\infty} x[k] \cdot z^{-k} = z^{-n_0} \cdot X(z)$$

En appliquant cette propriété à la récurrence de l'Éq 4-7:

Éq 4-18

$$Y(z) + Y(z) \sum_{i=1}^N a_i \cdot z^{-i} = X(z) \sum_{i=0}^M b_i \cdot z^{-i}$$

ce qui conduit à une autre formulation de  $H(z)$ :

Éq 4-19

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^M b_i \cdot z^{-i}}{1 + \sum_{i=1}^N a_i \cdot z^{-i}} = \frac{\sum_{i=0}^M b_i \cdot z^{-i}}{\sum_{i=0}^N a_i \cdot z^{-i}} \quad (\text{avec } a_0 = 1)$$

Cette relation définit la fonction de transfert en  $z$  d'un SLI qu'on peut comparer à l'équation donnant la forme de la fonction de transfert en  $p$  d'un système analogique. On définit les *zéros* et les *pôles* d'un SLI numérique comme les racines du numérateur et du dénominateur de  $H(z)$ .

Enfin, on peut établir un lien entre la position des pôles d'un système numérique et sa stabilité, et montrer qu'un SLI numérique est strictement stable si ses pôles sont tous à l'intérieur du cercle de rayon unité (cercle non compris), et stable si on accepte aussi les pôles simples sur le cercle de rayon unité. Le demi-plan de droite du plan complexe se trouve ainsi transformé en l'intérieur du cercle de rayon unité en substituant  $z$  à  $e^p$ .

En effet, le nombre complexe  $p=r+js$  correspond à la variable  $z = e^r \cdot e^{js}$  dont le module est :  $|z| = e^r$ . Le demi-plan de droite du plan complexe correspond à  $r < 0$ , qui rend  $|z| < 1$  (l'intérieur du disque de rayon unité).

L'axe imaginaire correspond à  $r=0$  qui donne  $|z|=1$ . D'où l'équivalence entre le demi-plan de droite et le disque unité illustrée dans la Figure 4-5.

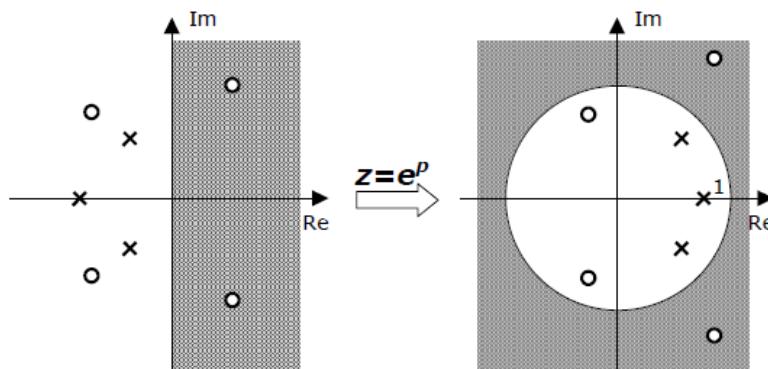


Figure 4-5. Zone de stabilité en  $z$  et lien avec la zone de stabilité en  $p$ .

### 4.3.1 Propriétés de la transformée en Z

#### Domaine de Convergence

La transformée en  $Z$  n'a pas de sens que si l'on précise le domaine des valeurs de  $z$  pour lesquelles cette série existe: sa région de convergence.

En effet, une série de la forme  $\sum_{n=0}^{+\infty} s_n$  converge, selon le critère de Cauchy sur la convergence des séries de puissance, si

$$\lim_{n \rightarrow \infty} |s_n|^{1/n} < 1$$

Pour appliquer ce critère, on décompose la série de  $X(z)$  en deux séries:

$$X(z) = \sum_{n=-\infty}^{+\infty} x[n] \cdot z^{-n} = \sum_{n=-\infty}^{-1} x[n] \cdot z^{-n} + \sum_{n=0}^{+\infty} x[n] \cdot z^{-n} = X_1(z) + X_2(z)$$

$X_2(z)$  converge si

$$\lim_{n \rightarrow \infty} |x[n] \cdot z^{-n}|^{1/n} = |z^{-1}| \lim_{n \rightarrow \infty} |x[n]|^{1/n} < 1 \Rightarrow |z| > \lim_{n \rightarrow \infty} |x[n]|^{1/n}$$

Alors  $X_2(z)$  converge si  $|z| > R_-$  avec

$$R_- = \lim_{n \rightarrow \infty} |x[n]|^{1/n}.$$

Pour  $X_1(z)$ , on fait un changement de variables  $m = -n$

$$X_1(z) = \sum_{n=-\infty}^{-1} x[n].z^{-n} = \sum_{n=1}^{+\infty} x[-m].z^m$$

$X_1(z)$  converge si

$$\lim_{m \rightarrow \infty} |x[-m].z^m|^{1/m} = |z| \lim_{m \rightarrow \infty} |x[-m]|^{1/m} < 1 \Rightarrow |z| < R_+ = \frac{1}{\lim_{m \rightarrow \infty} |x[-m]|^{1/m}}$$

Ainsi le domaine de convergence de  $X(z)$  est en général dans un anneau du plan complexe des  $z$  constitué de l'intersection de ces deux domaines de convergence.

Eq 4-20

$$0 \leq R_- < |z| < R_+ \leq \infty, \text{ avec}$$

$$R_- = \lim_{n \rightarrow \infty} |x[n]|^{1/n} \text{ et } R_+ = \frac{1}{\lim_{m \rightarrow \infty} |x[-m]|^{1/m}}$$

Si cette intersection existe, on a une couronne de convergence (Figure 4-6), sinon la transformée ne converge pas.

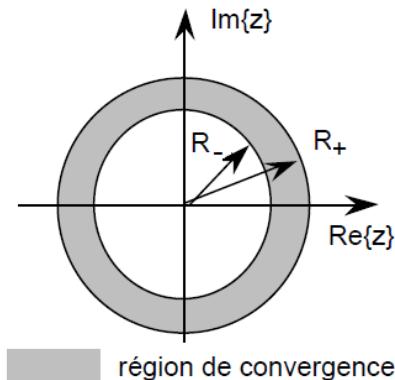


Figure 4-6. Couronne de convergence de la transformée en Z

D'une façon générale et pour les  $X(z)$  causales, le domaine de convergence est borné inférieurement par le plus grand pôle. En d'autres termes, les valeurs de  $z$  doivent se situer à l'extérieur d'un cercle qui passe par le pôle le plus loin de l'origine.

### Exemples :

- La transformée en Z de l'impulsion unité est donnée par :

$$\text{Eq 4-21} \quad Z\{\delta[n]\} = \sum_{n=-\infty}^{+\infty} \delta[n].z^{-n} = 1. \text{Convergente } \forall z.$$

- La transformée en Z de l'échelon unité est donnée par :

$$\text{Eq 4-22}$$

$$U(z) = Z\{u[n]\} = \sum_{n=0}^{+\infty} u[n].z^{-n} = 1 + z^{-1} + z^{-2} + z^{-3} + \dots = \frac{1}{1 - z^{-1}} = \frac{z}{z - 1}.$$

converge pour  $|z| > 1$  ( $R_- = 1$  et  $R_+ = +\infty$ ). Notons que la racine du dénominateur est  $z = 1$ . Il

s'agit d'un "pôle" de  $U(z)$ .

3. La transformée en Z du signal "rampe"  $r[n] = \begin{cases} n & \text{pour } n \geq 0 \\ 0 & \text{pour } n < 0 \end{cases}$  est donnée par :

Éq 4-23

$$R(z) = Z\{r[n]\} = \sum_{n=0}^{+\infty} r[n] \cdot z^{-n} = z^{-1} + 2z^{-2} + 3z^{-3} + \dots = \frac{z^{-1}}{(1 - z^{-1})^2} = \frac{z}{(z - 1)^2}.$$

converge pour  $|z| > 1$  ( $R_- = 1$  et  $R_+ = +\infty$ ).

4. La transformée en Z de  $x[n] = a^n \cdot u[n]$  est donnée par :

Éq 4-24

$$X(z) = \sum_{n=0}^{+\infty} x[n] \cdot z^{-n} = \sum_{n=0}^{+\infty} a^n \cdot z^{-n} = \sum_{n=0}^{+\infty} (a \cdot z^{-1})^n = \frac{1}{1 - a \cdot z^{-1}} = \frac{z}{z - a}.$$

converge pour  $|z| > |a|$  ( $R_- = |a|$  et  $R_+ = +\infty$ ).

Pour vérifier, vous pouvez calculer  $X(z)$  pour  $z=2a$  et pour  $z=a/2$  :

$$X(2a) = \sum_{n=0}^{+\infty} (a \cdot (2a)^{-1})^n = \sum_{n=0}^{+\infty} \left(\frac{1}{2}\right)^n = 2, \text{ et } X\left(\frac{a}{2}\right) = \sum_{n=0}^{+\infty} \left(a \cdot \left(\frac{a}{2}\right)^{-1}\right)^n = \sum_{n=0}^{+\infty} 2^n = +\infty$$

Notons que pour  $x[n] = a^n$ ,  $R_- = |a|$  et  $R_+ = |a|$ , ce qui signifie que la suite  $X(z)$  ne converge pas.

## Linéarité

Éq 4-25

$$Z\{a \cdot x_1[n] + b \cdot x_2[n]\} = a \cdot Z\{x_1[n]\} + b \cdot Z\{x_2[n]\} = a \cdot X_1(z) + b \cdot X_2(z)$$

Avec convergence, **au moins**, dans la région commune de convergence de  $X_1(z)$  et  $X_2(z)$  :

$$R_- = \max \{R_{x1-}, R_{x2-}\} \text{ et } R_+ = \min \{R_{x1+}, R_{x2+}\}$$

Si la combinaison compense certains pôles, la région de convergence peut être plus grande.

## Dérivation

Éq 4-26

$$\frac{dX(z)}{dz} = \sum_{n=-\infty}^{+\infty} -n \cdot x[n] \cdot z^{-n-1} = -z^{-1} \cdot \sum_{n=-\infty}^{+\infty} n \cdot x[n] \cdot z^{-n}$$

$$\text{Ce qui donne: } Z\{n \cdot x[n]\} = -z \cdot \frac{dX(z)}{dz}$$

Le calcul de  $Z\{r[n]\}$  de l'exemple 3 ci-dessus en est une application.

Cette propriété peut être généralisée à l'ordre  $k$  comme suit :

$$\text{Éq 4-27} \quad Z\{n^k \cdot x[n]\} = (-z)^k \cdot \frac{d^k X(z)}{dz^k}$$

## Inter-corrélation

L'inter-corrélation entre deux signaux discrets  $x[n]$  et  $y[n]$  est donnée par :

Éq 4-28

$$\varphi_{xy}[n] = \sum_{k=-\infty}^{+\infty} x[k] \cdot y[n+k] = x[-n] * y[n]$$

Sa transformée en Z est :

$$\text{Éq 4-29} \quad \Phi_{xy}(z) = Z\{x[-n]\}.Y(z) = X\left(\frac{1}{z}\right).Y(z)$$

### Théorème de la valeur initiale

Pour une suite causale  $x[n]$  dont la transformée en Z est  $X(z)$ , le théorème de la valeur initiale montre que :

$$\text{Éq 4-30} \quad x[0] = \lim_{|z| \rightarrow \infty} X(z)$$

### Théorème de la valeur finale

De même, le théorème de la valeur finale montre que, lorsque la limite existe :

$$\text{Éq 4-31} \quad \lim_{n \rightarrow +\infty} x[n] = \lim_{z \rightarrow +1} (z - 1).X(z)$$

### Table des transformées en Z usuelles:

$x[n]$ (suites causales)	$X(z)$	Région de convergence
$\delta[n]$	1	$\mathbb{C}$
$\delta[n-k]$	$z^{-k}$	$\mathbb{C}^*$
1	$\frac{z}{z-1}$	$ z >1$
n	$\frac{z}{(z-1)^2}$	$ z >1$
$a^n$	$\frac{z}{z-a}$	$ z > a $
$n.a^n$	$\frac{a.z}{(z-a)^2}$	$ z > a $
$\cos(n\omega)$	$\frac{z^2 - z \cdot \cos(\omega)}{z^2 - 2z \cdot \cos(\omega) + 1}$	$ z >1$
$\sin(n\omega)$	$\frac{z \cdot \sin(\omega)}{z^2 - 2z \cdot \cos(\omega) + 1}$	$ z >1$
$a^n \cdot \cos(n\omega)$	$\frac{z^2 - az \cdot \cos(\omega)}{z^2 - 2az \cdot \cos(\omega) + a^2}$	$ z > a $
$a^n \cdot \sin(n\omega)$	$\frac{az \cdot \sin(\omega)}{z^2 - 2az \cdot \cos(\omega) + a^2}$	$ z > a $
$a^n \cdot x[n]$	$X(z/a)$	
$x[n-k]$	$z^{-k} \cdot X(z)$	
$x[n+k]$	$z^k \cdot X(z) - \sum_{j=0}^{k-1} x[j] \cdot z^{k-j}$	
$n^k \cdot x[n]$	$(-z)^k \cdot \frac{d^k X(z)}{dz^k}$	

### Systèmes en cascade

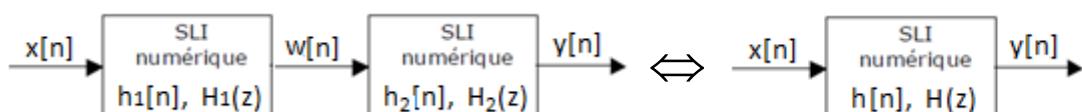


Figure 4-7. Systèmes en cascade

Lorsque on met deux systèmes discrets linéaires et invariants dans le temps en cascade (la sortie du premier est connectée à l'entrée de l'autre, (Figure 4-7) de réponses impulsionales

respectives  $h_1[n]$  et  $h_2[n]$ , le système résultant est aussi linéaire et invariant de réponse impulsionnelle  $h[n] = h_1[n]*h_2[n]$ , dont la transformée en Z est donnée par  $H(z) = H_1(z).H_2(z)$ , où  $H_1(z)$  et  $H_2(z)$  sont les transformées en Z des deux systèmes.

### 4.3.2 Réponse en fréquence d'un SLI numérique

#### Réponse de régime à une exponentielle imaginaire numérique

Soit un SLI numérique excité par  $x[n] = z^n$  ( $n = -\infty, \dots, +\infty$ ). La réponse du système est donnée par:

Éq 4-32

$$y[n] = x[n] * h[n] = \sum_{l=-\infty}^{+\infty} h[l].x[n-l] = \sum_{l=-\infty}^{+\infty} h[l].z^{n-l} = z^n \sum_{l=-\infty}^{+\infty} h[l].z^{-l} = H(z).z^n$$

Éq 4-33

$$y[n] = H(z).z^n|_{z=e^{j\omega}} = H(\omega).e^{jn\omega}$$

où  $H(\omega)$  n'est autre que la transformée de Fourier de  $h[n]$  (avec  $T_e=1$ ), donnée par (voir Éq 2-3) :

Éq 4-34

$$H(\omega) = H(z)|_{z=e^{j\omega}} = \sum_{l=-\infty}^{+\infty} h[l].e^{-jl\omega}$$

On constate donc que l'exponentielle imaginaire numérique  $x[n] = e^{jn\omega}$  ( $n = -\infty, \dots, +\infty$ ) est une *fonction propre* de tout SLI numérique: la réponse d'un SLI à une exponentielle imaginaire n'est autre que l'exponentielle imaginaire d'entrée, multiplié par un facteur complexe  $H(\omega)$  qui dépend de la pulsation  $\omega$  de l'exponentielle.

$H(\omega)$  est appelée *réponse en fréquence*, ou *transmittance isochrone*, du SLI. Elle est la TF de sa réponse impulsionnelle  $h[n]$  et égale à sa transformée en Z calculée sur le cercle de rayon unité.

La réponse en régime sinusoïdal d'un SLI est évidemment directement liée à la réponse en fréquence. En effet, pour  $x[n] = A.\cos(n\omega + \theta) = \frac{A}{2}[e^{j(n\omega + \theta)} + e^{-j(n\omega + \theta)}]$ ,  $y[n]$  s'écrit:

Éq 4-35

$$\begin{aligned} y[n] &= \frac{A}{2}[H(\omega).e^{j(n\omega + \theta)} + H(-\omega).e^{-j(n\omega + \theta)}] = \frac{A}{2}[H(\omega).e^{j(n\omega + \theta)} + \{H(\omega).e^{j(n\omega + \theta)}\}^*] \\ &= A.\operatorname{Re}\{H(\omega).e^{j(n\omega + \theta)}\} = A.|H(\omega)|.\cos(n\omega + \theta + \operatorname{Arg}(H(\omega))) \end{aligned}$$

#### Trois méthodes pour déterminer la réponse en fréquence d'un système SLI numérique

$H(\omega)$  peut être déterminée par l'une de ces 3 méthodes :

- La TF de la réponse impulsionnelle  $h[n]$  selon l'Éq 4-34
- À partir de la réponse du système au signal  $e^{jn\omega}$  (donnée par l'Éq 4-34) en remplaçant cette réponse dans l'équation aux différences.
- À partir de la fonction de transfert en Z du système en remplaçant  $z$  par  $e^{j\omega}$ .

Nous allons appliquer ces 3 méthodes à travers l'exemple suivant :

Soit le système LI donné par la structure de la Figure 4-8 (où  $|a|<1$ ) et dont l'équation aux différences est donnée par :  $y[n]=a.y[n-1]+x[n]$ .

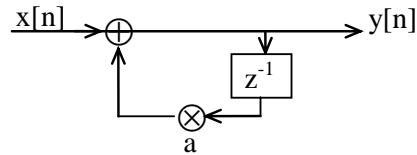


Figure 4-8. Système SLI simple

#### 1<sup>ère</sup> méthode :

La réponse impulsionnelle  $h[n]$  est donnée par :  $h[n]=a.h[n-1]+\delta[n]$ .

$h[0]=1$ ,  $h[1]=a$ ,  $h[2]=a^2$ , ... et  $h[n]=a^n.u[n]$ .

En appliquant l'Éq 4-34, la réponse en fréquence  $H(\omega)$  est alors :

$$H(\omega) = \sum_{l=0}^{+\infty} a^l \cdot e^{-jl\omega} = \sum_{l=0}^{+\infty} (a \cdot e^{-j\omega})^l = \frac{1}{1 - a \cdot e^{-j\omega}}$$

#### 2<sup>ème</sup> méthode :

Prenons le signal d'entrée  $x[n]=e^{jn\omega}$ . Le signal de sortie est alors  $y[n]=H(\omega).e^{jn\omega}$  et par conséquence,  $y[n-1]=H(\omega).e^{j(n-1)\omega}$ . En remplaçant chaque terme par sa valeur dans l'équation aux différences, nous obtenons:  $H(\omega).e^{jn\omega} = a \cdot H(\omega).e^{j(n-1)\omega} + e^{jn\omega}$ , qui donne  $H(\omega) = \frac{1}{1-a.e^{-j\omega}}$ .

#### 3<sup>ème</sup> méthode :

En effectuant la transformée en Z de l'équation aux différences, nous obtenons:  $Y(z) = a.z^{-1}.Y(z) + X(z)$ , qui donne  $H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1-a.z^{-1}}$ . D'où,  $H(\omega) = H(z)|_{z=e^{j\omega}} = \frac{1}{1-a.e^{-j\omega}}$ .

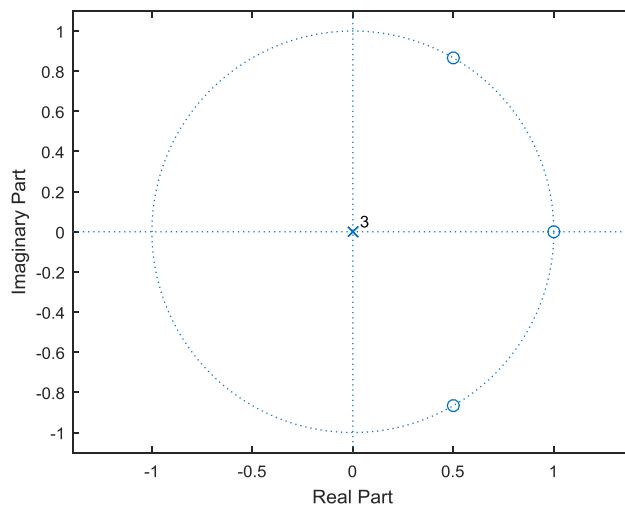


Figure 4-9. Zéros (o) et pôles (x) d'une  $H(z)$  dans le plan complexe avec le cercle unité

### 4.3.3 Pôles et Zéros de $H(z)$

Nous avons vu dans l'Éq 4-19 que la forme générale de la fonction de transfert en z d'un SLI est

un quotient de deux polynômes en  $z$  dont le degré du numérateur  $M$  et celui du dénominateur est  $N$ .

On appelle "zéros" les racines du numérateur et "pôles" celles du dénominateur. La connaissance des pôles et des zéros détermine, à un facteur multiplicatif près, la fonction  $H(z)$ .

Il est utile de représenter graphiquement les zéros et les pôles de  $H(z)$  comme des points dans le plan complexe (plan des  $z$ ).

La Figure 4-9 montre les pôles et les zéros d'une fonction  $H(z) = 1 - 2z^{-1} + 2z^{-2} - z^{-3}$  en utilisant la fonction `zplane` de Matlab comme suit : `zplane([1 -2 2 -1], [1])`.

Le cercle unité est le lieu des valeurs de  $z$  pour lesquelles on évalue la réponse fréquentielle du système en remplaçant  $z$  par  $e^{j\omega}$ :  $\omega$  est la pulsation normalisée qui correspond aux fréquences normalisées (par rapport à la fréquence d'échantillonnage  $\frac{1}{T}$ ) comprises entre  $-\frac{\pi}{2}$  et  $+\frac{\pi}{2}$ .  $\omega$  varie de  $-\pi$  à  $+\pi$ .

En fonction des pôles  $p_i$  et des zéros  $z_i$ ,  $H(z)$  peut être mise sous la forme suivante :

Éq 4-36

$$H(z) = K \cdot z^{-(M-N)} \frac{\prod_{i=1}^M (z - z_i)}{\prod_{i=1}^N (z - p_i)} = K_0 \frac{\prod_{i=1}^M (1 - z_i \cdot z^{-1})}{\prod_{i=1}^N (1 - p_i \cdot z^{-1})}$$

avec  $K_0 = \lim_{z \rightarrow +\infty} H(z)$ . Cette écriture se prête à une interprétation géométrique permettant de calculer la réponse en fréquence d'un système numérique en fonction de la position de ses pôles et zéros. On en conclut que le module de la réponse en fréquence est obtenu (à  $K_0$  près) en multipliant les normes des vecteurs joignant les zéros du système à un point du cercle de rayon unité, et en divisant le résultat par le produit des normes des vecteurs joignant les pôles à ce même point.

De même la réponse en phase correspond à la somme des arguments des vecteurs issus des zéros moins la somme des arguments des vecteurs issus des pôles.

À partir de la représentation graphique des pôles et des zéros de la Figure 4-9, on peut déduire que la réponse aux fréquences 0,  $+1/6$  et  $-1/6$  est nulle.

Le pôle  $p_1=0$  est d'ordre 3. En effet,  $H(z) = 1 - 2z^{-1} + 2z^{-2} - z^{-3} = \frac{(z-1)(z - e^{-j\frac{\pi}{3}})(z - e^{j\frac{\pi}{3}})}{z^3}$ .

Notons que lorsque  $H(z)$  ne comporte que des puissances de  $z$  négatives, on peut écrire les facteurs sous la forme  $(1 - a.z^{-1})$  dont la racine est  $a$ . Ce facteur n'est autre que  $\frac{z-a}{z}$ , ce qui justifie la présence des pôles 0 dans la transformée en Z d'un système à réponse impulsionnelle finie.

Il est à remarquer que la gamme de fréquences dans laquelle l'influence des pôles et des zéros est la plus forte est celle qui correspond aux parties du cercle unité les plus proches de ces pôles ou de ces zéros.

De plus, l'influence d'un pôle ou d'un zéro sur la réponse en fréquence est d'autant plus grande qu'il se trouve plus proche du cercle unité. Dans le cas extrême où un zéro se trouve sur le cercle unité, l'amplitude de la réponse en fréquence s'annule pour la fréquence correspondante à ce zéro. À l'opposé, la réponse en fréquence a une amplitude infinie à la fréquence qui correspond à un pôle se trouvant sur le cercle unité.

Si tous les zéros sont à l'intérieur du cercle unité, on se trouve devant un système "à phase minimale", qui correspond à la plus faible valeur possible du temps de propagation de groupe ( $-d\phi/d\omega$ ) pour toutes les fréquences. Notons qu'on peut toujours améliorer la caractéristique de phase d'un système ayant des zéros à l'extérieur du cercle unité en les remplaçant par des zéros dont les modules sont inversés (sans changer leurs arguments respectifs) : un zéro donné par  $|r| \cdot e^{j\theta}$  est remplacé par  $|1/r| \cdot e^{j\theta}$  (où  $|r| > 1$ ).

Un système est dit "tout pôle" s'il ne possède que des pôles (et éventuellement des zéros à l'origine).

Si un système possède uniquement des zéros symétriques par paires par rapport au cercle unité ou sur le cercle unité même (avec éventuellement des pôles à l'origine), alors il est dit "à phase linéaire".

Pour les systèmes mis en cascade, les pôles et les zéros peuvent se coïncider. Dans ce cas, un pôle et un zéro coïncidés s'annulent mutuellement.

Pour un système causal, le nombre de pôles est supérieur ou égal à celui des zéros (y compris les pôles et zéros à l'origine). L'ordre du système est le degré du polynôme du dénominateur.

#### 4.3.4 L'inversion de la transformée en Z

Pour retrouver la séquence numérique relative à une fonction en Z donnée, on peut utiliser la Transformation en Z inverse dont la définition formelle est la suivante :

Éq 4-37

$$x[n] = \frac{1}{j2\pi} \oint_C X(z) \cdot z^{n-1} dz$$

C'est une intégrale sur un chemin quelconque fermé du plan des Z parcouru dans le sens inverse des aiguilles d'une montre et situé dans la région de convergence et renfermant l'origine.

Ceci paraît bien compliqué ce qui la rend rarement utilisée dans la pratique. Plusieurs méthodes plus pratiques peuvent être utilisées pour identifier les  $x[n]$ . Nous en présentons 4 méthodes dont deux aboutissent à des solutions analytiques alors que les deux autres donnent lieu à des solutions numériques.

#### Méthode des Résidus:

Le théorème de Cauchy permet d'écrire la relation de l'Éq 4-37 sous forme d'une somme des résidus de la fonction à intégrer  $X(z) \cdot z^{n-1}$  dans le contour C, relatifs à tous les pôles de  $X(z) \cdot z^{n-1}$ :

Éq 4-38

$$x[n] = \frac{1}{j2\pi} \oint_C X(z) \cdot z^{n-1} dz = \sum_{p_i} \lim_{z \rightarrow p_i} \left\{ \frac{1}{(m_i - 1)!} \frac{d^{m_i-1}}{dz^{m_i-1}} [z^{n-1} X(z) \cdot (z - p_i)^{m_i}] \right\}$$

où  $p_i$  est l'un des pôles de  $X(z) \cdot z^{n-1}$  de multiplicité  $m_i$ .

#### Exemple 1:

Soit à inverser la fonction  $X(z) = \frac{z}{(z-1)^2}$ , qui a un pôle double ( $m=2$ ) en  $z=1$ . En appliquant l'Éq 4-38, pour  $p_1=1$  et  $m_1=2$ , nous obtenons :

$$x[n] = \lim_{z \rightarrow 1} \left\{ \frac{d}{dz} \left[ z^{n-1} \frac{z}{(z-1)^2} \cdot (z-1)^2 \right] \right\} = \lim_{z \rightarrow 1} \left\{ \frac{d}{dz} [z^n] \right\} = n \cdot z^{n-1}|_{z=1} = n$$

**Exemple 2:** Soit  $X(z) = \frac{1-2.1z^{-1}}{1-0.3z^{-1}-0.4z^{-2}}$  avec  $D(z) = 1 - 0.3z^{-1} - 0.4z^{-2} = (1 + 0.5z^{-1})(1 - 0.8z^{-1})$

Ce qui permet d'écrire  $X(z)$  sous la forme :  $X(z) = \frac{z(z-2.1)}{(z+0.5)(z-0.8)}$ . En appliquant la méthode des résidus, nous obtenons :

$$\begin{aligned} x[n] &= \lim_{z \rightarrow 0.8} \left\{ z^{n-1} \frac{z(z-2.1)}{(z+0.5)} \right\} + \lim_{z \rightarrow -0.5} \left\{ z^{n-1} \frac{z(z-2.1)}{(z-0.8)} \right\} \\ &= \lim_{z \rightarrow 0.8} \left\{ \frac{z^n(z-2.1)}{(z+0.5)} \right\} + \lim_{z \rightarrow -0.5} \left\{ \frac{z^n(z-2.1)}{(z-0.8)} \right\} = -(0.8)^n + 2(-0.5)^n \end{aligned}$$

### Décomposition en fractions rationnelles

Cette méthode consiste à décomposer  $X(z)$  en éléments simples dont on connaît les inverses (en utilisant les tables des fonctions usuelles, par exemple). Ces éléments sont en général des fractions rationnelles.

L'idée de base consiste à exprimer  $X(z) = \frac{N(z)}{D(z)}$  sous forme d'une somme de fractions simples en factorisant le dénominateur.

Si  $D(z)$  est un polynôme de degré  $N$ , alors on peut l'exprimer avec ses  $K$  racines ( $K \leq N$ ) sous forme de produits de  $K$  facteurs comme suit :

Éq 4-39

$$D(z) = \prod_{k=1}^K (z - p_k)^{m_k}$$

Les fractions simples relatives au pôle  $p_k$  sont constituées de  $m_k$  termes comme suit :

Éq 4-40

$$\frac{c_1^k}{(z - p_k)} + \frac{c_2^k}{(z - p_k)^2} + \dots + \frac{c_{m_k}^k}{(z - p_k)^{m_k}}$$

Dans le cas où  $m_k$  est égal à 1 (pôle simple), cette somme se réduit au premier terme uniquement. Ainsi  $X(z)$  s'exprime par :

Éq 4-41

$$X(z) = \sum_{k=1}^K \sum_{i=1}^{m_k} \frac{c_i^k}{(z - p_k)^i}$$

Les  $c_i^k$  doivent être calculés en utilisant l'une des deux méthodes suivantes :

1. Par développement de la somme de manière à avoir un dénominateur commun qui n'est autre que  $D(z)$  de l'Éq 4-39. Les coefficients du numérateur résultant seront identifiés aux coefficients correspondants du polynôme  $N(z)$  pour obtenir les valeurs des  $c_i^k$ .

2. Par la méthode des résidus qui donne les coefficients  $c_i^k$  par la formule suivante :

$$c_i^k = \frac{1}{(m_k - i)!} \left. \frac{d^{m_k-i}}{dz^{m_k-i}} [(z - p_k)^{m_k} \cdot X(z)] \right|_{z=p_k}$$

Notons que, dans le cas où  $m_k$  est égal à 1 (cas le plus fréquent),  $c_1^k$  se simplifie pour devenir :  $c_1^k = [(z - p_k) \cdot X(z)]|_{z=p_k}$

### Exemples :

#### Exemple 1 : Pôles réels simples

Soit  $X(z) = \frac{1-2.1 z^{-1}}{1-0.3 z^{-1}-0.4 z^{-2}}$

$$D(z) = 1 - 0.3 z^{-1} - 0.4 z^{-2} = (1 + 0.5 z^{-1})(1 - 0.8 z^{-1})$$

Les 2 pôles  $p_1=-0.5$  et  $p_2=0.8$  sont simples.  $X(z)$  s'écrit alors :

$$X(z) = \frac{c_1^1}{(1 + 0.5 z^{-1})} + \frac{c_1^2}{(1 - 0.8 z^{-1})}$$

#### **Selon la 1<sup>ère</sup> méthode :**

$$X(z) = \frac{c_1^1 \cdot (1 - 0.8 z^{-1}) + c_1^2 \cdot (1 + 0.5 z^{-1})}{(1 + 0.5 z^{-1}) \cdot (1 - 0.8 z^{-1})} = \frac{c_1^1 + c_1^2 + (0.5 c_1^2 - 0.8 c_1^1) z^{-1}}{(1 + 0.5 z^{-1}) \cdot (1 - 0.8 z^{-1})}$$

Par identification :

$$c_1^1 + c_1^2 = 1 \quad \text{et} \quad 0.5 c_1^2 - 0.8 c_1^1 = -2.1$$

Deux équations à deux inconnus dont la solution est  $c_1^1=2$  et  $c_1^2=-1$

Ce qui donne :  $X(z) = \frac{2}{(1+0.5 z^{-1})} - \frac{1}{(1-0.8 z^{-1})}$

dont la transformée inverse est :  $x[n]=[2(-0.5)^n - (0.8)^n]u[n]$ .

#### **Selon la 2<sup>ème</sup> méthode :**

$$c_1^1 = [(1 + 0.5 z^{-1}) \cdot X(z)]|_{z=-0.5} = \left. \frac{1 - 2.1 z^{-1}}{1 - 0.8 z^{-1}} \right|_{z=-0.5} = \frac{1 + 4.2}{1 + 1.6} = 2$$

$$c_1^2 = [(1 - 0.8 z^{-1}) \cdot X(z)]|_{z=0.8} = \left. \frac{1 - 2.1 z^{-1}}{1 + 0.5 z^{-1}} \right|_{z=0.8} = \frac{1 - 2.1/0.8}{1 + 0.5/0.8} = -1$$

#### Exemple 2 : Pôles complexes conjugués

- Soit  $X(z)=\frac{1}{1+z^{-2}}$  avec  $D(z)=1 + z^{-2} = (1 + j z^{-1})(1 - j z^{-1}) = \left(1 - e^{-j\frac{\pi}{2}}z^{-1}\right)\left(1 - e^{j\frac{\pi}{2}}z^{-1}\right)$

$$X(z) = \frac{c_1^1}{(1 + j z^{-1})} + \frac{c_1^2}{(1 - j z^{-1})} = \frac{1}{2} \left[ \frac{1}{\left(1 - e^{-j\frac{\pi}{2}}z^{-1}\right)} + \frac{1}{\left(1 - e^{j\frac{\pi}{2}}z^{-1}\right)} \right]$$

dont la transformée inverse est:  $x[n] = \frac{1}{2} \left\{ e^{-jn\frac{\pi}{2}} + e^{jn\frac{\pi}{2}} \right\} u[n] = \cos\left(n\frac{\pi}{2}\right) \cdot u[n]$

- Soit  $X(z) = \frac{2z^2+6z-26}{z(z^2+4z+13)}$ ,  $D(z)=z(z^2 + 4z + 13) = z(z + 2 + 3j)(z + 2 - 3j)$

Les 3 pôles  $p_1 = 0$ ,  $p_2 = -2 - 3j = \sqrt{13}e^{j\varphi}$  et  $p_3 = -2 + 3j = \sqrt{13}e^{-j\varphi}$  sont simples ( $\varphi = \arctg(3/2)$ ).  $X(z)$  s'écrit alors :  $X(z) = \frac{c_1^1}{z} + \frac{c_1^2}{(z+2+3j)} + \frac{c_1^3}{(z+2-3j)}$

**Selon la 1<sup>ère</sup> méthode :**

$$X(z) = \frac{(c_1^1 + c_1^2 + c_1^3)z^2 + (4c_1^1 + (2-3j)c_1^2 + (2+3j)c_1^3)z + 13c_1^1}{z(z+2+3j)(z+2-3j)}$$

Par identification :

$$c_1^1 + c_1^2 + c_1^3 = 2, \quad 4c_1^1 + (2-3j)c_1^2 + (2+3j)c_1^3 = 6 \quad \text{et} \quad 13c_1^1 = -26$$

Trois équations à trois inconnus dont la solution est :

$$c_1^1 = -2, \quad c_1^2 = 2+j = \sqrt{5}e^{j\theta} \quad \text{et} \quad c_1^3 = 2-j = \sqrt{5}e^{-j\theta} \quad \text{avec } \theta = \arctg(\frac{j}{2})$$

Ce qui donne:

$$\begin{aligned} X(z) &= \frac{-2}{z} + \frac{2+j}{(z+2+3j)} + \frac{2-j}{(z+2-3j)} \\ &= z^{-1} \left[ -2 + \frac{\sqrt{5}e^{j\theta}}{(1-\sqrt{13}e^{j\varphi}z^{-1})} + \frac{\sqrt{5}e^{-j\theta}}{(1-\sqrt{13}e^{-j\varphi}z^{-1})} \right] \end{aligned}$$

dont la transformée inverse est:

$$\begin{aligned} x[n] &= \left\{ -2\delta[n-1] + \sqrt{5}e^{j\theta}(\sqrt{13}e^{j\varphi})^{n-1} + \sqrt{5}e^{-j\theta}(\sqrt{13}e^{-j\varphi})^{n-1} \right\} u[n] \\ &= -2 \left\{ \delta[n-1] - \sqrt{5}(\sqrt{13})^{n-1} \cos[(n-1)\varphi + \theta] \right\} u[n] \end{aligned}$$

$$\text{Selon la 2<sup>ème</sup> méthode: } c_1^1 = [z \cdot X(z)]|_{z=0} = \frac{2z^2 + 6z - 26}{(z^2 + 4z + 13)}|_{z=0} = \frac{-26}{13} = -2$$

$$\begin{aligned} c_1^2 &= [(z+2+3j) \cdot X(z)]|_{z=-2-3j} = \frac{2z^2 + 6z - 26}{z(z+2-3j)}|_{z=-2-3j} \\ &= \frac{2(-2-3j)^2 - 12 - 18j - 26}{(-2-3j)(-6j)} = \frac{2(-5+12j) - 12 - 18j - 26}{(12j-18)} \\ &= \frac{-48+6j}{-18+12j} = \frac{-8+j}{-3+2j} = \frac{26+13j}{9+4} = 2+j \end{aligned}$$

$$\begin{aligned} c_1^3 &= [(z+2-3j) \cdot X(z)]|_{z=-2+3j} = \frac{2z^2 + 6z - 26}{z(z+2+3j)}|_{z=-2+3j} \\ &= \frac{2(-2+3j)^2 - 12 + 18j - 26}{(-2+3j)(6j)} = \frac{2(-5-12j) - 12 + 18j - 26}{(-12j-18)} \\ &= \frac{-48-6j}{-18-12j} = \frac{8+j}{3+2j} = \frac{26-13j}{9+4} = 2-j \end{aligned}$$

## Division polynomiale suivant les puissances croissantes de $z^{-1}$

On effectue une division euclidienne du polynôme en  $z^{-1}$  du numérateur et le polynôme en  $z^{-1}$  du dénominateur de  $X(z)$ . Le résultat sera un polynôme en  $z^{-1}$  dont les coefficients sont les valeurs de  $x[n]$ .

Notons que cette méthode ne donne pas nécessairement une expression analytique de  $x[n]$  comme c'est le cas des deux méthodes précédentes. Ceci peut entraîner une forme incomplète de  $x[n]$  du fait que le résultat de la division peut être infini (cas d'une fonction de transfert dont la réponse impulsionnelle est infinie, par exemple).

#### Exemple :

$$\text{Soit } X(z) = \frac{1-2.1z^{-1}}{1-0.3z^{-1}-0.4z^{-2}}$$

La division euclidienne donne :

$$X(z) = \frac{1 - 2.1 z^{-1}}{1 - 0.3 z^{-1} - 0.4 z^{-2}} = 1 - 1.8 z^{-1} - 0.14 z^{-2} - 0.762 z^{-3} - 0.2846 z^{-4} + \dots$$

Comme  $z^{-k}$  est la transformée en  $z$  de  $\delta[n-k]$ , alors :

$$x[n] = \{\delta[n] - 1.8\delta[n-1] - 0.14\delta[n-2] - 0.762\delta[n-3] - 0.2846\delta[n-4] + \dots\}u[n]$$

$$\begin{array}{r|l} 1 - 2.1 z^{-1} & 1 - 0.3 z^{-1} - 0.4 z^{-2} \\ 1 - 0.3 z^{-1} - 0.4 z^{-2} & \hline \\ 0 - 1.8 z^{-1} + 0.4 z^{-2} & 1 - 1.8 z^{-1} - 0.14 z^{-2} - 0.762 z^{-3} - 0.2846 z^{-4} + \dots \\ -1.8 z^{-1} + 0.54 z^{-2} + 0.72 z^{-3} & \hline \\ -0.14 z^{-2} - 0.72 z^{-3} & -0.762 z^{-3} - 0.056 z^{-4} \\ -0.14 z^{-2} + 0.042 z^{-3} & -0.762 z^{-3} + 0.2286 z^{-4} + 0.3048 z^{-5} \\ \hline & \hline \\ & -0.2846 z^{-4} - 0.3048 z^{-5} \\ & \dots\dots\dots \end{array}$$

#### **Par l'équation aux différences (équation récurrente pour une impulsion unité)**

Cette méthode consiste à déduire les valeurs de  $x[n]$  en utilisant l'équation aux différences correspondante à  $\frac{X(z)}{\Delta(z)}$ ,  $\Delta(z)$  étant la transformée en  $z$  de l'impulsion unité et qui vaut 1.

Comme la méthode précédente, celle-ci ne donne pas une expression analytique de  $x[n]$ .

#### Exemple :

Pour  $X(z) = \frac{1-2.1z^{-1}}{1-0.3z^{-1}-0.4z^{-2}}$ , on détermine l'équation aux différences de la fonction de transfert  $\frac{X(z)}{\Delta(z)} = \frac{1-2.1z^{-1}}{1-0.3z^{-1}-0.4z^{-2}}$  qui est donnée par :

$$x[n] = \delta[n] - 2.1\delta[n-1] + 0.3x[n-1] + 0.4x[n-2]$$

À partir de cette équation, on peut calculer les valeurs de  $x[n]$  à partir de  $n=0$ , sachant que  $x[n]$  est causal, c'est-à-dire nulle pour les  $n<0$ .

Pour cet exemple, nous obtenons :

$$x[0] = \delta[0] = 1. \quad x[1] = \delta[1] - 2.1\delta[0] + 0.3x[0] = -2.1 + 0.3 = -1.8$$

$$x[2] = \delta[2] - 2.1\delta[1] + 0.3x[1] + 0.4x[0] = 0.3(-1.8) + 0.4 = -0.14$$

$$x[3] = 0.3x[2] + 0.4x[1] = 0.3(-0.14) + 0.4(-1.8) = -0.762$$

$$x[4] = 0.3x[3] + 0.4x[2] = 0.3(-0.762) + 0.4(-0.14) = -0.2846$$

$$x[5]=\dots$$

Ainsi on retrouve la même solution de la méthode précédente :

$$x[n] = \{\delta[n] - 1.8\delta[n-1] - 0.14\delta[n-2] - 0.762\delta[n-3] - 0.2846\delta[n-4] + \dots\}u[n]$$

# Chapitre 5 – Les Filtres Numériques

Le filtrage consiste à modifier la distribution fréquentielle d'un signal selon des spécifications données. Cette opération est appelée **filtrage numérique** lorsqu'elle est effectuée sur un signal numérique et à l'aide d'un système numérique en utilisant des opérations arithmétiques.

Pour des raisons historiques, les filtres numériques ont constitué la plus branche la plus étudiée du traitement numérique de signaux. Ils ont été développés et étudiés dans le but de pouvoir simuler les filtres analogiques sur ordinateur. Ceci a permis de vérifier les performances et d'optimiser les paramètres de ces filtres avant leur éventuelle réalisation.

Comme on l'a déjà mentionné à maintes reprises, la réponse fréquentielle (ou harmonique)  $H(\omega)$  d'un système linéaire invariant dans le temps est liée aux transformées de Fourier  $X(\omega)$  et  $Y(\omega)$  des signaux d'entrée et de sortie par:

$$\text{Eq 5-1} \quad Y(\omega) = H(\omega).X(\omega)$$

Cette relation indique que les répartitions fréquentielles de l'amplitude et de la phase du signal d'entrée  $x[n]$  sont modifiées par le système selon la forme particulière de la fonction complexe  $H(\omega)$  pour satisfaire au mieux les exigences du traitement particulier envisagé.

Puisque  $H(\omega)$  détermine les atténuations ou les amplifications des composantes de diverses fréquences du signal d'entrée, le système correspondant est appelé filtre.

Dans le domaine temporel, la relation correspondante à l'Eq 5-1 est le produit de convolution numérique donné par:  $y[n] = x[n] * h[n]$  où  $h[n]$  est la réponse impulsionnelle du filtre  $H$ .

## 5.1 Classification des filtres:

Selon la durée de la réponse impulsionnelle, on distingue deux larges catégories des filtres qui sont:

- Filtres à Réponse Impulsionnelle Infinie (RII): les valeurs de  $h[n]$  sont non nulles sur un intervalle infini  $n_0 \leq n \leq +\infty$ .
- Filtres à Réponse Impulsionnelle Finie (RIF): dans ce cas, les valeurs de  $h[n]$  sont non nulles seulement sur un intervalle de durée finie  $L$  avec  $n_0 \leq n \leq n_0 + L - 1$ .

Selon leur réponse fréquentielle, on classe les filtres dans quatre grandes catégories qui sont:

- Filtres Passe-bas
- Filtres Passe-haut
- Filtres Passe-bande
- Filtres Coupe-bande.

Chacune de ces catégories est représentée par un filtre idéalisé dans le sens où tout ce qui doit être atténué doit l'être complètement et tout ce qui doit être transféré à la sortie doit l'être sans

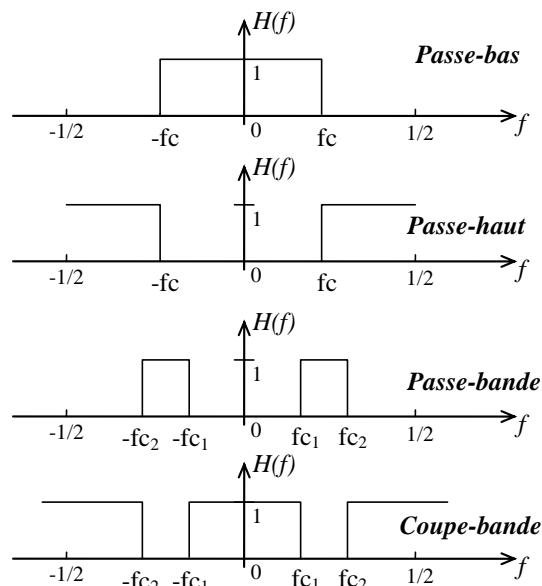


Figure 5-1. Réponses fréquentielles des filtres idéaux

modification.

Pour un signal discret, la répartition fréquentielle est périodique et de période  $f_e = \frac{1}{T} = \text{nombre d'échantillons par seconde}$ . Sur la période principale comprise entre  $\pm \frac{f_e}{2} = \pm \frac{1}{2T}$ , on peut distinguer les hautes et les basses fréquences tout en gardant en mémoire que ces fréquences se répètent tout au long de l'axe des fréquences.

La Figure 5-1 représente les 4 types de filtres (en considérant  $T$  égale à 1).

Pratiquement, les filtres idéaux ne peuvent être réalisés que d'une manière approximative. C'est pourquoi, dans des cas pratiques, les spécifications sont données avec des tolérances qui constituent ce qu'on appelle gabarit de la réponse fréquentielle du filtre.

Ce "gabarit" porte sur le module d'un filtre  $|H(f)|$ . Il peut être défini par les trois paramètres suivants:

- L'erreur d'approximation maximale sur l'atténuation dans la bande bloquante  $\delta_1$  (au lieu d'un gain nul dans cette bande)
- L'erreur d'approximation dans la bande passante  $\pm \delta_2$  (au lieu d'un gain égal à 1 dans cette bande): Le gain varie alors entre  $1-\delta_2$  et  $1+\delta_2$ .
- La largeur de la bande de transition dans laquelle  $|H(f)|$  passe de la bande passante à la bande bloquante ou vice-versa.

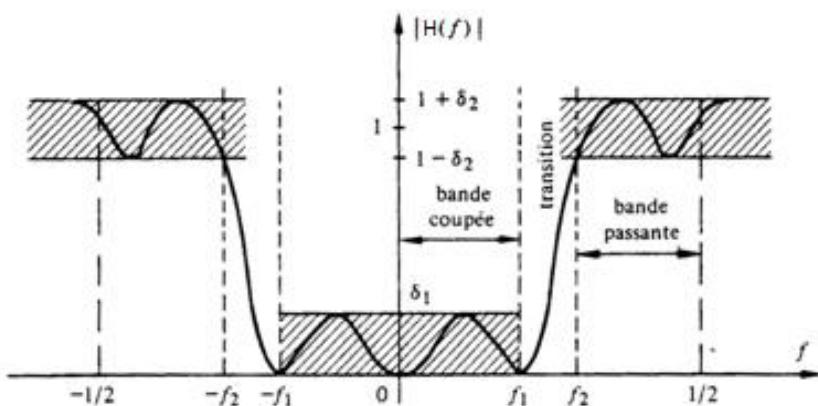


Figure 5-2. Gabarit d'un filtre

Plus ces trois paramètres s'approchent de zéro, plus la sélectivité du filtre s'approche de l'idéale. La Figure 5-2 présente un tel gabarit pour un filtre passe-haut.

La plupart des filtres pratiques sont spécifiés sur le module sans faire intervenir la phase qui est généralement spécifiée par des contraintes de stabilité et de causalité. Ces contraintes découlent des restrictions posées sur la fonction de transfert  $H(z)$  qui doit avoir ses pôles à l'intérieur du cercle unité.

Une fois le gabarit est fixé, l'étape suivante consiste à trouver le système linéaire dont la réponse satisfait ces conditions. Deux approches sont possibles:

- Développer des méthodes d'approximation en se basant sur les mathématiques appliquées.
- Utiliser les méthodes déjà utilisées en analogique pour la synthèse des filtres. Dans ce cas, on peut souvent utiliser des relations analytiques pré-établies et faciles à manipuler. Ceci permet en même temps de pouvoir simuler numériquement les filtres analogiques.

## 5.2 Filtres à Réponse Impulsionnelle de durée Finie (RIF):

Ces sont des systèmes linéaires discrets invariants dans le temps (coefficients indépendants du temps) non récursifs (un échantillon du signal filtré est obtenu par une sommation pondérée d'un ensemble fini des échantillons du signal à filtrer). Les coefficients de la sommation pondérée constituent la réponse impulsionnelle du filtre.

### 5.2.1 Réalisations

La mise en œuvre de l'opération de filtrage, peut être effectuée:

- soit par une réalisation transversale ou non récursive (Figure 5-3) en appliquant la relation de

convolution suivante: 
$$y[n] = \sum_{l=n_0}^{n_0+L-1} h[l].x[n-l]$$

Le filtre RIF (ou FIR de sa nomination anglaise Finite Impulse Response) est dit "à mémoire finie": sa sortie ne dépend que d'un nombre fini d'échantillons d'entrée d'ancienneté limitée. On note que le nombre d'éléments mémoire nécessaires est égal à (L-1). Lorsqu'une réalisation utilise un nombre d'éléments mémoire supérieur ou égal à L, le système est dit non-canonical.

D'une façon générale, un filtre discret est dit "Canonique" s'il contient le nombre minimal d'éléments de retard nécessaires pour réaliser son équation aux différences.

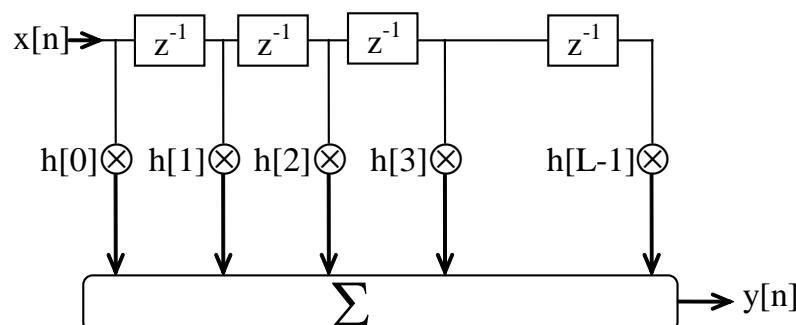


Figure 5-3. Réalisation transversale ( $n_0 = 0$  ici).

Un tel filtre est toujours stable pour autant que les valeurs de  $h[n]$  soient toutes finies. Sa fonction de transfert du filtre  $H(f)$  est donnée par:

Éq 5-2 
$$H(f) = \sum_{n=n_0}^{n_0+L-1} h[n].e^{-j2\pi fn}.$$

Notons que  $H(z)$  correspondante ne possède que des zéros (et un nombre identique des pôles à l'origine).

- soit par TFD: en multipliant la TFD du signal d'entrée  $X(k)$  par celle du filtre  $H(k)$ , on obtient la TFD du signal filtré  $Y(k)$  qui, par TFD inverse, donne les échantillons  $y[n]$  du signal de sortie (figure 9).

si  $N > L$ , alors il faut compléter le filtre  $h$  par  $N - L$  échantillons de valeur nulle pour pouvoir calculer les TFD de  $x$  et  $h$  sur  $N$  points et donc obtenir à la fin un signal filtré de  $N$  échantillons.

Du point de vue nombre d'opérations, la réalisation par la TFD à l'aide de la TFR (FFT) est plus

avantageuse que la réalisation non récursive si la longueur L de la réponse impulsionnelle du filtre dépasse les 30 coefficients.

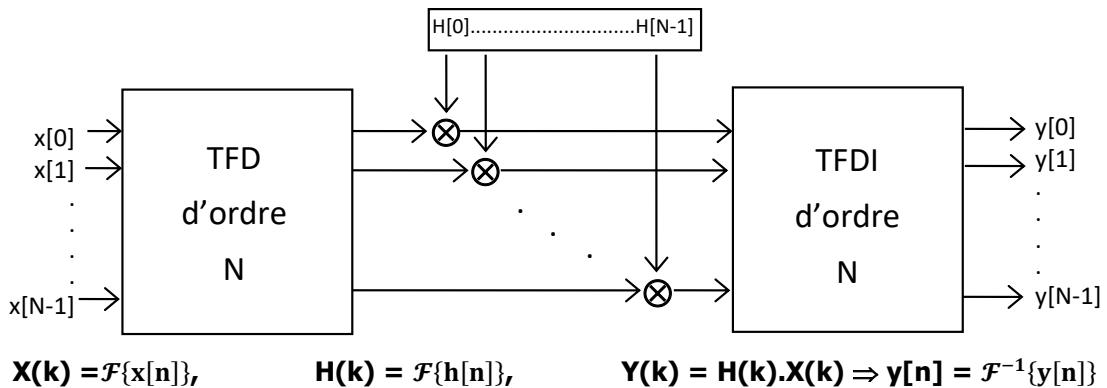


Figure 5-4. Réalisation par TFD

### Déférence entre les deux réalisations :

#### *Cas des signaux de longueur finie*

Le produit de convolution est une opération très importante en traitement des signaux, aussi bien sur le plan théorique que sur le plan de réalisation pratique de certains systèmes linéaires.

On peut montrer que, lorsque cette opération est effectuée entre un signal  $x$  de durée finie ( $N$  échantillons) et un filtre de longueur  $L$ , le signal  $y$  résultant sera constitué de  $N+L-1$  échantillons, ce qui veut dire que le signal de sortie possède  $L-1$  échantillons supplémentaires par rapport au signal d'entrée.

Or la réalisation par TFD, telle qu'elle est présentée ci-dessus, fournit à la sortie un signal de  $N$  échantillons seulement, ce qui montre que les deux réalisations ne sont pas équivalentes.

En effet, les TFD de  $x$  et  $h$  sur  $N$  points permettent d'avoir une TFD de  $y$  sur  $N$  points et donc d'obtenir à la fin un signal filtré  $y[n]$  de  $N$  échantillons et non pas de  $L+N-1$  échantillons.

Ceci s'explique par le fait que le calcul des TFDs suppose que les signaux concernés sont périodiques et de période  $N$ , ce qui sera aussi le cas de  $y[n]$ :  $y[n]$  est donc le résultat d'une convolution dite "circulaire" qui consiste, dans le domaine temporel, à périodiser le signal à filtrer avant de procéder à la convolution.

La convolution qui ne fait pas impliquer des signaux périodisés est appelée convolution linéaire.

Les deux réalisations "transversale" et "par TFD" peuvent devenir équivalentes lorsqu'on effectue la TFD sur  $N+L-1$  points au lieu de  $N$  et ce, en ajoutant  $L-1$  échantillons nuls au signal d'entrée  $x$  et  $N-1$  points nuls à la réponse impulsionnelle  $h$ . Le résultat des produits des 2 TFD sera alors une TFD sur  $N+L-1$  points dont la transformée inverse donne le signal  $y[n]$  de  $N+L-1$  échantillons qui sera identique à celui obtenu par une simple convolution linéaire.

#### *Cas des signaux de longueur très grande (voire infinie)*

En pratique, il n'est pas rare de vouloir calculer le produit de convolution linéaire de deux signaux dont l'un a une durée finie  $L$  et l'autre d'une durée très longue ("infinie"). Cette situation se présente si on veut, par exemple, filtrer une version échantillonnée  $x[n]$  d'un signal de parole avec

un filtre de réponse impulsionnelle  $h[n]$  de longueur  $L$ . Le signal résultant sera alors :

$$\text{Éq 5-3} \quad y[n] = \sum_{l=0}^{+\infty} x[l]h[n-l]$$

Bien évidemment, on peut obtenir, par "calcul direct",  $y[n]$  pour tout  $n$ , en utilisant l'Éq 5-3, puisque le calcul, pour chaque point d'indice  $n$ , n'implique que  $L$  échantillons du signal  $x$  seulement.

Cependant, il est tout à fait possible d'effectuer cette opération de filtrage en utilisant la TFD. À première vue, la longueur "infinie" du signal  $x$  pose des problèmes pour le calcul de sa TFD. Il est évident qu'on ne peut pas calculer d'un seul coup la TFD du signal: d'une part, la détermination de cette TFD demanderait beaucoup de calcul (et donc une grande capacité mémoire) et d'autre part, il faudrait attendre que la totalité du signal soit à notre disposition avant de démarrer le calcul de  $y[n]$  ce qui introduirait un temps de retard énorme.

La solution à ce problème consiste à procéder par partie en considérant le signal  $x$  comme une succession de plusieurs signaux  $x_1[n]$ ,  $x_2[n]$ ,  $x_3[n]$ , ... de durée finie  $N_s$ , juxtaposés et disjoints, tel que :

$$x[n] = x_1[n] + x_2[n] + x_3[n] + \dots \quad \text{pour } n=0, \dots, +\infty$$

avec

$$\text{Éq 5-4} \quad x_i[n] = \begin{cases} x[n] & \text{pour } n = (i-1)N_s, \dots, i.N_s - 1 \\ 0 & \text{ailleurs} \end{cases}$$

ce qui permet d'écrire l'Éq 5-3 comme suit :

Éq 5-5

$$y[n] = \sum_{l=0}^{+\infty} x_1[l].h[n-l] + \sum_{l=0}^{+\infty} x_2[l].h[n-l] + \sum_{l=0}^{+\infty} x_3[l].h[n-l] + \dots = y_1[n] + y_2[n] + y_3[n] + \dots$$

où  $y_i[n]$  est le produit de convolution linéaire de  $x_i[n]$  avec  $h[n]$  et dont la longueur est  $N_s+L-1$ .

Ainsi, les  $y_i[n]$  peuvent être calculés séparément en utilisant la TFD sur  $N_s+L-1$ , puis additionnés les uns avec les autres pour donner  $y[n]$ .

Il est normal que les  $y_i[n]$  adjacents se recouvrent et s'additionnent sur  $L-1$  échantillons. C'est la raison pour laquelle cette méthode est appelée méthode d'addition-recouvrement.

## 5.2.2 Filtres à phase linéaire

Une propriété importante des filtres RIF est qu'ils peuvent donner des filtres dont la caractéristique de phase est exactement linéaire.

Un filtre est dit à phase linéaire lorsque le temps de propagation de groupe  $\tau_g$  est constant.  $\tau_g$  est défini par l'opposé de la dérivée de la phase du filtre  $\varphi(\omega)$  par rapport à  $\omega$  :

$$\text{Éq 5-6} \quad \tau_g(\omega) = -\frac{d\varphi(\omega)}{d\omega}$$

Un temps de propagation de groupe constant signifie que toutes les composantes fréquentielles constituant le signal à l'entrée du filtre (ou du système en général) sont affectées par

le même retard lors de leur passage dans le filtre, ce qui préserve la forme du signal en fonction du temps. Ceci revêt une grande importance dans des domaines où il existe des spécifications particulières à ce sujet (comme en télévision et en transmission numérique).

Cela veut dire aussi que la phase varie linéairement (avec une pente constante) avec la possibilité d'avoir des sauts de  $\pi$  radians. Des tels sauts se produisent uniquement lorsque la fonction de transfert possède des zéros sur le cercle unité dans le plan des  $z$ . Aux fréquences où se produisent ces sauts de phase, la réponse en fréquence a une amplitude nulle et le fait que  $\tau_g(\omega)$  ne soit pas défini à ces fréquences n'a pas d'importance.

Donc, il est toujours intéressant de ne pas modifier la phase du signal filtré par rapport à celle du signal d'entrée; or ceci n'est pas toujours possible pour des raisons de stabilité et/ou de causalité.

Reprendons, par exemple, l'expression de l'Éq 5-2 donnant  $H(f)$  en fonction de  $h[n]$ ; en séparant la partie réelle et la partie imaginaire, on obtient:

$$\text{Éq 5-7} \quad H(f) = \sum_{n=n_0}^{n_0+L-1} h[n] \cos(2\pi f n) - j \sum_{n=n_0}^{n_0+L-1} h[n] \sin(2\pi f n)$$

Pour enlever toute modification de la phase entre le signal de sortie et celui d'entrée, il faut que  $\text{Im}\{H(f)\}$  soit identiquement nulle. Ceci n'est possible que si la réponse impulsionnelle  $h[n]$  est une fonction paire autour de l'origine, puisque, dans ce cas, les termes  $h[n] \sin(2\pi f n)$  et  $h[-n] \sin(-2\pi f n)$  s'annulent deux à deux; d'où une partie imaginaire nulle de  $H(f)$ . Cette condition se traduit par:

$$\text{Éq 5-8} \quad h[n] = h[-n] \quad n_0 \leq n \leq n_0 + L - 1$$

Pour satisfaire la condition de symétrie autour de  $n = 0$ , il faut que  $n$  soit compris entre  $-(L-1)/2$  et  $(L-1)/2$  soit,

$$\text{Éq 5-9} \quad h[n] = h[-n] \quad |n| \leq (L-1)/2$$

**Si  $L$  est impair, alors  $(L-1)/2$  est un entier et  $n_0$  doit être égal à  $-(L-1)/2$ . Dans ce cas, le nombre d'échantillons non nuls de  $h[n]$  d'indice positif est égal au nombre d'échantillons non nuls de  $h[n]$  d'indice négatif, et comme la fonction cosinus est paire aussi, l'expression de  $H(f)$  devient alors:**

$$\text{Éq 5-10} \quad H(f) = h[0] + 2 \sum_{n=1}^{(L-1)/2} h[n] \cos(2\pi f n)$$

**Si  $L$  est pair, la condition de symétrie de l'Éq 5-9 reste valable mais  $(L-1)/2$  n'est plus un entier, et  $n_0$  doit être égal à l'entier supérieur le plus proche de  $-(L-1)/2$  qui est  $-(L-2)/2$ . Dans ce cas, un déséquilibre apparaît entre le nombre d'échantillons d'indice positif et le nombre d'échantillons d'indice négatif, et la symétrie n'est plus assurée autour de  $n = 0$  mais elle peut l'être autour de  $n = \frac{1}{2}$ . Pour ramener la symétrie autour de l'origine, on considère le filtre  $h_1[n]$  déduit de  $h[n]$  par un décalage de  $-\frac{1}{2}$ :**

$$\text{Eq 5-11} \quad h_1[n - \frac{L}{2}] = h[n] \quad \text{pour } -\frac{L}{2} + 1 \leq n \leq \frac{L}{2}$$

Les TFD des deux termes, en se servant de la propriété de translation, s'écrivent:

$$\text{Eq 5-12} \quad H_1(f) \cdot e^{-j2\pi f(\frac{1}{2})} = H(f) \Rightarrow H_1(f) = H(f) \cdot e^{j\pi f} = 2 \sum_{n=1}^{L/2} h[n] \cdot \cos(2\pi f(n - \frac{1}{2}))$$

D'où un filtre  $h_1[n]$  satisfaisant la relation de symétrie et possédant une phase identiquement nulle.

Les filtres ainsi obtenus (que ce soit pour  $L$  pair ou impair) sont des filtres non causaux car la moitié de leur réponse impulsionnelle se trouve dans la partie négative de l'axe des  $n$ . Pour rendre  $h[n]$  (ou  $h_1[n]$ ) causal tout en conservant la symétrie, on décale la réponse impulsionnelle de  $(L-1)/2$  pas vers les valeurs positives de l'axe des  $n$ , de manière à ramener à l'origine la première valeur non nulle de  $h[n]$ .

En vertu du théorème de retard, ce décalage correspond dans le domaine fréquentiel au produit de  $H(f)$  par le terme  $e^{-j2\pi f(L-1)/2}$ .

Comme  $H(f)$  est réelle, la caractéristique de phase du filtre après décalage est donnée par:

$$\text{Eq 5-13} \quad \varphi_h(f) = -(L-1)\pi f$$

qui est une fonction linéaire en  $f$ ; d'où la nomination du "filtre à phase linéaire".

Après décalage, la symétrie peut s'écrire de la manière suivante:

$$\text{Eq 5-14} \quad h[n] = h[L-1-n] \quad n = 0, \dots, L-1$$

Ceci montre la possibilité d'élaborer tout filtre à phase linéaire par synthèse d'un filtre à réponse impulsionnelle finie (FIR). Ainsi, un signal se trouvant dans la bande passante de ce filtre est reproduit exactement à la sortie, avec un retard donné par la pente de la réponse fréquentielle de phase.

Remarquons que lors du filtrage par un filtre à phase linéaire, on peut éviter la moitié des multiplications en utilisant la symétrie ce qui représente un gain considérable et une propriété importante pour ce type de filtre.

### 5.2.3 Exemple des Filtres en peigne

Un type de filtre intéressant est connu sous le nom de "filtre en peigne", qui peut être obtenu à partir de n'importe quel filtre discret de fonction de transfert  $H(z)$  en remplaçant  $z$  par  $z^N$ . Ceci se traduit, dans la réalisation par remplacer chaque élément de mémoire (retardateur) par  $N$  éléments en cascade, ce qui donne lieu à un filtre dont la fonction de transfert  $G(z)$  est égale à  $H(z^N)$ .

La réponse en fréquence du filtre résultant est une répétition de  $N$  fois de la réponse fréquentielle du filtre de départ  $H(e^{j\omega})$  dans l'intervalle principal  $[-\pi, +\pi[$ .

Prenons l'exemple d'un filtre passe-haut de fonction de transfert  $H(z)=1-z^{-1}$  dont la réponse fréquentielle et la répartition des pôles et des zéros sont données par la Figure 5-5. Celles du filtre en peigne résultant pour  $N=2$ ,  $N=10$  et  $N=20$  sont données respectivement par la Figure 5-6, la Figure 5-7 et la Figure 5-8. Sa structure transversale est donnée par la Figure 5-9.

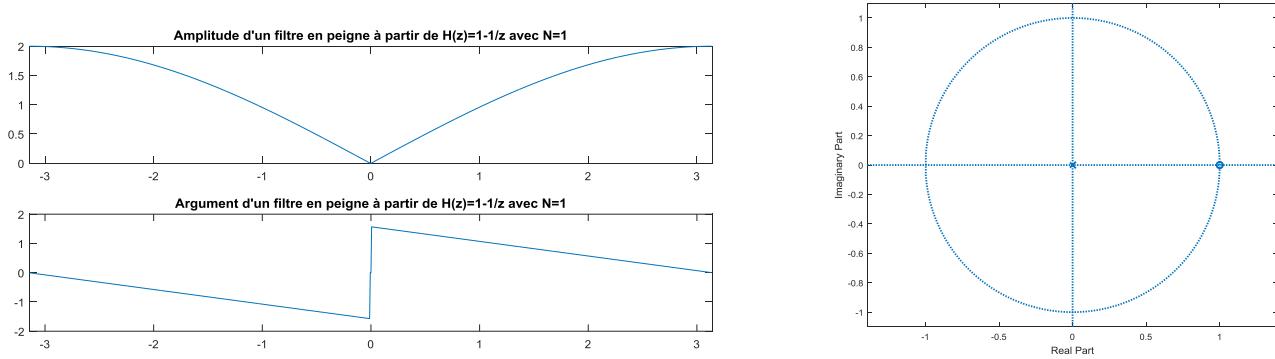


Figure 5-5. Réponse fréquentielle et Répartition des pôles et des zéros de  $H(z)=1-z^{-1}$ .

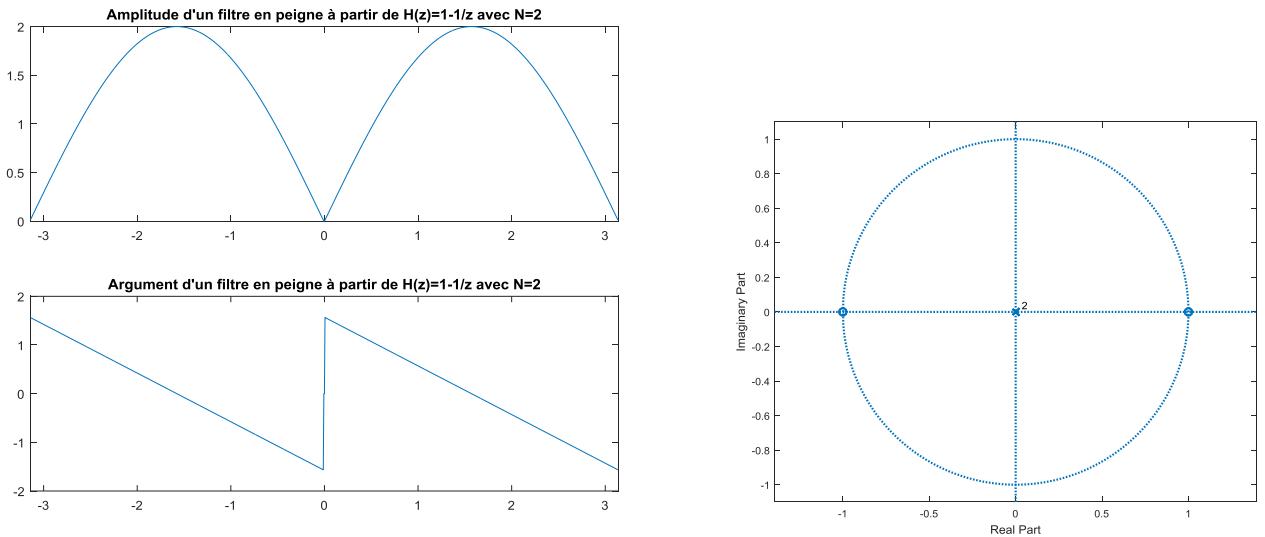


Figure 5-6. Réponse fréquentielle et Répartition des pôles et des zéros d'un filtre en peigne  $G(z)=1-z^{-2}$ .

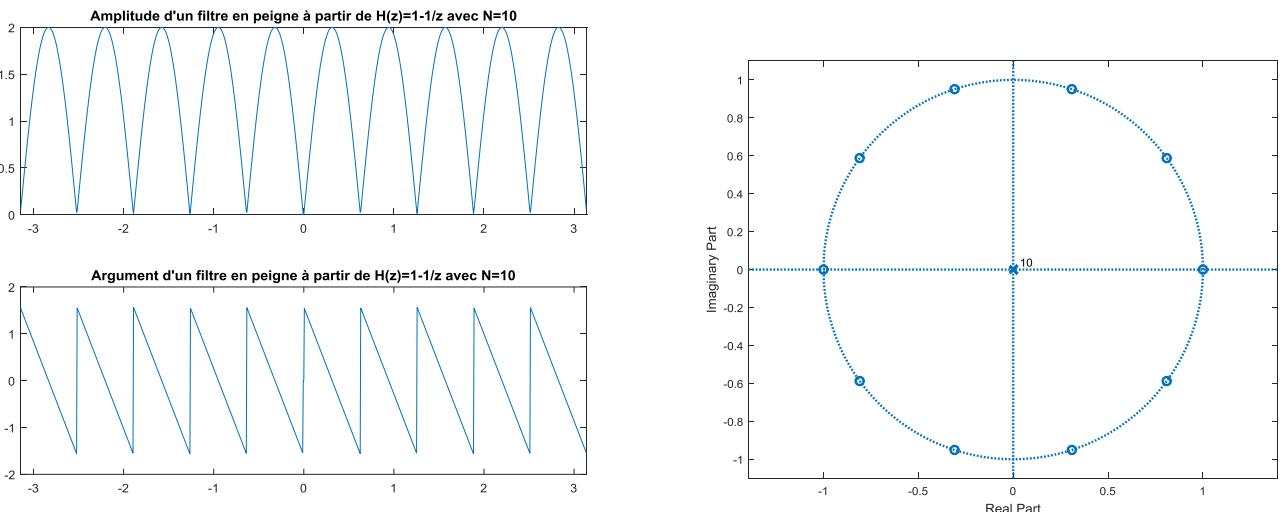


Figure 5-7. Réponse fréquentielle et Répartition des pôles et des zéros d'un filtre en peigne  $G(z)=1-z^{-10}$ .

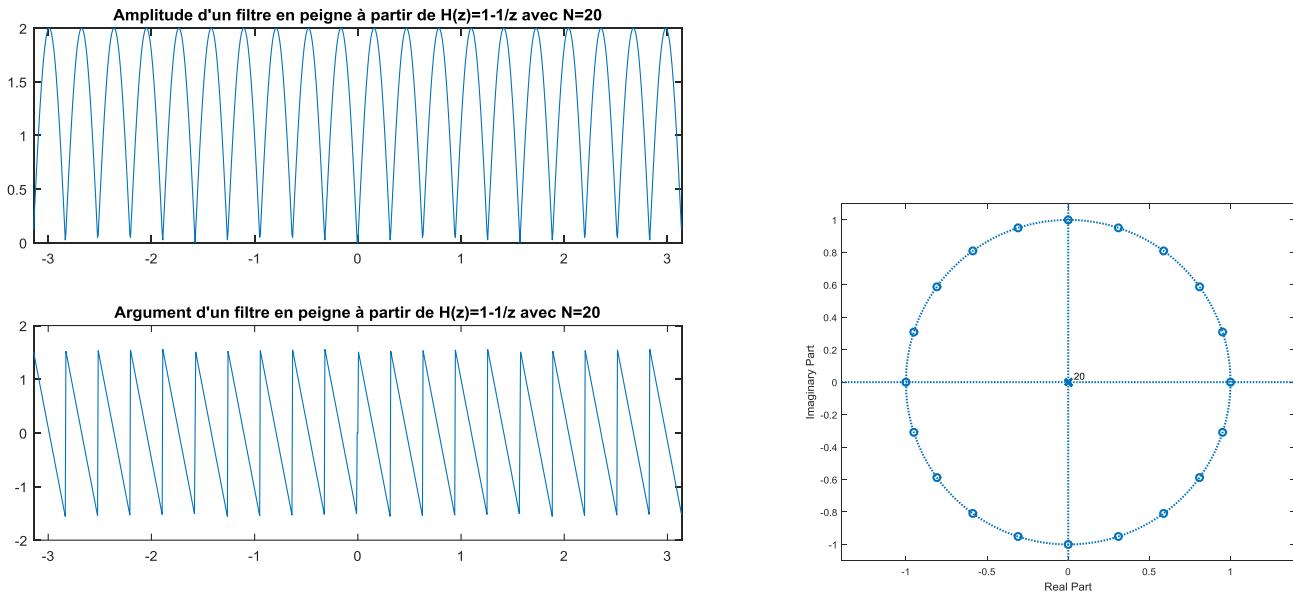


Figure 5-8. Réponse fréquentielle et Répartition des pôles et des zéros d'un filtre en peigne  $G(z)=1-z^{-20}$ .

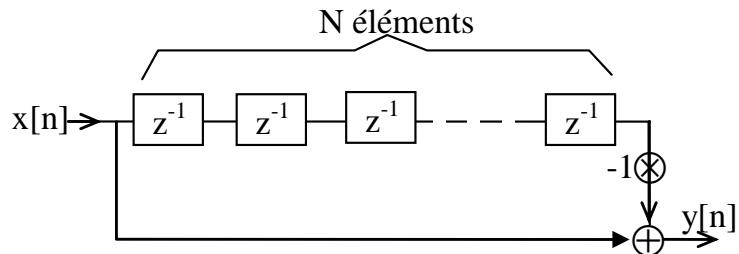


Figure 5-9. Structure transversale d'un filtre en peigne de la forme  $G(z)=1-z^{-N}$ .

### 5.3 Filtres à Réponse Impulsionnelle de durée Infinie (RII):

Ces sont des systèmes linéaires discrets invariants dans le temps dont le fonctionnement est régi par une équation de convolution portant sur une infinité de termes.

En principe, ils conservent une trace des signaux qui leur ont été appliqués pendant une durée infinie: ils sont dits "à mémoire infinie". Une telle mémoire est réalisée par une boucle de réaction de la sortie vers l'entrée, d'où la dénomination courante de filtre récursif.

Chaque élément de la suite des nombres de sortie est calculé par une sommation pondérée d'un certain nombre d'éléments de la suite d'entrée et d'un certain nombre d'éléments précédents de la suite de sortie.

Le fait d'avoir cette réponse impulsionnelle infinie permet d'obtenir en général un filtrage beaucoup plus sélectif que celui d'un filtre RIF à quantité de calculs équivalente. Cependant la boucle de réaction complique l'étude des propriétés et la conception de ces filtres et amène des phénomènes parasites.

L'équation générale qui gère un filtre RII est donnée par l'équation aux différences donnée par l'Eq 4-6 et rappelée ici:

$$y[n] + a_1 \cdot y[n-1] + a_2 \cdot y[n-2] + \dots + a_N \cdot y[n-N] \\ = b_0 \cdot x[n] + b_1 \cdot x[n-1] + \dots + b_M \cdot x[n-M]$$

soit:

$$y[n] = \sum_{i=0}^M b_i \cdot x[n-i] - \sum_{i=1}^N a_i \cdot y[n-i]$$

dont la fonction de transfert en z:

$$\text{Eq 5-15} \quad H(z) = \frac{\sum_{m=0}^M b_m z^{-m}}{1 + \sum_{n=1}^N a_n z^{-n}}$$

C'est le quotient de deux polynômes en z, qui sont souvent tels que  $M \leq N$ . Les coefficients  $a_n$  et  $b_m$  sont réels et  $H(z)$  est un nombre complexe tel que:  $H^*(z) = H(z^*)$  et la réponse en fréquence du filtre s'écrit:  $H(\omega) = [H(z)]_{z=e^{j\omega}} = |H(\omega)| \cdot e^{-j\varphi(\omega)}$  avec  $|H(\omega)|^2 = [H(z) \cdot H(z^{-1})] \Big|_{z=e^{j\omega}}$  et,  $\varphi(\omega) =$

$$\frac{j}{2} \ln \left( \frac{H(z)}{H(z^{-1})} \Big|_{z=e^{j\omega}} \right)$$

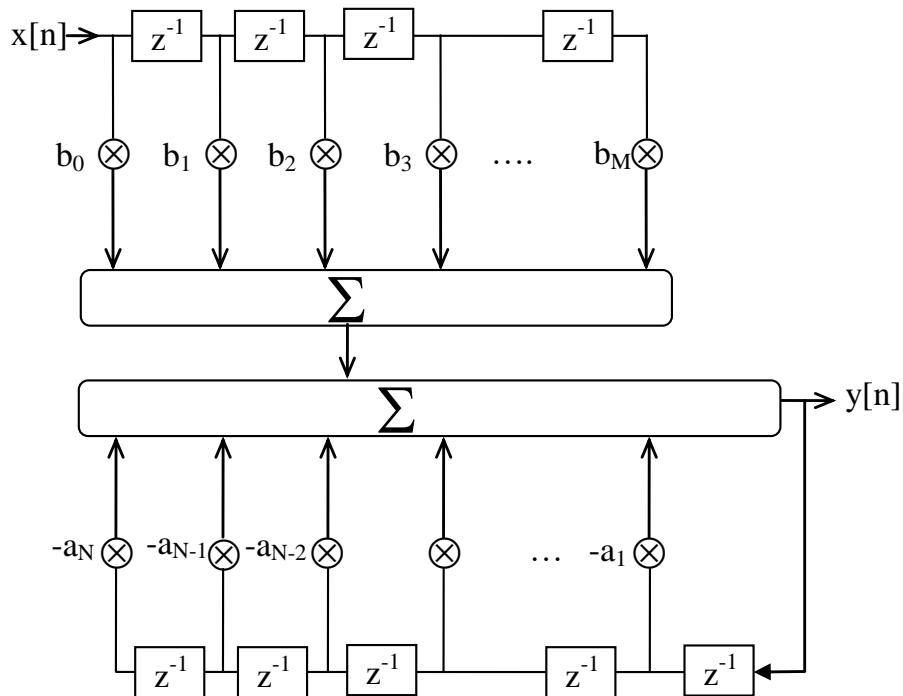


Figure 5-10. Réalisation récursive de forme directe 1.

Le filtre est donc complètement défini par les coefficients  $a_n$  et  $b_m$ . L'opération de filtrage est réalisée par une structure dite récursive dont le schéma bloc peut être celui de la Figure 5-10.

Cette structure de réalisation est dite directe (forme 1) car elle traduit directement l'équation aux différences établi ci-dessus. On remarque qu'elle contient  $M+N$  éléments de mémoire et  $M+N+1$  multiplicateurs.  $M+N+1$  valeurs doivent être additionnées ensemble dans les additionneurs.

Cette forme 1 est non-canonical parce que le nombre d'éléments de mémoire peut être réduit. C'est ce qui est réalisé par la forme directe 2.

En effet, la structure de la forme 1 peut être vue comme étant la cascade de deux systèmes :

- L'un est constitué par la partie non-réursive dont l'entrée est  $x[n]$  et la fonction de transfert est le numérateur de  $H(z)$  de l'Éq 5-15
- L'autre est constitué par la partie récursive dont l'entrée est la sortie du 1<sup>er</sup> système et la fonction de transfert est l'inverse du dénominateur de  $H(z)$ .

Puisque la mise en cascade peut être inter-changée sans modifier le filtre, on peut alors avoir la forme de la Figure 5-11. Dans cette nouvelle forme on remarque que les  $M$  éléments mémoire qui mémorisent les échantillons  $v[n]$ ,  $v[n-1]$ , ...  $v[n-M+1]$  peuvent être utilisés une seule fois pour les deux parties ce qui donne la forme directe 2 de la Figure 5-12 qui représente un système canonique avec le minimum d'éléments mémoire nécessaires.

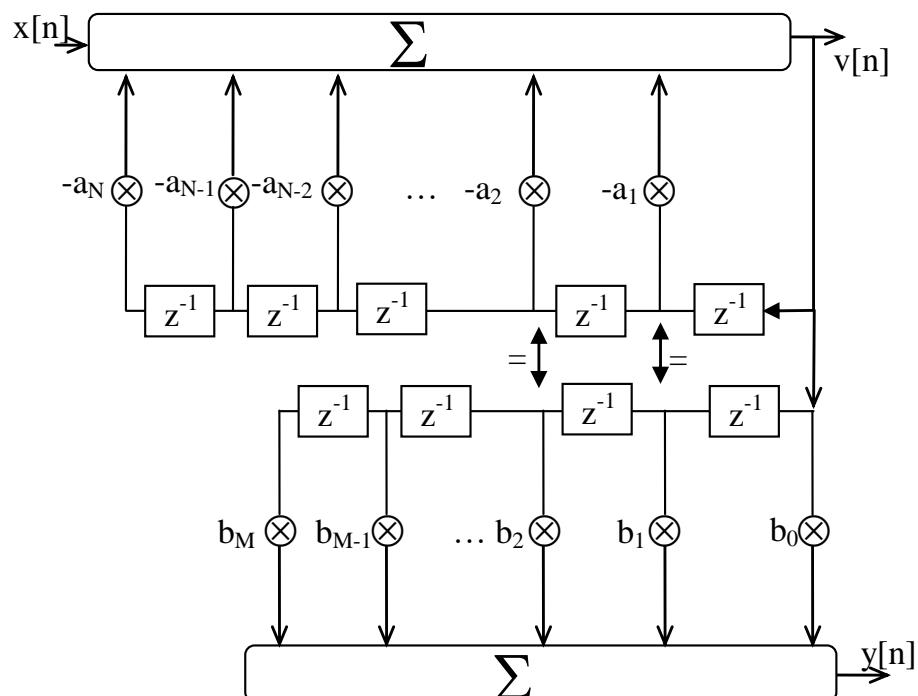


Figure 5-11. Réalisation récursive obtenue en inter-changeant la partie récursive et non-réursive de forme directe 1.

Un inconvénient majeur des systèmes discrets est leur forte sensibilité aux valeurs des coefficients  $b_m$  et  $a_n$ . En effet, une faible variation dans l'un des coefficients  $b_m$  affecte la localisation des  $M$  zéros du système alors qu'une faible variation dans l'un des coefficients  $a_n$  affecte la localisation des  $N$  pôles du système. Cela signifie que la réponse en fréquence dans son intégralité peut varier considérablement.

Pour réduire cet impact, surtout pour des grandes valeurs de  $M$  et  $N$ , on divise la fonction de transfert  $H(z)$  en  $K$  fonctions de transfert d'ordre inférieur présentant chacune un nombre limité de pôles et de zéros. Ainsi, le système sera constitué de plusieurs sous-systèmes mis soit en cascade soit en parallèle.

## Structure en cascade

Pour réaliser un système de fonction de transfert  $H(z)$  en cascade, il faut mettre  $H(z)$  sous la forme d'un produit de  $K$  facteurs  $H_i(z)$  ( $i=1,..,K$ ):  $H(z) = H_1(z) \cdot H_2(z) \cdot H_3(z) \dots \cdot H_K(z)$

Rappelons que pour des systèmes à réponse impulsionnelle réelle, les pôles et les zéros sont soit réels soit complexes conjugués deux à deux.

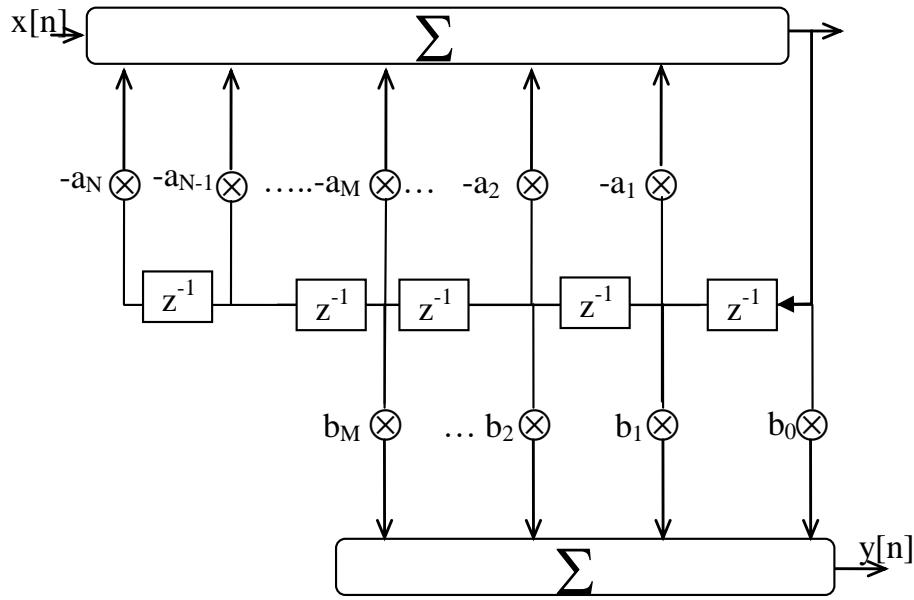


Figure 5-12. Réalisation récursive et canonique de forme directe 2.

Ceci permet alors de montrer que  $H_i(z)$  peut avoir l'une deux formes suivantes:

Éq 5-16

$$H_i(z) = \frac{c_i + d_i \cdot z^{-1}}{1 + e_i \cdot z^{-1}}$$

ou bien,

Éq 5-17

$$H_i(z) = \frac{c_i + d_i \cdot z^{-1} + e_i \cdot z^{-2}}{1 + f_i \cdot z^{-1} + g_i \cdot z^{-2}}$$

Pour la 1<sup>ère</sup> forme (du 1<sup>er</sup> ordre),  $H_i(z)$  possède au plus un zéro et un pôle tous les deux réels, alors  $H_i(z)$  de la 2<sup>nde</sup> forme (du 2<sup>nd</sup> ordre) possède deux pôles complexes conjugués et, au plus, deux zéros réels ou complexes conjugués.

**Exemple :** Soit

$$\begin{aligned} H(z) &= \frac{23 + 40z^{-1} + 36z^{-2} + 19z^{-3}}{10 + 9z^{-1} + 8z^{-2} + 3z^{-3}} = \frac{(1 + z^{-1}) \cdot (23 + 17z^{-1} + 19z^{-2})}{(2 + z^{-1}) \cdot (5 + 2z^{-1} + 3z^{-2})} \\ &= \frac{1 + z^{-1}}{1 + 0.5z^{-1}} \times \frac{2.3 + 1.7z^{-1} + 1.9z^{-2}}{1 + 0.4z^{-1} + 0.6z^{-2}} \end{aligned}$$

Ce système peut être mis en cascade avec une réalisation canonique des deux sous-systèmes comme le montre la Figure 5-14.

Notons que, la décomposition de  $H(z)$  en des facteurs  $H_i(z)$  se fait librement en choisissant de regrouper les pôles et les zéros des sous-fonctions comme bon il nous semble.

### Structure en parallèle

Pour réaliser un système de fonction de transfert  $H(z)$  en parallèle (Figure 5-13), il faut mettre  $H(z)$  sous la forme d'une somme de  $K$  facteurs  $H_i(z)$  ( $i=1,..,K$ ):  $H(z) = H_0 + H_1(z) + H_2(z) + H_3(z) + \dots + H_K(z)$  où  $H_0$  est une constante et  $H_i(z)$  sont des facteurs du 1<sup>er</sup> ou 2<sup>nd</sup> ordre de la forme :

Éq 5-18

$$H_i(z) = \frac{c_i}{1 + d_i \cdot z^{-1}}$$

ou bien,

Eq 5-19

$$H_i(z) = \frac{c_i + d_i \cdot z^{-1}}{1 + e_i \cdot z^{-1} + f_i \cdot z^{-2}}$$

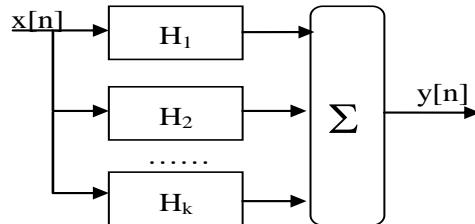


Figure 5-13. Mise en parallèle d'un système discret

L'ensemble des pôles des  $H_i(z)$  représente tous les pôles de  $H(z)$  alors que leurs zéros ne sont pas les mêmes.

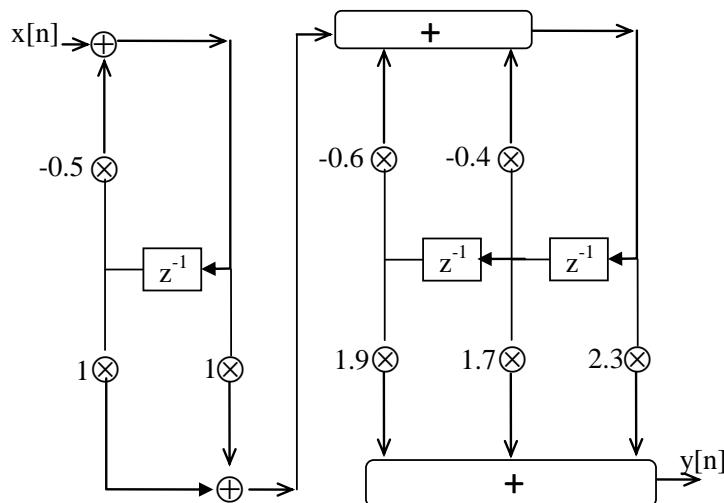


Figure 5-14. Réalisation canonique d'un système en cascade formé d'un sous-système d'ordre 1 et d'un autre d'ordre 2.

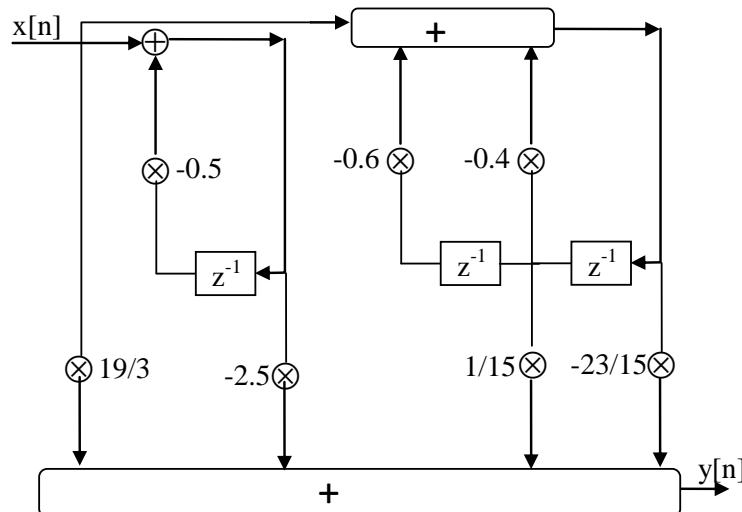


Figure 5-15. Réalisation canonique d'un système en parallèle formé d'un sous-système d'ordre 1 et d'un autre d'ordre 2.

**Exemple :** Soit  $H(z)$  de l'exemple précédent dont la décomposition en facteurs simples peut se faire en utilisant la méthode décrite dans le paragraphe 4.3.4 du chapitre précédent:

$$\begin{aligned}
 H(z) &= \frac{23 + 40z^{-1} + 36z^{-2} + 19z^{-3}}{10 + 9z^{-1} + 8z^{-2} + 3z^{-3}} = H_0 + \frac{c_1}{1 + 0.5z^{-1}} + \frac{c_2 + c_3z^{-1}}{1 + 0.4z^{-1} + 0.6z^{-2}} \\
 &= \frac{19}{3} + \frac{-2.5}{1 + 0.5z^{-1}} + \frac{-\frac{23}{15} + \frac{1}{15}z^{-1}}{1 + 0.4z^{-1} + 0.6z^{-2}}
 \end{aligned}$$

Ce système peut être mis en parallèle avec une réalisation canonique des deux sous-systèmes comme le montre la Figure 5-15.

## 5.4 Le principe de transposition

Une méthode générale d'obtention, à partir de tout filtre discret, d'un autre filtre possédant exactement la même réponse en fréquence, est basée sur le "théorème de transposition". Ce théorème montre que la réponse fréquentielle d'un système discret linéaire et invariant dans le temps est invariante par transposition qui consiste à :

- Remplacer les additionneurs par des nœuds et les nœuds par des additionneurs
- Inverser le sens de circulation (l'entrée devient sortie et la sortie entrée).

Un exemple est donné par la Figure 5-16.

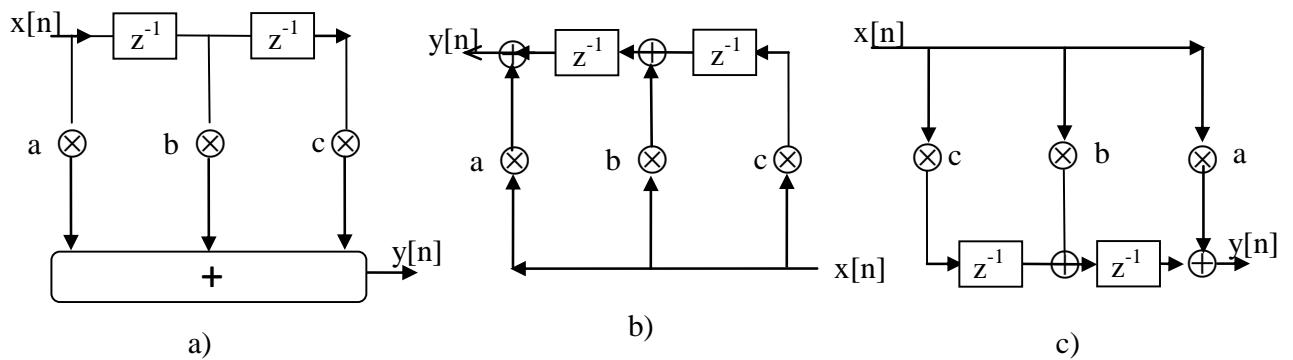


Figure 5-16. a) Filtre transversal simple b) Version transposée de ce filtre. c) Version transposée avec l'entrée à gauche.

## 5.5 Le principe du filtrage adaptatif

Le filtrage adaptatif consiste à adapter les valeurs des coefficients du filtre en fonction d'un ou plusieurs critères fixées à priori. Pour ce faire, le filtre adaptatif comporte deux parties distinctes :

- Le filtre proprement dit (la structure classique du filtre dans laquelle les coefficients modifiables sont notés  $c_i[n]$ , qui reflètent leurs valeurs fixés à l'instant  $n$ )
- Une unité de commande dont le rôle est de calculer et mettre à jour les  $c_i[n]$  à chaque instant, en accord avec les critères de commande préétablis (habituellement basés sur la minimisation de la différence entre le signal de sortie et un signal de référence).

Un exemple d'utilisation de ce type de filtrage est celui utilisé en télévision pour la réception de la télédiffusion terrestre pour fournir un signal amélioré  $y[n]$  en fonction du signal reçu  $x[n]$  dégradé par différents facteurs dont les échos dus à des réflexions. Le signal de référence dans ce cas peut être le signal transmis durant le temps de retour-ligne et dont le récepteur connaît exactement à quoi il ressemble ce qui permet à l'unité de commande de savoir les effets de dégradation sur le signal reçu et donc de déterminer les meilleures valeurs des coefficients du filtre qui permettent de compenser partiellement ou totalement ces effets.

L'égalisation des effets des canaux de transmission de données en est un autre exemple. L'adaptation se fait en fonction des caractéristiques variantes du canal utilisé.

## 5.6 Synthèse des Filtres Numériques

La conception d'un filtre discret commence habituellement avec la spécification du comportement en fréquence requis. La forme de la caractéristique de phase est parfois laissée complètement non-spécifiée. Dans d'autres cas, on peut souhaiter une caractéristique de phase linéaire, par exemple.

À partir de ces spécifications, le processus de conception comprend les étapes suivantes:

- On décide si l'on veut approcher la caractéristique de fréquence désirée par un filtre RIF ou RII.
- On choisit l'ordre du filtre pour calculer ensuite les coefficients de la fonction de transfert de la meilleure façon possible (c'est là tout le problème).
- On valide le résultat obtenu en vérifiant qu'il répond aux spécifications de départ et en tenant compte de l'effet de quantification des coefficients calculés. Si le résultat n'est satisfaisant, on reprend la procédure en adoptant un choix différent de type de filtre, de son ordre et/ou de la précision de la quantification utilisés.

### Synthèse des Filtres RIF

Lorsqu'un filtre est spécifié dans le domaine fréquentiel par sa réponse  $H(f)$ , cette réponse doit être considérée comme périodique pour réaliser le filtrage par voie numérique. Dans l'intervalle principal, la réponse impulsionnelle est donnée par le développement en série de Fourier de la fonction périodique  $H(f)$ :

$$\text{Eq 5-20} \quad h[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f) e^{j2\pi f n} df$$

Cette synthèse par série de Fourier nécessite une expression analytique ou des caractéristiques correspondant aux filtres idéaux, ce qui limite l'emploi de cette méthode au cas des réponses fréquentielles de forme relativement simple pour lesquelles on peut évaluer l'intégrale analytiquement.

#### Cas où l'intégrale peut être évaluée analytiquement:

Dans ce cas,  $h[n]$  a une durée très grande (voire infinie) surtout lorsque les zones de transition sont idéales. De plus, la réponse impulsionnelle résultante est non causale ce qui rend le filtre obtenu non réalisable.

Pour le rendre réalisable, on devra d'une part limiter sa longueur à une valeur finie acceptable  $L$  et, d'autre part, introduire un décalage suffisant (un retard) pour obtenir une réponse impulsionnelle causale.

En effectuant ces deux ajustements, nous obtenons un filtre, de réponse impulsionnelle  $h[n]$ , directement réalisable qui est une approximation plus ou moins précise du filtre de départ souhaité.

En effet, la différence la plus importante par rapport à la caractéristique souhaitée se produit au voisinage des transitions raides. La troncature de la réponse impulsionnelle correspond à une

multiplication de la réponse impulsionnelle infinie par une fenêtre rectangulaire de largeur L. Ceci se traduit dans le domaine fréquentiel par un produit de convolution de la caractéristique initiale du filtre avec la transformée de Fourier de ce rectangle qui est une fonction "sinus cardinal" dont la largeur de la lobe principal est inversement proportionnelle à L.

Ceci se traduit par des oscillations dans la caractéristique fréquentielle résultante dont le nombre augmente avec L.

Pour réduire ces oscillations, on doit choisir une autre fonction-fenêtre dont la transformée en fréquence ait une lobe principale la plus étroite possible et des lobes secondaires qui contiennent aussi peu d'énergie que possible.

C'est le cas, par exemple, des fenêtres telles que celles de Hanning, de Hamming, de Bartlett ou de Blackman (Figure 5-17):

$$\text{Fenêtre de Hanning (ou Hann): } w[n] = \frac{1}{2} \left[ 1 - \cos \left( \frac{2n\pi}{L-1} \right) \right] \quad \text{pour } 0 \leq n \leq L-1$$

$$\text{Fenêtre de Hamming: } w[n] = 0.54 - 0.46 \cos \left( \frac{2n\pi}{L-1} \right) \quad \text{pour } 0 \leq n \leq L-1$$

$$\text{Fenêtre de Bartlett: } w[n] = \begin{cases} \frac{2n}{L-1} & \text{pour } 0 \leq n \leq \frac{L-1}{2} \\ 2 - \frac{2n}{L-1} & \text{pour } \frac{L-1}{2} < n \leq L-1 \end{cases}$$

$$\text{Fenêtre de Blackman: } w[n] = 0.42 - 0.5 \cos \left( \frac{2n\pi}{L-1} \right) + 0.08 \cos \left( \frac{4n\pi}{L-1} \right) \quad \text{pour } 0 \leq n \leq L-1$$

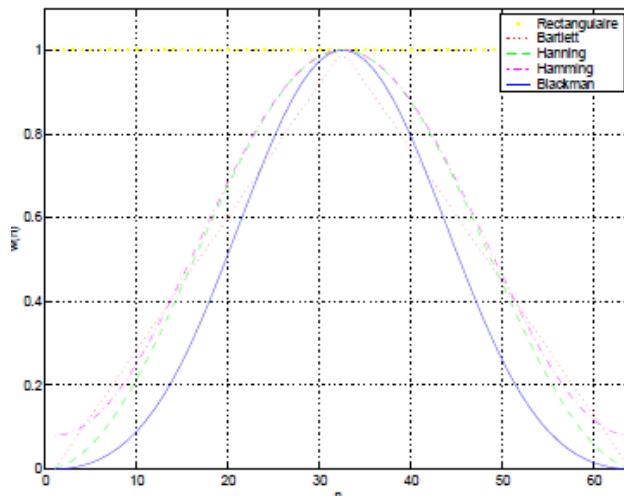
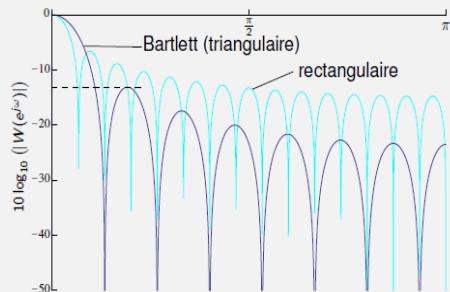


Figure 5-17. Réponses temporales des principales fenêtres

Ces fenêtres sont les principales fonctions utilisées en analyse spectrale et en synthèse de filtre RIF. Leurs réponses fréquentielles sont données à la Figure 5-18 avec L = 64. On voit que selon la fenêtre, la largeur du lobe principal et l'amplitude du plus grand lobe secondaire diffèrent. Par exemple, la fenêtre rectangulaire possède le lobe principal le plus étroit ( $4\pi/L$ ), mais le lobe secondaire le moins atténué (-13dB). Le Tableau 2 résume ces caractéristiques. On y voit en particulier qu'en utilisant une fenêtre sans discontinuité et plus lisse (comme celle de Hamming ou Blackman), on peut réduire de manière importante l'amplitude des lobes secondaires parasites.

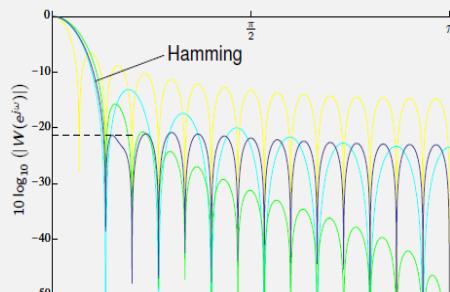
### Spectre – fenêtre de Bartlett

Lobe secondaire à -12.5 dB pour la fenêtre de Bartlett.



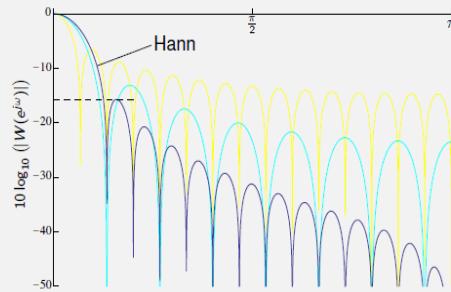
### Spectre – fenêtre de Hamming

Lobe secondaire à -20.5 dB pour la fenêtre de Hamming.



### Spectre – fenêtre de Hann

Lobe secondaire à -15.5 dB pour la fenêtre de Hann.



### Spectre – fenêtre de Blackman

Lobe secondaire à -28.5 dB pour la fenêtre de Blackman.

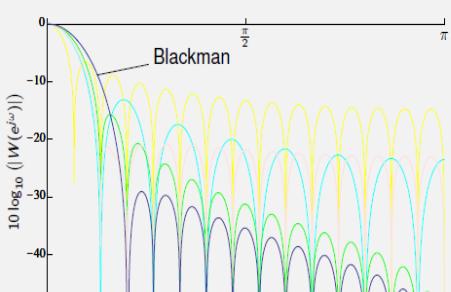


Figure 5-18. Réponses fréquentielles des principales fenêtres

Tableau 2 : Caractéristiques des principales fenêtres (N=L)

Fenêtre	Largeur du lobe principal	Hauteur maximale des lobes secondaires (dB)
Rectangulaire	$\frac{4\pi}{M}$	-6.5
Bartlett	$\frac{8\pi}{M}$	-12.5
Hann	$\frac{8\pi}{M}$	-15.5
Hamming	$\frac{8\pi}{M}$	-20.5
Blackman	$\frac{12\pi}{M}$	-28.5

#### Cas où l'intégrale ne peut pas être évaluée analytiquement:

Pour surmonter ce problème, on recourt à un calcul numérique de cette intégrale. Dans ce cas, la surface de la fonction  $H(f)e^{j2\pi fn}$  sur la période principale est approximée par la somme des surfaces d'un grand nombre de rectangles étroits et juxtaposés.

Pour faciliter les calculs, il est préférable de choisir une même largeur pour tous les rectangles. Les longueurs des rectangles sont les valeurs prises par la fonction à intégrer au début de chaque intervalle correspondant à la largeur. Si N représente le nombre de ces rectangles, la forme approchée de l'Éq 5-20 est donnée par:

$$\text{Éq 5-21} \quad \hat{h}[n] = \frac{1}{N} \sum_{k=-N/2}^{N/2-1} H\left(\frac{k}{N}\right) e^{j2\pi nk/N}$$

On retrouve l'expression de la TFD inverse de  $H(k)$ . La relation liant  $h[n]$  à son approximation

$\hat{h}[n]$  est donnée par:

$$\text{Éq 5-22} \quad \hat{h}[n] = \sum_{i=-\infty}^{+\infty} h[iN + n]$$

Pour  $N$  suffisamment grand par rapport à la durée présumée  $L$  de la réponse impulsionnelle cherchée, on peut écrire:  $h[n] = \hat{h}[n] \quad pour n = 0, \dots, L-1$

Dans quelle mesure peut-on approcher d'une façon satisfaisante la caractéristique fréquentielle souhaitée ? Aux fréquences qui coïncident exactement avec les échantillons  $H\left(\frac{k}{N}\right)$ , l'approximation est exacte, mais aux fréquences intermédiaires il s'introduit des erreurs qu'on ne peut pas directement contrôler. Pour valider le résultat, on doit calculer la TF (et non la TFD) de la réponse impulsionnelle  $h[n]$  obtenue et la comparer avec la caractéristique fréquentielle de départ.

### ***Conception à ondulations constantes (equiripple characteristic)***

Les méthodes de conception des filtres FIR décrites ci-dessus fournissent des filtres dans lesquels la divergence avec les caractéristiques idéales souhaitées se produit principalement au voisinage des transitions.

La conception des filtres à ondulations constantes cherche à distribuer les erreurs d'approximation d'une façon aussi équitable que possible sur toutes les fréquences au lieu qu'elles soient concentrées dans la (ou les) zone(s) de transition.

Ainsi, cette méthode utilise un gabarit de forme identique à celle de la Figure 5-2. En fixant les valeurs de  $\delta_1$ ,  $\delta_2$ ,  $f_1$  et  $f_2$ , on peut calculer les  $L$  coefficients du filtre en utilisant des méthodes d'optimisation itérative pour minimiser un critère d'erreur entre la courbe réelle du gabarit et le filtre idéal.

Plusieurs algorithmes ont été développés pour réaliser de tels filtres. Le plus connu est celui de Parks et McClellan qui reformule le problème de synthèse de filtre sous la forme d'une approximation polynomiale. Il fournit une approximation optimale du filtre FIR de Tchebychev et pour lequel l'utilisateur doit fournir les 5 paramètres cités ci-dessus ( $\delta_1$ ,  $\delta_2$ ,  $f_1$ ,  $f_2$ , et  $L$ ). Pour avoir une idée sur la valeur de  $L$  à utiliser, une estimation grossière peut être obtenue en utilisant la formule empirique suivante:  $L = 1 - \frac{10 \log_{10}(\delta_1 \cdot \delta_2) + 15}{14(f_2 - f_1)}$ .

### **Synthèse des filtres RII:**

Les méthodes de synthèse, les plus utilisées, des filtres à réponse impulsionnelle de durée infinie sont basées sur les techniques de transposition des filtres analogiques aux filtres numériques, en établissant une correspondance appropriée entre les deux domaines analogique et numérique.

Le problème général de synthèse consiste à déterminer l'ensemble des coefficients  $\{a_n\}$  et  $\{b_m\}$  de manière à ce que la réponse fréquentielle du filtre obtenue satisfasse le gabarit donné, et que le filtre soit réalisable, c'est à dire qu'il soit causal et stable tout en se rappelant que la contrainte de stabilité des filtres RII est plus sévère que celle des filtres RIF.

Il faut non seulement que les valeurs de  $h[n]$  soient finies mais aussi que  $\sum_{n=0}^{+\infty} |h[n]| < +\infty$  qui se traduit dans le domaine en  $z$  par le fait que les pôles de la fonction de transfert  $H(z)$  soient à l'intérieur du cercle unité.

Dans la transformation d'un filtre analogique en un filtre numérique, il est essentiel que les propriétés principales du filtre analogique soient conservées.

En fait, la transformation de Laplace est aux systèmes analogiques ce que la transformation en  $z$  est aux systèmes numériques. La transposition consiste à trouver un pont entre ces deux transformations.

De point de vue mathématique, un tel pont est une application du plan des  $p$  de la transformée de Laplace au plan des  $z$ , que l'on note  $p = f(z)$ . Dans une telle application, il est souhaitable que l'axe imaginaire du plan des  $p$  soit appliqué sur le cercle unité et que le demi-plan négatif des  $p$  soit appliqué à l'intérieur du cercle unité dans le plan des  $z$ . Ceci garantit qu'un filtre analogique stable se transforme en un filtre numérique stable.

Le problème de synthèse revient donc à trouver d'abord un filtre analogique qui remplit les spécifications du problème donné. On détermine ensuite sa fonction de transfert  $H_a(p)$  qui est la transformée de Laplace de sa réponse impulsionnelle  $h_a(t)$ . En substituant  $p=f(z)$  dans l'expression de  $H_a(p)$ , on obtient la fonction de transfert  $H(z)$  du filtre numérique correspondant:

$$\text{Eq 5-23} \quad H(z) = H_a(p) \Big|_{p=f(z)}$$

que l'on peut mettre sous la forme de l'Eq 5-15 pour en déduire les  $\{a_n\}$  et  $\{b_m\}$ .

L'application  $p = f(z)$  peut être obtenue par deux manières distinctes:

#### *Par équivalence de la dérivation :*

En approximant la différentiation, dans le domaine numérique, d'un signal  $x[n]$  à l'instant  $n$  par le signal  $y[n]$  donné par:

$$\text{Eq 5-24} \quad y[n] = \frac{x[n] - x[n-1]}{T}$$

où  $T$  est l'intervalle de temps entre les deux instants  $n-1$  et  $n$ , ou période d'échantillonnage dans le cas d'un signal échantillonné. La transformée en  $z$  de cette relation donne:

$$\text{Eq 5-25} \quad Y(z) = \frac{1-z^{-1}}{T} X(z)$$

Dans le plan de la transformée de Laplace, la dérivation s'écrit:  $Y(p) = p X(p)$

En approximant la différentielle par une différence, nous obtenons l'équivalence suivante:

$$\text{Eq 5-26} \quad p = \frac{1-z^{-1}}{T} \quad \text{ou} \quad z = \frac{1}{1-T.p}$$

Cette équivalence présente un désavantage du fait qu'il ne transpose pas l'axe imaginaire et le demi-plan gauche du plan de  $p$  sur et dans le cercle unité du plan de  $z$ , mais sur et dans un

cercle tangent intérieurement au cercle unité, de rayon  $\frac{1}{2}$  et de centre  $\frac{1}{2}$  (voir Figure 5-19).

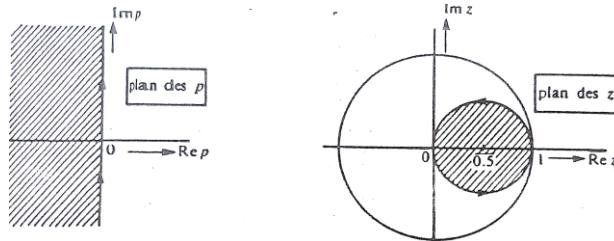


Figure 5-19. Application de  $z=1/(1-pT)$  du plan des  $p$  dans le plan des  $z$

En effet,

$$\text{Eq 5-27} \quad \left| z - \frac{1}{2} \right| = \left| \frac{1}{1-T.(r+j\omega_a)} - \frac{1}{2} \right| = \left| \frac{1+T.(r+j\omega_a)}{2(1-T.(r+j\omega_a))} \right| = \frac{\sqrt{(1+T.r)^2 + \omega_a^2}}{2\sqrt{(1-T.r)^2 + \omega_a^2}} \leq \frac{1}{2};$$

l'égalité est obtenue lorsque  $r$  est égale à 0, c.à.d. lorsque  $p$  décrit l'axe des imaginaires.

Ceci garantit la stabilité dans les deux domaines. Cependant, la précision de l'approximation augmente avec la fréquence d'échantillonnage ( $T$  petit).

De plus, cette application n'est pas idéale dans le sens où le système discret résultant n'a pas le même comportement en fréquence que le système analogique de départ dans l'intervalle fondamental  $[-\frac{1}{2}, +\frac{1}{2}]$ , car l'image de l'axe vertical du plan des  $p$  (axe des fréquences  $p=j\omega$ ) n'est pas le cercle unité du plan des  $z$  ( cercle des fréquences  $z=e^{j\omega}$ ) : c'est seulement dans la région où le petit cercle se rapproche du cercle unité (c.à.d pour les fréquences très basses :  $|\omega| < \pi$ ) qu'il existe une bonne concordance entre comportements en fréquence le système continu et le système discret équivalent. C'est la raison pour laquelle on préfère utiliser la méthode suivante basée sur l'équivalence de l'intégration.

#### Par équivalence de l'intégration

En approximant l'intégrale, dans le domaine numérique, d'un signal  $x[n]$  à l'instant  $n$  par le signal  $y[n]$  donné par:

$$\text{Eq 5-28} \quad y[n] - y[n-1] = T \cdot \frac{x[n] + x[n-1]}{2} = \text{Surface du trapèze formé par les échantillons } x[n-1] \text{ et } x[n]$$

qui, par transformée en  $z$ , donne:  $Y(z) = \frac{T}{2} \frac{1+z^{-1}}{1-z^{-1}} X(z)$

Dans le plan de la transformée de Laplace, l'intégration s'écrit:  $Y(p) = \frac{1}{p} X(p)$

d'où l'équivalence:

$$\text{Eq 5-29} \quad \frac{1}{p} = \frac{T}{2} \frac{1+z^{-1}}{1-z^{-1}} \Rightarrow p = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}} \quad \text{ou} \quad z = \frac{1+(T/2)p}{1-(T/2)p}$$

Cette relation définit ce qu'on appelle "**transformation bilinéaire**" du fait qu'elle est linéaire en  $z$  et en  $p$ . L'image de l'axe imaginaire du plan  $p$  dans le plan  $z$  est obtenue en remplaçant  $p$  par  $j\omega_a$  dans l'Eq 5-29, ce qui donne:

$$\text{Eq 5-30} \quad z = \frac{1 + j(T/2)\omega_a}{1 - j(T/2)\omega_a} \quad \text{avec } |z| = 1 \text{ et } \text{Arg}(z) = 2\text{Arctg}(\frac{T}{2}\omega_a) \Rightarrow z = e^{j2\text{Arctg}(\frac{T}{2}\omega_a)}$$

Ceci montre que lorsque  $p$  décrit l'axe imaginaire  $z$  décrit le cercle unité et que si  $p = -r + j\omega_a$  (avec  $r > 0$ ) alors module de  $z$  devient inférieur à 1, ce qui signifie que le demi-plan gauche de  $p$  est appliqué à l'intérieur du cercle unité dans le plan de  $z$ .

Ainsi, l'application de l'Éq 5-29 satisfait parfaitement la condition de transformer un filtre analogique stable en un filtre numérique stable.

Remarquons que  $H(f)$ , transformée de Fourier de  $h[n]$ , peut être obtenue à partir de

$$H(z) = \sum_{n=0}^{+\infty} h[n]z^{-n} \quad \text{en remplaçant } z \text{ par } e^{j\omega} = e^{j2\pi fT}; \quad f \text{ étant la fréquence du signal numérique.}$$

En identifiant cet exponentiel à celui de l'Éq 5-30, on peut écrire:

$$\text{Eq 5-31} \quad f_a = \frac{1}{\pi T} \text{tg}(\pi T f) \text{ soit } \omega_a = \frac{2}{T} \text{tg}(\frac{\omega}{2})$$

où  $\omega$  est la pulsation normalisée du domaine discret et  $\omega_a = 2\pi f_a$  la pulsation analogique.

D'où la possibilité d'obtenir directement  $H(f)$  à partir de  $H_a(f_a)$  en substituant  $f_a$  par son expression de l'Éq 5-31.

### Par invariance de la réponse impulsionnelle

Notons qu'il existe une méthode directe de passage du domaine analogique au domaine discret. Cette méthode, dite "à invariance de la réponse impulsionnelle" est relativement simple dans le cas où la fonction de transfert  $H_a(p)$  peut se mettre sous la forme d'une somme des facteurs simples d'ordre 1 uniquement :

$$\text{Eq 5-32} \quad H_a(p) = \sum_{k=1}^N \frac{A_k}{p - B_k}$$

avec  $B_k$  les pôles (réels ou complexes) de  $H_a(p)$ .

En partant de la réponse impulsionnelle du filtre analogique  $h_a(t)$ , l'objectif est de déterminer le filtre discret à réponse impulsionnelle infinie dont la réponse impulsionnelle  $h[n]$  telle que  $h[n] = h_a(nT)$  où  $T$  est la période d'échantillonnage du système discret.

Pour expliquer cette méthode, prenons l'exemple simple d'un filtre analogique dont la fonction de transfert  $H_a(p)$  est donnée par :

$$\text{Eq 5-33} \quad H_a(p) = \frac{A}{p - B}$$

dont la réponse impulsionnelle  $h_a(t)$  est donnée par la transformée inverse de Laplace par :

$$\text{Eq 5-34} \quad h_a(t) = \begin{cases} Ae^{Bt} & \text{pour } t \geq 0 \\ 0 & \text{pour } t < 0 \end{cases}$$

La réponse impulsionnelle discrète  $h[n]$  correspondante est alors :  $h[n] = h_a(nT) = Ae^{nBT} \cdot u[n]$  dont la transformée en  $z$  est :

Éq 5-35

$$H(z) = A \sum_{n=0}^{+\infty} (e^{BT} \cdot z^{-1})^n = \frac{A}{1 - e^{BT} \cdot z^{-1}}$$

On peut considérer que  $H(z)$  est l'approximation discrète de  $H_a(p)$ . En comparant les deux fonctions de transfert, on voit que  $H_a(p)$  a un pôle en  $p=B$  et que  $H(z)$  en a un à  $z=e^{BT}$ . C'est l'illustration d'une propriété générale de cette méthode de réalisation :

Chaque pôle simple  $p=p_k$  du filtre analogique est converti en un pôle simple  $z=e^{p_k T}$ .

En général, à la fonction de transfert  $H_a(p)$  donnée par l'Éq 5-32, on associe la fonction de transfert du filtre discret  $H(z)$  donnée par :

Éq 5-36

$$H(z) = \sum_{k=1}^N \frac{A_k}{1 - e^{p_k T} \cdot z^{-1}}$$

Avec ce passage selon les pôles, la stabilité dans le domaine analogique est conservée dans le domaine discret. En fait, si  $\operatorname{Re}\{p_k\} \leq 0$ , alors  $|e^{p_k T}| \leq 1$ .

En ce qui concerne les zéros, il n'existe pas une relation simple entre les zéros de  $H_a(p)$  et ceux de  $H(z)$ .

**Exemple :**

$$\text{Soit } H_a(p) = \frac{2p+22}{(p+1)(p^2+4p+13)} = \frac{2}{(p+1)} - \frac{2p+4}{(p^2+4p+13)} = \frac{2}{(p+1)} - \frac{1}{(p+2+3j)} - \frac{1}{(p+2-3j)}$$

En appliquant la méthode ci-dessus, on obtient  $H(z)$  :

$$\begin{aligned} H(z) &= \frac{2}{1 - e^{-T} \cdot z^{-1}} - \frac{1}{1 - e^{-(2+3j)T} \cdot z^{-1}} - \frac{1}{1 - e^{-(2-3j)T} \cdot z^{-1}} \\ &= \frac{2z[(e^{-T} - 2e^{-2T} \cos(3T)) \cdot z + e^{-4T} - e^{-3T} \cos(3T)]}{(z - e^{-T}) \cdot (z^2 - 2e^{-2T} \cos(3T) \cdot z + e^{-4T})} \end{aligned}$$

On remarque que les pôles sont convertis tels que nous l'avons décrit ci-dessus alors que les zéros sont complètement différents :  $H_a(p)$  possède un zéro à  $p=-11$ , alors que  $H(z)$  en possède deux à  $z=0$  et  $z=-\frac{[e^{-4T}-e^{-3T} \cos(3T)]}{[e^{-T}-2e^{-2T} \cos(3T)]}$ .

## 5.7 Comparaison entre les filtres FIR et IIR

La première question qui se pose pour la réalisation d'un filtre discret est celle du choix d'un filtre FIR ou d'un filtre IIR. Un grand nombre de facteurs entrent en jeu de sorte qu'il n'est toujours clair à l'avance de savoir quel sera le choix final.

Il se peut parfois être utile de considérer les deux solutions et d'évaluer en suite les deux possibilités pour savoir laquelle donnera la meilleure solution pour l'application considérée. Ce sont ici des facteurs très pratiques qu'il faut examiner comme la complexité, la consommation d'énergie, la rapidité de calcul, la facilité d'intégration et la disponibilité de certains modules de circuits.

Dans ce qui suit, nous résumons les facteurs plus ou moins théoriques qui différencient les

deux types de filtres :

Filtres FIR	Filtres IIR
<b>Fonction de transfert :</b> Elle possède des zéros uniquement	– Elle possède des pôles et des zéros.
<b>Méthodes de conception :</b> – Applicables à tout gabarit (filtres à plusieurs bandes passantes par exemple) – Nécessitent un ordinateur pour les procédures itératives de calcul	– Adaptées aux filtres classiques (Passe-bas, Passe-Haut, Passe-bande, Coupe-bande et Passe-tout). – Il n'est pas nécessaire d'utiliser un ordinateur lorsqu'on utilise des solutions toutes faites en analogique.
<b>Caractéristiques de phase :</b> – Possibilité d'une phase parfaitement linéaire. – Déphaseurs impossibles (Passe-tout)	– Les spécifications se portent uniquement sur les caractéristiques d'amplitude. On peut approcher une phase linéaire grâce à un égaliseur en cascade. – Possibilité d'un Passe-tout.
<b>Stabilité:</b> Toujours stables	Instables pour des pôles à l'extérieur du cercle unité
<b>Complexité:</b> Proportionnelle à la longueur	– Pas de relation directe entre la complexité et la longueur (théoriquement infinie). – On peut réaliser des filtres de sélectivité élevée avec un hardware de complexité relativement basse.
<b>Structure:</b> Réalisation transversale	– Réalisation récursive: la forme la plus largement utilisée est la connexion en cascade de cellules du 1 <sup>er</sup> et de 2 <sup>nd</sup> ordre.

# Chapitre 6 - Synthèse des filtres analogiques<sup>1</sup>

## 6.1 Rappel sur les filtres analogiques

Un filtre linéaire analogique est caractérisé par sa fonction de transfert isochrone ou réponse en fréquence:

Éq 6-1

$$H(j\omega) = \frac{Y(j\omega)}{X(j\omega)}$$

qu'on la décompose en réponse en amplitude  $A(\omega)$  et réponse en phase  $\varphi(\omega)$ :

Éq 6-2

$$H(j\omega) = A(\omega) \cdot e^{j\varphi(\omega)}$$

On définit également l'affaiblissement  $A_f(\omega)$ , mesuré en dB, et le délai de groupe  $\tau(\omega)$ :, mesuré en secondes:

Éq 6-3

$$A_f(\omega) = -20 \cdot \log_{10}(A(\omega))$$

Éq 6-4

$$\tau(\omega) = -\frac{\partial \varphi(\omega)}{\partial \omega}$$

Une transformation n'apporte pas de distorsion au signal auquel elle est appliquée, si elle restitue en sortie un signal  $y(t)$  de même forme que le signal d'entrée  $x(t)$ . Par contre, le signal d'entrée peut subir une amplification et/ou un délai:

Éq 6-5

$$y(t) = Kx(t - t_0)$$

Ceci correspond, en transformée de Fourier, à une amplification du spectre d'amplitude et à un déphasage linéaire:

Éq 6-6

$$Y(j\omega) = K \cdot X(j\omega) \cdot e^{-j\omega t_0}$$

et donc à une fonction de transfert de type :

Éq 6-7

$$H(j\omega) = K \cdot e^{-j\omega t_0}$$

Si on considère maintenant un filtre, dont le rôle est de produire un signal de sortie correspondant à une plage de fréquences du signal d'entrée, il est clair que ce filtre doit, si on veut éviter toute distorsion, vérifier Éq 6-6. Il doit donc présenter une réponse en amplitude **constante** et une réponse en phase **linéaire et passant par 0**, du moins dans la plage de fréquences utile, appelée **bande passante** (Figure 6-1).

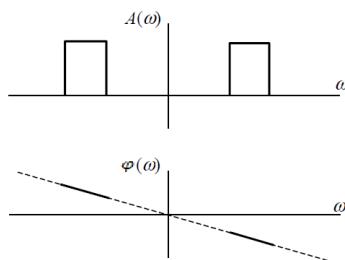


Figure 6-1. Réponse en amplitude et en phase d'un filtre n'introduisant pas de distorsion

En pratique, on admet parfois que le déphasage d'un filtre ne s'annule pas pour  $\omega=0$  :

Éq 6-8

$$H(j\omega) = K \cdot e^{-j(\omega t_0 + \alpha)} \text{ avec } \alpha \neq 2k\pi$$

<sup>1</sup> Ce chapitre est largement inspiré du polycopié "Introduction à la Synthèse des Filtres Actifs" de Thierry Dutoit (Faculté Polytechnique de Mons).

Ceci peut impliquer une distorsion de la forme du signal reçu.

Comme on l'a déjà vu dans le chapitre précédent pour les filtres numériques, on catégorise les filtres en fonction du type de modification qu'ils imposent sur leur entrée. Les filtres réalisant des modifications du spectre d'amplitude sont classés en filtres *passe-bas*, *passe-bande*, *passe-haut*, ou *coupe-bande*.

La forme générale de la fonction de transfert d'un filtre est:

Éq 6-9

$$H(p) = \frac{N(p)}{D(p)} = \frac{b_m p^m + \dots + b_1 p + b_0}{p^n + \dots + a_1 p + a_0}$$

L'*ordre* du filtre est  $n$ , qui doit bien entendu satisfaire à  $n >= m$ . Les zéros de  $N(p)$  sont les *zéros* du filtre; les zéros de  $D(p)$  sont les *pôles* du filtre. Les pôles du filtre doivent être situés à gauche de l'axe imaginaire (à parties réelles négatives ou nulles) pour que le filtre soit stable.

Nous étudions ici la manière de spécifier divers types de filtres. Nous verrons ensuite comment établir des fonctions de transfert qui permettent de respecter ces spécifications.

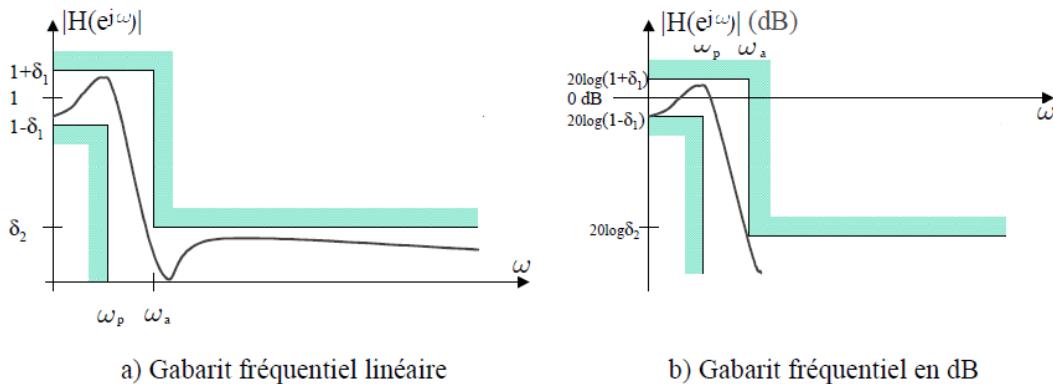


Figure 6-2. Gabarit fréquentiel d'un filtre passe-bas

Les spécifications d'un **filtre passe-bas** typique sont données à la Figure 6-2. Sa *bande passante* se situe entre 0 et la pulsation de coupure notée  $\omega_p$  et sa *bande atténuée* s'étend de  $\omega_a$  à *l'infini*. On accepte une certaine variation maximale  $\delta_1$  en bande passante (appelée ondulation en bande passante), et on impose que l'atténuation en bande atténuée soit inférieure à une valeur maximale  $\delta_2$  (appelée atténuation en bande atténuée).

Le passage entre zones passantes et zones atténuées se fait par des zones dites "de transition" dont la largeur exprime la sélectivité du filtre. Ainsi, la réponse en amplitude d'un filtre passe-bas possède trois zones distinctes : la bande passante ( $0 \leq \omega \leq \omega_p$ ), la bande de transition ( $\omega_p \leq \omega \leq \omega_a$ ) et la bande atténuée ( $\omega_a \leq \omega \leq +\infty$ ).

Les bandes passante et atténuée sont inversées pour le **filtre passe-haut** ce qui se traduit par le fait que  $\omega_p > \omega_a$ .

Pour le **filtre passe-bande** (Figure 6-3), on aura une bande passante ( $\omega_{p-} \leq \omega \leq \omega_{p+}$ ), deux bandes de transition ( $\omega_{a-} \leq \omega \leq \omega_{p-}$  et  $\omega_{p+} \leq \omega \leq \omega_{a+}$ ) et deux atténuées ( $0 \leq \omega \leq \omega_{a-}$  et  $\omega_{a+} \leq \omega \leq +\infty$ ).

Pour le **filtre coupe-bande**, la bande passante et les bandes atténuées sont inversées par rapport à celles du filtre passe-bande.

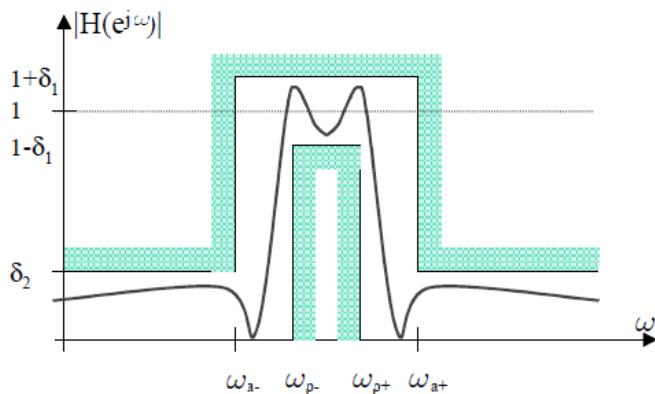


Figure 6-3. Gabarit fréquentiel d'un filtre passe-bande

La sélectivité  $s$  d'un filtre (toujours  $< 1$ ) est mesurée comme suit :

- Pour le filtre passe-bas :  $s = \frac{\omega_p}{\omega_a}$
- Pour le filtre passe-haut :  $s = \frac{\omega_a}{\omega_p}$
- Pour le filtre passe-bande :  $s = \frac{\omega_{p+} - \omega_{p-}}{\omega_{a+} - \omega_{a-}}$
- Pour le filtre coupe-bande :  $s = \frac{\omega_{a+} - \omega_{a-}}{\omega_{p+} - \omega_{p-}}$

## 6.2 Synthèse des filtres analogiques

Une fois le gabarit du filtre analogique est défini (valeurs de  $\omega_p$ ,  $\omega_a$ ,  $\delta_1$ , et  $\delta_2$  sont connues), la synthèse de la fonction de transfert du filtre  $H(p)$  suit la démarche suivante :

- Normalisation du gabarit
- Approximation de la fonction de transfert normalisé  $H(P)$  en utilisant une des méthodes d'approximation connues (les plus utilisées sont celles de Butterworth, de Chebychev et elliptique).
- Dé-normalisation de  $H(P)$  pour obtenir  $H(p)$ .

À partir de  $H(p)$ , on peut obtenir  $H(z)$  du filtre numérique correspondant aux spécifications du filtre analogique conçu.

### La normalisation du gabarit

Cette étape permet de transformer le gabarit défini au départ en un gabarit d'un filtre passe-bas prototype (normalisé – voir Figure 6-4) ayant une pulsation de coupure normalisée  $\Omega_p$  égale à 1 et une pulsation atténuee normalisée  $\Omega_a = \frac{\omega_a}{\omega_p}$  égale à l'inverse de la sélectivité ( $1/s$ ); La pulsation normalisée étant  $\Omega = \frac{\omega}{\omega_p}$ , on définit la variable de Laplace normalisée correspondante  $P = j\Omega$  (P majuscule).

Pour passer des spécifications réelles d'un filtre quelconque aux spécifications correspondantes d'un filtre passe-bas normalisé, on utilise une fonction  $\Omega = f(\omega)$  associant à toute fréquence  $\omega$  des spécifications réelles une fréquence  $\Omega$  des spécifications du passe-bas normalisé.

Cette fonction dépend du type du filtre réel à concevoir. Elle est donnée par le Tableau 3.

Tableau 3 : Fonctions de transformation des 4 types de filtre en filtre passe-bas normalisé

Passe-bas	Passe-haut	Passe-bande	Coupe-bande
$\Omega = \frac{\omega}{\omega_p}$	$\Omega = -\frac{\omega_p}{\omega}$	$\Omega = \frac{\omega_0}{B} \left( \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)$ $= \frac{\omega^2 - \omega_0^2}{B\omega}$ <p>où <math>B = \omega_{p+} - \omega_{p-}</math> est la largeur de la bande et <math>\omega_0 = \sqrt{\omega_{p+}\omega_{p-}}</math> est la pulsation centrale (moyenne géométrique)</p>	$\Omega = \left[ \frac{\omega_0}{B} \left( \frac{\omega_0}{\omega} - \frac{\omega}{\omega_0} \right) \right]^{-1}$ $= \frac{B\omega}{\omega_0^2 - \omega^2}$
À la bande passante $[0, \omega_p]$ du filtre passe-bas de départ correspond la bande passante $[0, 1]$ du filtre normalisé	La bande passante du filtre passe-haut $[\omega_p, \infty[$ est transformée en la bande passante $[0, -1]$ du filtre normalisé	La bande passante $[\omega_{p-}, \omega_{p+}]$ du filtre passe-bande ainsi que celle $[0, \omega_{p-}] \cup [\omega_{p+}, \infty[$ du filtre coupe-bande sont transformées en la bande $(-1, 1)$ du filtre normalisé. Notons qu'on ne pourra engendrer qu'une catégorie particulière des filtres passe-bande et coupe-bande : ceux pour lesquels les spécifications sont quasi les mêmes et en symétrie géométrique.	

Les valeurs de l'ondulation  $\delta_1$  et de l'atténuation  $\delta_2$  restent inchangées.  $A_p = -10\log_{10}(1 - \delta_1)^2$  et  $A_a = -10\log_{10}\delta_2^2$  sont respectivement les atténuations maximale admise en bande passante et minimale requise en bande atténuee; elles sont exprimées en dB.

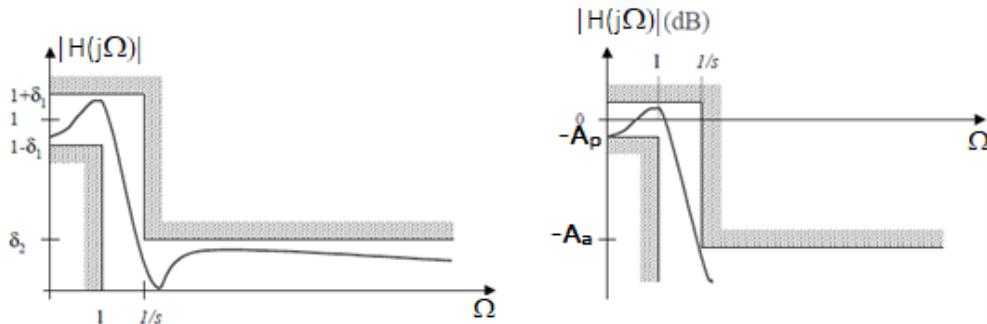


Figure 6-4. Gabarit prototype passe-bas (Normalisé)

### L'approximation de la fonction de transfert normalisée

Cette étape consiste à trouver la fonction de transfert  $H(P)$  qui correspond au gabarit normalisé du prototype passe-bas obtenu dans l'étape de normalisation.

En fait, le problème de l'approximation analytique consiste à positionner les pôles et les zéros de  $H(P)$  de façon à respecter les spécifications sur  $|H(j\Omega)|$ .

L'idéal est d'avoir  $A_p = 0$  dB et  $A_a$  infinie. Or, imposer une atténuation infinie correspond à placer les zéros de  $|H(j\Omega)|$  dans la bande atténuee (c.à.d. les zéros de  $H(P)$  sur l'axe imaginaire). Par contre, imposer une valeur unitaire à  $|H(j\Omega)|$  ne correspondent pas à placer les pôles de  $H(P)$  à un endroit particulier (les placer sur l'axe imaginaire donne un gain infini et non pas unitaire).

Pour simplifier le problème, on passe donc plutôt par le calcul d'une fonction  $K(j\Omega)$ , appelée

fonction caractéristique, dont on va s'arranger pour que ses zéros et ses pôles correspondent précisément aux fréquences pour lesquelles  $H(j\Omega)$  vaut 1 ou 0 (que l'on appelle parfois zéros et pôles d'affaiblissement, respectivement). Il suffit pour cela de poser:

**Éq 6-10**

$$|H(j\Omega)|^2 = \frac{1}{1 + |K(j\Omega)|^2} = \frac{|N(j\Omega)|^2}{|D(j\Omega)|^2} \Rightarrow |K(j\Omega)|^2 = \frac{|D(j\Omega)|^2 - |N(j\Omega)|^2}{|N(j\Omega)|^2}$$

Il est clair que les zéros de  $K(j\Omega)$  correspondent aux fréquences où l'atténuation vaut 0 dB ( $|H(j\Omega)| = 1$ ) et que les pôles de  $K(j\Omega)$  correspondent aux fréquences où l'atténuation est infinie. ( $|H(j\Omega)|$  est infiniment petit. Donc ces pôles correspondent aux zéros de  $H(j\Omega)$ ).

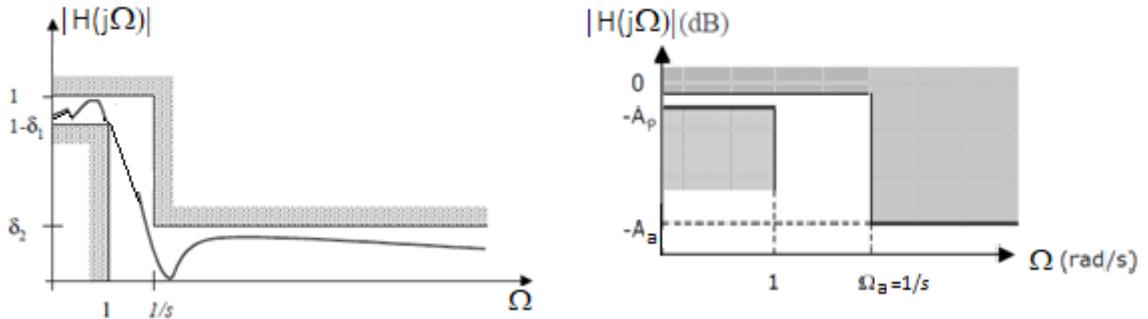


Figure 6-5. Gabarit prototype passe-bas normalisé et modifié selon la fonction caractéristique  $K(j\Omega)$

Le problème de l'approximation analytique devient alors:

- Positionner les pôles et les zéros de  $K(P)$  de façon à respecter les spécifications sur  $H(j\Omega)$  données à la Figure 6-4.
- Retrouver ensuite le  $H(P)$  correspondant.

Notons que la forme de  $|H(j\Omega)|^2$  donnée par l'Éq 6-10 indique que la valeur de  $|H(j\Omega)|$  ne peut dépasser la valeur 1, ce qui signifie que  $0 < |H(j\Omega)| \leq 1$ . Ceci amène à modifier légèrement le gabarit de la Figure 6-4 pour devenir celui de la Figure 6-5 où  $|H(j\Omega)|$  varie entre  $1-\delta_1$  et 1 (et non plus  $1+\delta_1$ ).

La Figure 6-6 donne l'allure correspondante de  $|K(j\Omega)|^2$  répondant à ces spécifications, ainsi que l'allure de la courbe de gain logarithmique correspondante.

Les relations entre  $\varepsilon$ ,  $\delta$ ,  $\delta_1$ ,  $\delta_2$ ,  $A_p$  et  $A_a$  sont données par:

**Éq 6-11**

$$(1 - \delta_1)^2 = \frac{1}{1 + \varepsilon^2} \Rightarrow \varepsilon = \frac{\sqrt{\delta_1(2 - \delta_1)}}{1 - \delta_1}$$

$$\delta_2^2 = \frac{1}{1 + \delta^2} \Rightarrow \delta = \frac{\sqrt{1 - \delta_2^2}}{\delta_2}$$

$$A_p = 10 \log_{10}(1 + \varepsilon^2) \Rightarrow \varepsilon^2 = 10^{A_p/10} - 1$$

et  $A_a = -10 \log_{10}(1 + \delta^2) \Rightarrow \delta^2 = 10^{A_a/10} - 1$

Les zéros et les pôles de  $K(P)$  sont situés sur l'axe imaginaire :

- les premiers sont situés dans la bande passante et sont appelés zéros de réflexion ( $\Omega_{r1}, \Omega_{r2}, \Omega_{r3}, \dots$ ) pour les fréquences desquels, le signal passe à travers le filtre sans être atténué

- les seconds sont situés en bande atténuée et sont appelés zéros de transmission ( $\Omega_{z1}, \Omega_{z2}, \Omega_{z3}, \dots$ ) pour les fréquences desquels, le signal est fortement atténué par le filtre. En fait, ils sont les zéros de  $H(j\Omega)$ .

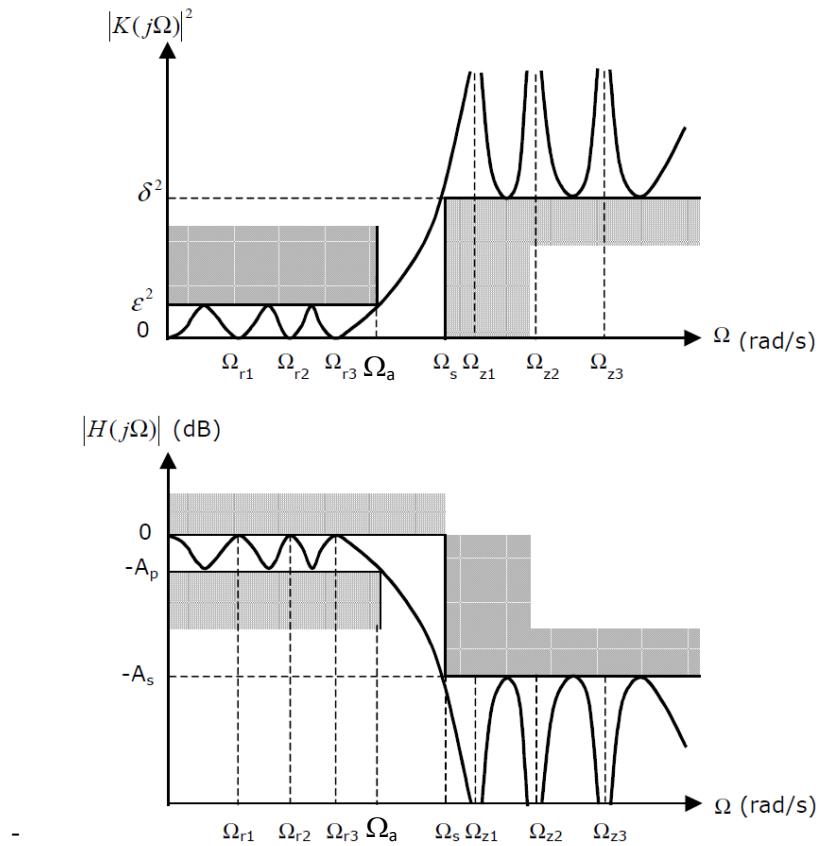


Figure 6-6. Allure de  $|K(j\Omega)|^2$  et du gain  $10\log_{10}|H(j\Omega)|^2$ .

Une fois la fonction  $K(j\Omega)$  respectant les spécifications est déterminée, il est facile de trouver  $H(j\Omega)$ . Et comme  $|H(j\Omega)|^2 = H(P).H(-P)|_{P=j\Omega}$ , il est toujours possible de retrouver une valeur de  $H(P)H(-P)$  en remplaçant  $\Omega$  par  $P/j$  dans  $|H(j\Omega)|^2$ .

Le dernier problème à résoudre consiste alors à répartir les zéros et les pôles de  $H(P).H(-P)$  entre  $H(P)$  et  $H(-P)$ . Ceci ne pose aucun problème pour les pôles puisque tous les pôles situés dans le demi-plan de gauche sont ceux de  $H(P)$ . La répartition des zéros de  $H(P).H(-P)$  est également univoque dans la mesure où ils seront dans la pratique tous situés sur l'axe imaginaire.

Il existe plusieurs fonctions d'approximation. Nous exposons ci-dessous les 3 méthodes d'approximation les plus utilisées :

- Approximation de Butterworth
- Approximation de Chebychev
- Approximation de Cauer (ou élliptique)

#### **Approximation de Butterworth**

La façon la plus simple pour respecter les spécifications sur  $|K(j\Omega)|^2$  (Figure 6-6) est de la mettre sous la forme:

$$K(j\Omega) = \epsilon \cdot \Omega^n$$

en fixant  $n$  de façon que

$$|K(j\Omega_a)|^2 \geq \delta^2.$$

Il s'agit d'une approximation polynomiale (la fonction caractéristique est un polynôme); Les zéros de réflexion se trouvent tous à l'origine et il n'y a pas de zéros de transmission. On parle d'approximation *méplate* (c.-à-d. extrêmement plate) à l'origine.

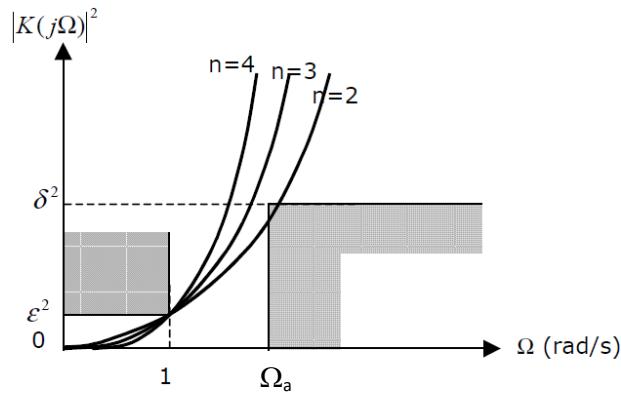


Figure 6-7. Allure des fonctions caractéristiques du filtre passe-bas de Butterworth

La détermination de l'ordre  $n$  du filtre est obtenue par la relation suivante :

$$\varepsilon^2 \cdot \Omega_a^{2n} \geq \delta^2 \Rightarrow n \geq \frac{\log \frac{\delta}{\varepsilon}}{\log \Omega_a} = \frac{\log \frac{10^{A_a/10}-1}{10^{A_p/10}-1}}{2 \cdot \log \Omega_a}$$

et puisque  $n$  est un entier, donc on prend l'entier juste supérieur à  $\frac{\log \frac{10^{A_a/10}-1}{10^{A_p/10}-1}}{2 \cdot \log \Omega_a}$ .

Eq 6-12

$$n = \left\lceil \frac{\log \frac{10^{A_a/10}-1}{10^{A_p/10}-1}}{2 \cdot \log \Omega_a} \right\rceil$$

L'expression de  $|H(j\Omega)|^2$  est alors:

$$\text{Eq 6-13} \quad |H(j\Omega)|^2 = \frac{1}{1+|K(j\Omega)|^2} = \frac{1}{1+\varepsilon^2 \cdot \Omega^{2n}}$$

et le remplacement de  $\Omega^2$  par  $-P^2$  donne :  $H(P) \cdot H(-P) = \frac{1}{1-\varepsilon^2 \cdot P^{2n}}$

On en conclut que  $H(P) \cdot H(-P)$  ne possède pas de zéros, et que ses  $2n$  pôles sont les racines de  $1 - \varepsilon^2 \cdot P^{2n} = 0$ .

Ces racines sont réparties sur un cercle de rayon  $\sqrt[n]{\frac{1}{\varepsilon}}$ . On retient pour  $H(P)$  les  $n$  pôles qui se trouvent à gauche de l'axe imaginaire (à partie réelle  $\leq 0$ ). Ces pôles sont donnés par l'expression suivante :

$$P_k = \sqrt[n]{\frac{1}{\varepsilon}} \cdot e^{j(\frac{\pi}{2} + \frac{2k-1}{2n}\pi)} \quad \text{pour } k = 1, \dots, n$$

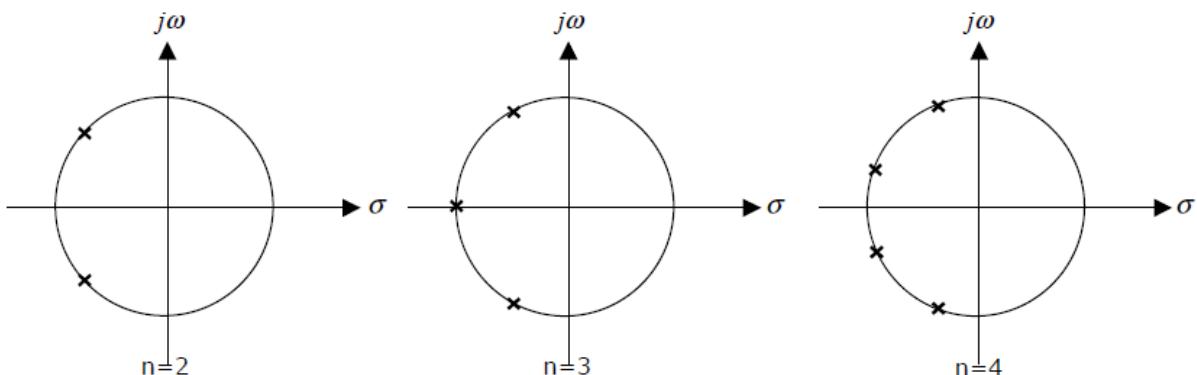


Figure 6-8. Localisation des racines du filtre passe-bas de Butterworth pour différentes valeurs de  $n$ .

Il est intéressant de calculer le comportement asymptotique ( $\Omega \rightarrow +\infty$ ) de la courbe de gain logarithmique:

Éq 6-14

$$20\log|H(j\Omega)||_{\Omega \rightarrow +\infty} = 10 \log \frac{1}{1 + \varepsilon^2 \cdot \Omega^{2n}} = -20 \log \varepsilon - 20n \cdot \log \Omega$$

On retrouve bien dans cette expression la classique chute de -20 dB/décade fois le nombre de pôles du filtre.

Exemple d'utilisation du Matlab pour obtenir un filtre de Butterworth

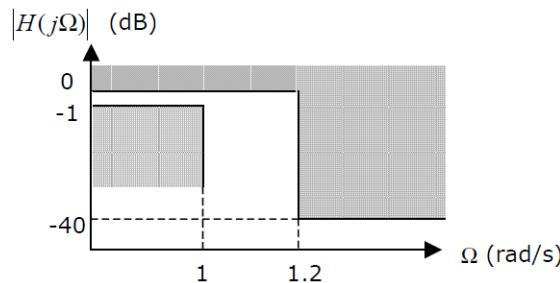


Figure 6-9. Gabarit d'un filtre passe-bas simple.

Pour calculer l'approximation de Butterworth pour le filtre passe-bas normalisé de la Figure 6-9, on peut utiliser Matlab avec les instructions suivantes:

```
wp=1; wa=1.2; Ap=1; Aa=40;
[n,wn]=buttord (wp,wa,Ap,Aa, 's') % Calcul de l'ordre du filtre normalisé
% de Butterworth pour Wp,Wa,Ap, et Aa.
% Wn est la pulsation dite "Naturelle" à -3dB
[z,p,k]=buttap(n)    % donne les zéros et les pôles ainsi que le gain de la
% fonction de transfert analogique du filtre d'ordre n
% buttap retourne l'approximation correspondant à une valeur unitaire de ε (c.-à.-d. à une
% valeur de Ap égale à 3 dB, et à des pôles sur le cercle de rayon unité). On obtient
% facilement l'approximation recherchée en divisant les pôles par ε^(1/n) (racine nème de ε).
eps=sqrt(10^(1/10)-1) % Calcul de ε en utilisant sa relation avec Ap
p=p/(eps^(1/n));       % On divise les pôles par ε^(1/n)
D=poly(p)               % On déduit les coefficients du polynôme de degré n dont les racines
% sont les éléments de p
N=D(n+1)                % D est le dénominateur de la fonction de transfert, N est le
% numérateur dont la valeur est telle que le gain statique est égal à k
G=tf(N,D)               % Affiche la fonction de transfert sous forme d'un rapport de deux
% polynômes dont les coefficients sont les éléments de N et D
figure
zplane([],D);           % Tracé des positions des pôles et des zéros
figure
freqs(N,D);             % Réponse en fréquence du filtre (Tracé de Bode)
[H,w]=freqs(N,D);        % H: Module w: vecteur des pulsations pour lesquelles H est calculé
% Tracé du délai de groupe
figure
semilogx (w(1:length(w)-1), -diff(unwrap(angle(H)))./diff(w));
xlabel('Frequency (radians)'); ylabel('Group delay (s)'); grid;
```

Le filtre obtenu est d'ordre n=29. Sa fonction de transfert H(P) est :

$$H(P) = \frac{1.965}{P^{29} + 18.91P^{28} + 178.7P^{27} + 1124P^{26} + 5282P^{25} + 19740P^{24} + 60960P^{23} + 159800P^{22} + 361800P^{21} + 717500P^{20} + 1259000P^{19} + 1967000P^{18} + 2754000P^{17} + 3466000P^{16} + 3932000P^{15} + 4025000P^{14} + 3717000P^{13} + 3094000P^{12} + 2315000P^{11} + 1552000P^{10} + 927100P^9 + 489800P^8 + 226600P^7 + 90590P^6 + 30730P^5 + 8615P^4 + 1921P^3 + 320P^2 + 35.46P + 1.965}$$

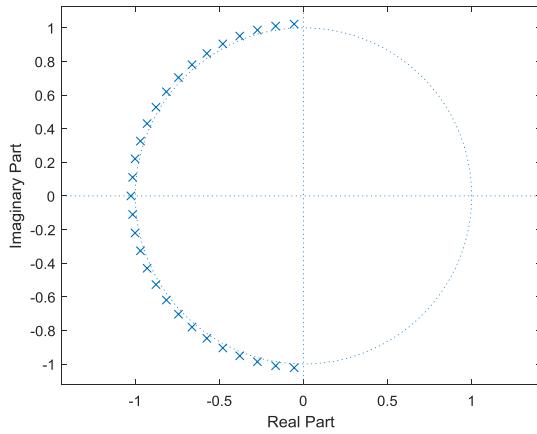


Figure 6-10. Répartition des pôles de l'approximation de Butterworth du passe-bas normalisé.

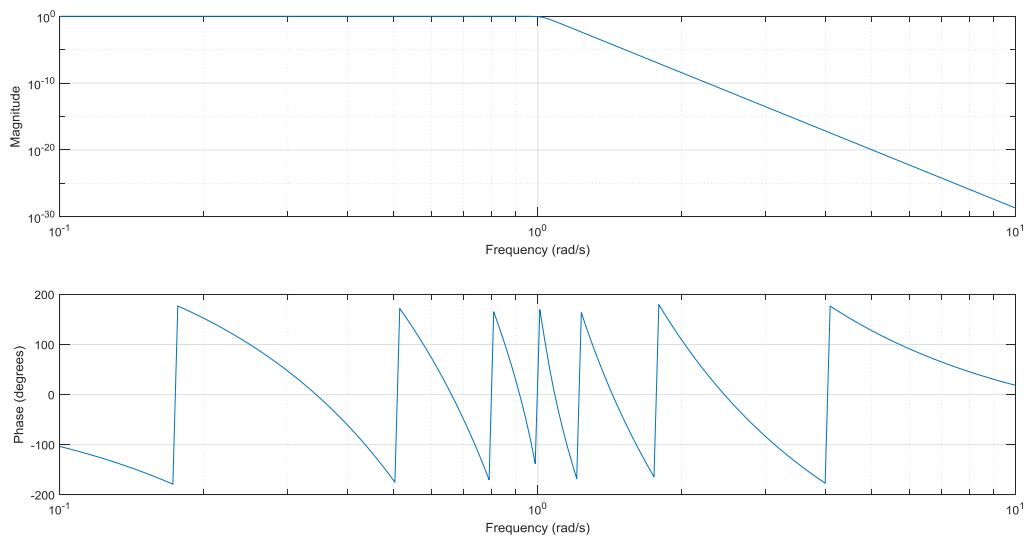


Figure 6-11. Réponse en fréquence de l'approximation de Butterworth du passe-bas normalisé.

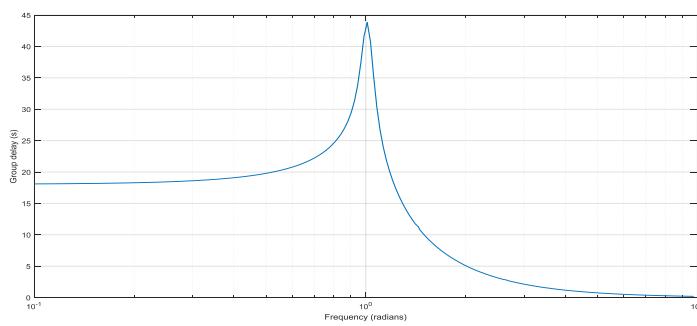


Figure 6-12. Gain et délai de groupe de l'approximation de Butterworth d'un passe-bas normalisé.

### Approximation de Chebychev

La courbe d'affaiblissement des filtres de Butterworth varie d'une façon monotone, ce qui implique que l'écart entre les spécifications et la courbe de gain dans la bande passante sera toujours minimale à la fréquence de coupure et maximale à l'origine. De même, cet écart est petit à droite de  $\Omega_a$  et plus grand partout ailleurs en bande atténuee. Bref, le filtre de Butterworth est trop bon presque partout, d'où son degré exagérément élevé.

spécifications, consiste à répartir l'erreur de façon plus uniforme dans la bande passante, en choisissant  $K(j\Omega) = \varepsilon \cdot C_n(\Omega)$  où  $C_n(\Omega)$  est un polynôme oscillant entre -1 et 1, de sorte que  $|K(j\Omega)|^2$  oscillerait entre 0 et  $\varepsilon^2$ . La valeur de  $n$  est fixée de façon que  $|K(j\Omega_a)|^2 \geq \delta^2$  soit toujours vérifiée.

Il s'agit là aussi d'une approximation polynomiale. Les polynômes  $C_n(\Omega)$  sont ceux de Chebychev. L'approximation correspondante est dite de type I; elle possède des zéros de réflexion dans la bande passante (Figure 6-13) et ne possède de zéros de transmission.

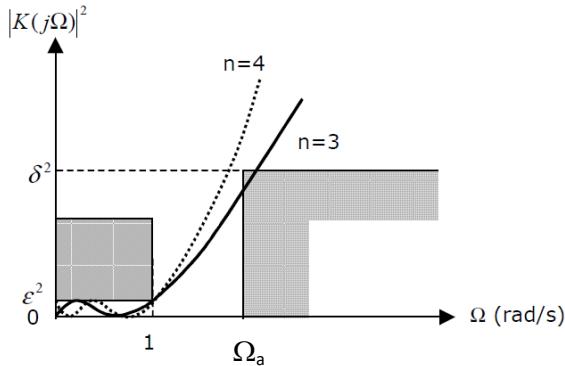


Figure 6-13. Allure des fonctions caractéristiques du filtre passe-bas de Chebychev (type I)

On appelle polynôme de Chebychev d'ordre  $n$  le polynôme défini par:

$$\text{Eq 6-15} \quad C_n(x) = \begin{cases} \cos[n \cdot \arccos(x)] & \text{pour } |x| \leq 1 \\ \cosh[n \cdot \operatorname{arccosh}(x)] & \text{pour } |x| > 1 \end{cases}$$

Contrairement à ce qu'il paraît, les  $C_n(x)$  sont bien des polynômes. On peut, en effet, montrer à l'aide de formules trigonométriques classiques que l'on a:

$$\text{Eq 6-16} \quad C_{n+1}(x) = 2xC_n(x) - C_{n-1}(x) \quad \text{avec } C_1(x) = x \text{ et } C_0(x) = 1$$

Les polynômes de Chebychev passent par les points caractéristiques suivants :

$$\text{Eq 6-17} \quad C_n(1) = \pm 1 \quad \text{et} \quad C_n(0) = \begin{cases} \pm 1 & \text{si } n \text{ pair} \\ 0 & \text{si } n \text{ impair} \end{cases}$$

Pour  $|x| \leq 1$ ,  $C_n(x)$  oscille  $n$  fois entre 1 et -1 (donc  $C_n(x)^2$  oscille entre 0 et 1) alors que pour  $|x| > 1$ , ces polynômes sont monotones croissants. La Figure 6-14 représente l'allure de  $C_n(x)^2$  pour plusieurs valeurs de  $n$ .

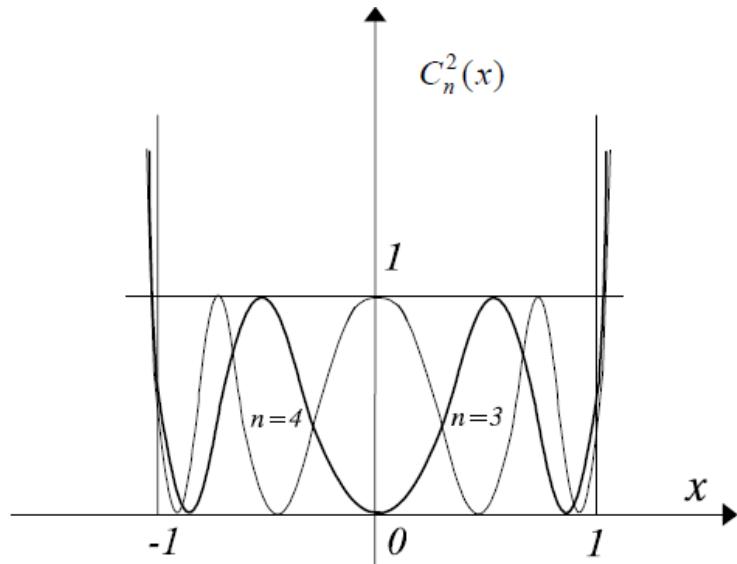


Figure 6-14. Allure des carrés des polynômes de Chebychev (pour  $n \leq 4$ )

La détermination de l'ordre n du filtre est obtenue par la relation suivante :

$$\varepsilon^2 \cdot C_n(\Omega_a)^2 \geq \delta^2 \Rightarrow \varepsilon^2 \cdot \cosh^2[n \cdot \operatorname{arccosh}(\Omega_a)] \geq \delta^2$$

$$\Rightarrow n \geq \frac{\operatorname{arccosh}\left(\frac{\delta}{\varepsilon}\right)}{\operatorname{arccosh}(\Omega_a)} = \frac{\operatorname{arccosh} \sqrt{\frac{10^{A_a/10}-1}{10^{A_p/10}-1}}}{\operatorname{arccosh} \Omega_a}$$

et puisque n est un entier, donc on prend l'entier juste supérieur à  $\frac{\operatorname{arccosh} \sqrt{\frac{10^{A_a/10}-1}{10^{A_p/10}-1}}}{\operatorname{arccosh} \Omega_a}$ :

**Eq 6-18**

$$n = \left\lceil \frac{\operatorname{arccosh} \sqrt{\frac{10^{A_a/10}-1}{10^{A_p/10}-1}}}{\operatorname{arccosh} \Omega_a} \right\rceil$$

sachant que  $\operatorname{arccosh}(x) = \ln(x + \sqrt{x^2 - 1})$ .

L'expression de  $|H(j\Omega)|^2$  est alors:

**Eq 6-19**

$$|H(j\Omega)|^2 = \frac{1}{1 + |K(j\Omega)|^2} = \frac{1}{1 + \varepsilon^2 \cdot C_n(\Omega)^2}$$

et le remplacement de  $\Omega$  par  $P/j$  donne  $H(P) \cdot H(-P)$  dont les  $2n$  pôles sont les racines de  $1 + \varepsilon^2 \cdot C_n(P/j)^2 = 0$ .

On montre que ces racines sont réparties sur une ellipse. Elles sont de la forme :

**Eq 6-20**

$$P_k = \left[ \sin\left(\frac{2k-1}{20}\pi\right) \cdot \sinh(\alpha) \right] + j \left[ \cos\left(\frac{2k-1}{20}\pi\right) \cdot \cosh(\alpha) \right] \quad \text{pour } k = 1, \dots, 2n$$

où  $\alpha = \frac{1}{n} \operatorname{arcsinh}\left(\frac{1}{\varepsilon}\right)$ .

Enfin, comme pour l'approximation de Butterworth, il est intéressant de calculer le comportement asymptotique de la courbe de gain logarithmique. On montrera à titre d'exercice qu'il vaut :

**Eq 6-21**

$$20 \log |H(j\Omega)| \Big|_{\Omega \rightarrow +\infty} = 10 \log \frac{1}{1 + \varepsilon^2 \cdot C_n(\Omega)^2} = -20 \log \varepsilon - 20n \cdot \log \Omega - 6.02(n-1)$$

On retrouve bien dans cette expression la valeur obtenue pour l'approximation de Butterworth (Eq 6-14) réduite de  $6.02(n-1)$ . On en conclut que, à degré égal, un filtre de Chebychev présente toujours une atténuation plus grande en bande atténuée qu'un filtre de Butterworth. Ainsi, pour respecter les mêmes spécifications, un filtre de Chebychev a toujours un degré inférieur ou égal à un filtre de Butterworth.

En fait, l'approximation de Butterworth n'est utilisée que lorsqu'il est fondamental d'avoir une courbe de gain très plate en bande passante. Son intérêt est donc plutôt didactique que pratique.

**Exemple d'utilisation du Matlab pour obtenir un filtre de Chebychev type I**

Pour calculer l'approximation de Chebychev type I pour le filtre passe-bas normalisé de la Figure 6-9, on peut utiliser Matlab avec les instructions suivantes:

```

wp=1; wa=1.2; Ap=1; Aa=40;
[n,wn]=cheblord (wp,wa,Ap,Aa, 's') % Calcul de l'ordre du filtre normalisé de Chebychev 1
% pour Wp, Wa, Ap et Aa.
% Wn est la pulsation dite "Naturelle" à -3dB
[z,p,k]=cheblap(n,Ap)           % donne les zéros, pôles et gain de la fonction de
% transfert analogique du filtre d'ordre n et
% d'atténuation Ap dans la bande passante
D=poly(p)                      % On déduit les coefficients du polynôme de degré n dont les racines
% sont les éléments de p. D est le dénominateur de la fonction de transfert
N=k*poly(z)                    % N est le numérateur en tenant compte du gain statique k

G=tf(N,D)                      % Affiche la fonction de transfert sous forme d'un rapport de deux
% polynômes dont les coefficients sont les éléments de N et D
figure
zplane(N,D);                  % Tracé des positions des pôles et des zéros
figure
freqs(N,D);                   % Réponse en fréquence du filtre (Tracé de Bode)
[H,w]=freqs(N,D);             % H: Module w: vecteur des pulsations pour lesquelles H est calculé
% Tracé du délai de groupe
figure
semilogx (w(1:length(w)-1), -diff(unwrap(angle(H)))./diff(w));
xlabel('Frequency (radians)'); ylabel('Group delay (s)'); grid;

```

Le filtre obtenu est d'ordre n=10. Sa fonction de transfert H(P) est :

$$H(P) = \frac{0.003838}{P^{10} + 0.9159P^9 + 2.919P^8 + 2.108P^7 + 2.982P^6 + 1.613P^5 + 1.244P^4 + 0.4554P^3 + 0.1825P^2 + 0.0345 s + 0.004307}$$

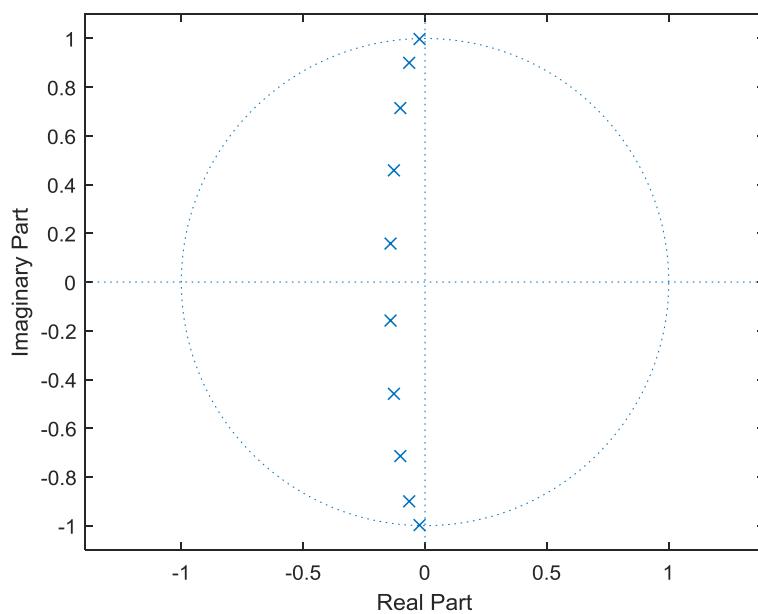


Figure 6-15. Répartition des pôles de l'approximation de Chebychev type 1 du passe-bas normalisé.

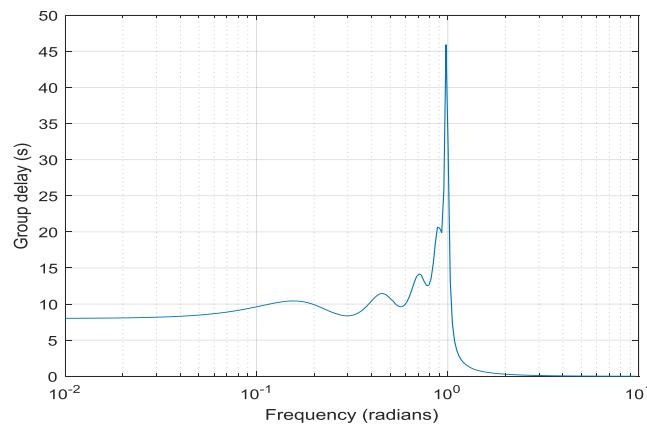


Figure 6-16. Gain et délai de groupe de l'approximation de Chebychev 1 d'un passe-bas normalisé.

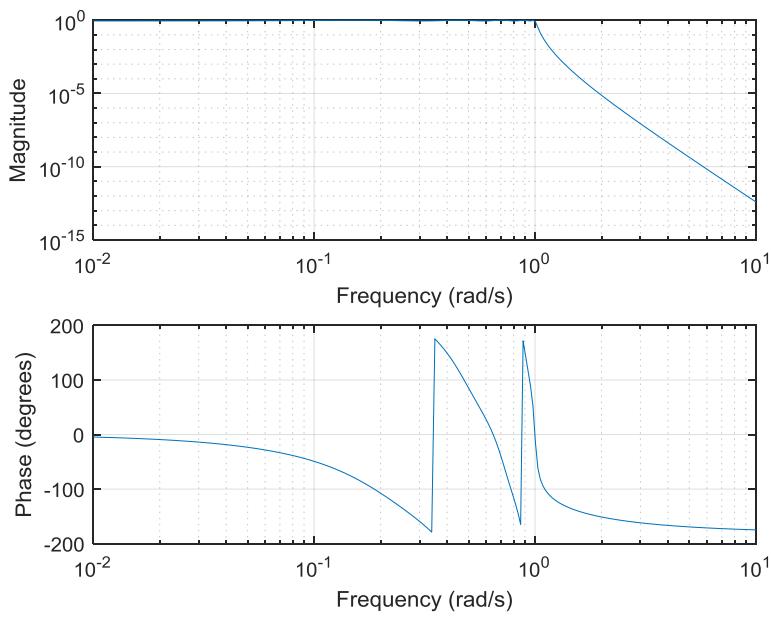


Figure 6-17. Réponse en fréquence de l'approximation de Chebychev 1 du passe-bas normalisé.

Une autre alternative utilisant les polynômes de Chebychev consiste à déplacer les ondulations dans la bande atténuee et à avoir la forme plate dans la bande passante.

Pour cela, on inverse le paramètre des polynômes  $C_n(x)$  en remplaçant  $\Omega$  par  $1/\Omega$ . Ainsi, pour  $\Omega < 1$ ,  $C_n(1/\Omega)$  est strictement monotone croissante.

Or, pour assurer cette continuité jusqu'au début de la bande atténuee et éviter l'ondulation dans la zone de transition, on modifie encore une fois le paramètre en le remplaçant par  $\Omega_a/\Omega$ . Ainsi, pour  $\Omega < \Omega_a$ ,  $C_n(\Omega_a/\Omega)$  reste monotone.

De plus, pour que  $|K(j\Omega_a)|$  soit égale à  $\delta$  et que  $|K(j\Omega)|$  soit supérieure à  $\delta$  pour  $\Omega > \Omega_a$ , on lui choisit l'expression suivante:

$$K(j\Omega) = \frac{\delta}{C_n\left(\frac{\Omega_a}{\Omega}\right)}$$

Cette approximation force la courbe de gain à passer par  $(\Omega_a \text{ rad/s}, -A_a \text{ dB})$  – voir Figure 6-18.

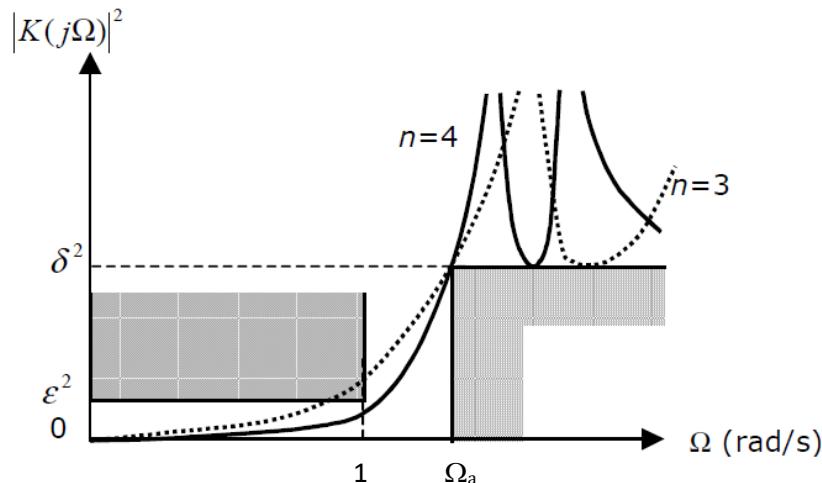


Figure 6-18. Allure des fonctions caractéristiques du filtre passe-bas de Chebychev type II.

On peut montrer que, pour des spécifications identiques, Chebychev I et Chebychev II sont de degrés identiques. Ils approximent donc aussi bien l'un que l'autre les spécifications en amplitude. Par contre, leurs réponses en phases sont très différentes. Les pôles de l'approximation de Chebychev I ont des facteurs de qualité plus élevés, ce qui conduit à des délais de groupes moins constants en fréquence.

### Exemple d'utilisation du Matlab pour obtenir un filtre de Chebychev type II

Pour calculer l'approximation de Chebychev type II pour le filtre passe-bas normalisé de la Figure 6-9, on peut utiliser Matlab avec les instructions suivantes:

```

wp=1; wa=1.2; Ap=1; Aa=40;
[n,wn]=cheb2ord (wp,wa,Ap,Aa, 's') % Calcul de l'ordre du filtre normalisé de Chebychev 2
                                         % pour Wp, Wa, Ap et Aa.
                                         % Wn est la pulsation dite "Naturelle" à -3dB
[z,p,k]=cheb1ap(n,Ap)             % donne les zéros, pôles et gain de la fonction de
                                         % transfert analogique du filtre d'ordre n et
                                         % d'atténuation Ap dans la bande passante
D=poly(p)                         % On déduit les coefficients du polynôme de degré n dont les racines
                                         % sont les éléments de p. D est le dénominateur de la fonction de transfert
N=k*poly(z)                       % N est le numérateur en tenant compte du gain statique k

[N2,D2]=lp2lp(N,D,wa); % cheb2ap donne l'approximation de Chebychev type 2
                           % avec wa comme fréquence normalisée.
                           % Pour corriger l'approximation pour une fréquence
                           % normalisée à wp, on applique la fonction lp2lp avec
                           % comme paramètre la valeur normalisée de wa.
G=tf(N2,D2)                      % Afficher la fonction de transfert sous forme d'un
                                         % rapport de deux polynômes dont les coefficients
                                         % sont les éléments de N2 et D2

figure
zplane(N2,D2);                  % Tracé des positions des pôles et des zéros
figure
freqs(N2,D2);                   % Réponse en fréquence du filtre (Tracé de Bode)

[H,w]=freqs(N2,D2);           % H: Module w: vecteur des pulsations pour lesquelles H est calculé
% Tracé du délai de groupe
figure
semilogx (w(1:length(w)-1), -diff(unwrap(angle(H)))./diff(w));
xlabel('Frequency (radians)'); ylabel('Group delay (s)'); grid;

```

Le filtre obtenu est d'ordre n=10. Sa fonction de transfert H(P) est :

$$H(P) = \frac{P^{10} + 9.69 \times 10^{-17} P^9 + 72 P^8 + 6.834 \times 10^{-15} P^7 + 829.4 P^6 + 4.082 \times 10^{-14} P^5 + 3344 P^4 + 2.628 \times 10^{-13} P^3 + 5504 P^2 + 2.081 \times 10^{-13} s + 31700}{100(P^{10} + 8.142 P^9 + 33.15 P^8 + 88.77 P^7 + 173.6 P^6 + 259.2 P^5 + 303.4 P^4 + 275 P^3 + 193.6 P^2 + 93.72 s + 31.7)}$$

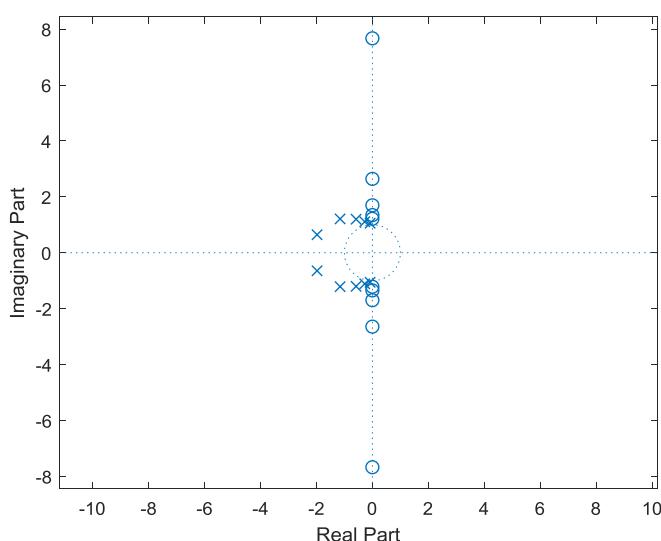


Figure 6-19. Répartition des pôles de l'approximation de Chebychev type 2 du passe-bas normalisé.

Notons que la courbe de délai de groupe de Chebychev I ressemble assez à celle de Chaouki DIAB

Butterworth (à spécifications inchangées) alors que celle de Chebychev II est beaucoup plus plate (Figure 6-21), si l'on fait abstraction des changements de signe brutaux de la phase dus à la présence de zéros sur l'axe imaginaire. On la préférera donc à l'approximation de Chebychev I pour les signaux sensibles à un décalage de phase non linéaire (signaux vidéo, signaux informatiques). Notons que, à chaque fois que la courbe de gain passe par une fréquence correspondant à un zéro sur l'axe imaginaire, on observe un brusque changement de signe de la courbe de phase, et donc un pic de délai de groupe. Ces pics n'ont pas d'importance en pratique, puisqu'ils sont situés en bande atténuee, et n'influencent donc pas beaucoup les phases des composantes spectrales en sortie du filtre.

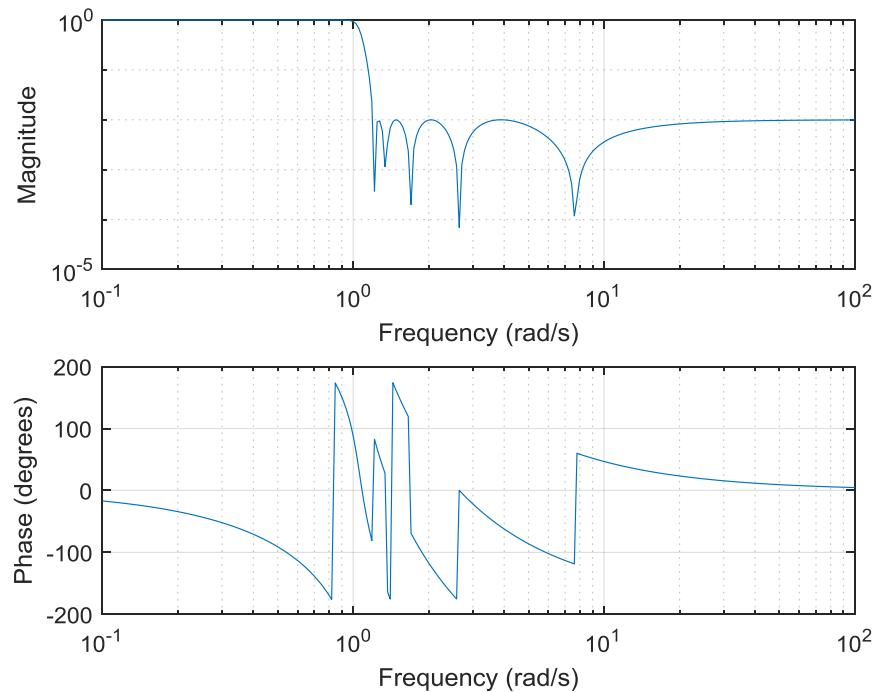


Figure 6-20. Réponse en fréquence de l'approximation de Chebychev 2 du passe-bas normalisé.

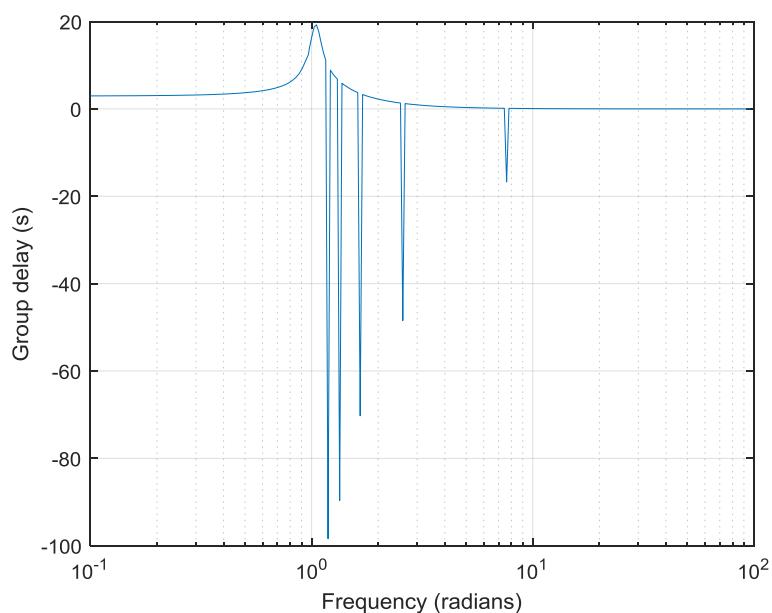


Figure 6-21. Gain et délai de groupe de l'approximation de Chebychev 2 d'un passe-bas normalisé.

### Approximation de Cauer (ou elliptique)

L'approximation est meilleure si on parvient à répartir l'erreur d'approximation de façon plus égale dans la bande passante ou dans la bande atténueée.

On doit donc pouvoir obtenir une approximation plus efficace encore en acceptant des ondulations de courbe de gain dans la bande passante *et* dans la bande atténueée.

La fonction caractéristique correspondante doit donc être cette fois une fraction rationnelle, présentant des zéros de réflexion *et* de transmission.

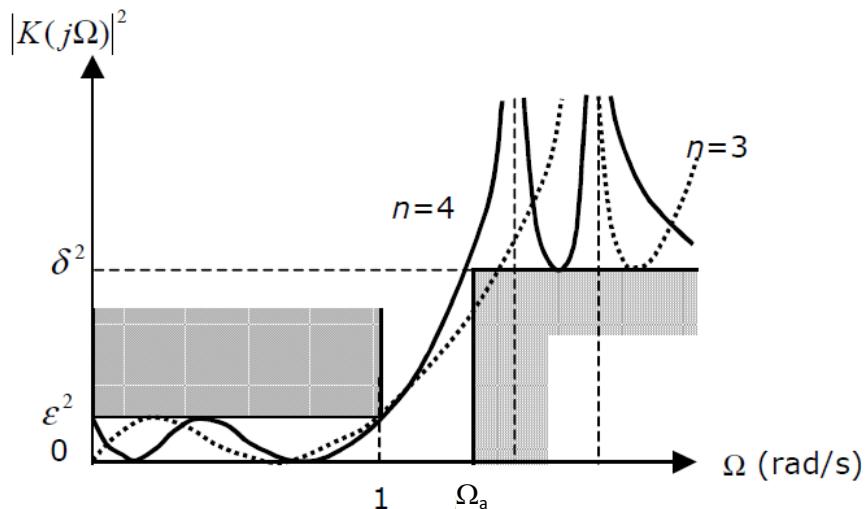


Figure 6-22. Allure des fonctions caractéristiques elliptiques de Cauer d'un filtre passe-bas.

Pour cela, on choisit  $K(j\Omega) = \varepsilon \cdot R_n(\Omega)$  où  $R_n(\Omega)$  est une fraction rationnelle dont l'élaboration a conduit à une théorie faisant intervenir les fonctions elliptiques, d'où le nom *d'approximation elliptique* (ou de Cauer, du nom de l'ingénieur qui l'a mise au point). L'allure de la fonction caractéristique correspond assez bien à une combinaison de l'allure d'une approximation de Chebychev I en bande passante, et d'une approximation Chebychev II en bande atténuee (Figure 6-22). L'estimation des paramètres de cette fonction est cependant nettement plus complexe. On se sert aujourd'hui systématiquement d'outils logiciels pour l'obtenir.

Pour calculer l'approximation de Cauer pour le filtre passe-bas normalisé de la Figure 6-9, on peut utiliser Matlab avec les instructions suivantes:

```

wp=1; wa=1.2; Ap=1; Aa=40;
[n,wn]=ellipord (wp,wa,Ap,Aa,'s') % Calcul de l'ordre du filtre normalisé de Cauer
% pour Wp, Wa, Ap et Aa.
% Wn est la pulsation dite "Naturelle" à -3dB
[z,p,k]=ellipap(n,Ap,Aa)      % donne les zéros, pôles et gain de la fonction de
% transfert analogique du filtre d'ordre n et d'atténuation Ap
% dans la bande passante et Aa dans la bande atténuee
D=poly(p)                      % On déduit les coefficients du polynôme de degré n dont les racines
% sont les éléments de p. D est le dénominateur de la fonction de transfert
N=k*poly(z)                    % N est le numérateur en tenant compte du gain statique k

G=tf(N,D)                      % Afficher la fonction de transfert sous forme d'un rapport de deux
% polynômes dont les coefficients sont les éléments de N et D
figure
zplane(z,p);                  % Tracé des positions des pôles et des zéros
figure
freqs(N,D);                   % Réponse en fréquence du filtre (Tracé de Bode)

[H,w]=freqs(N,D);            % H: Module w: vecteur des pulsations pour lesquelles H est calculé
% Tracé du délai de groupe
figure
semilogx (w(1:length(w)-1), -diff(unwrap(angle(H)))./diff(w));
xlabel('Frequency (radians)'); ylabel('Group delay (s)'); grid;

```

Le filtre obtenu est d'ordre n=6. Sa fonction de transfert H(P) est :

$$P^6 + 11.73P^4 + 28P^2 + 18.6$$

$$H(P) = \frac{P^6 + 11.73P^4 + 28P^2 + 18.6}{100(P^6 + 0.9154P^5 + 2.238P^4 + 1.48P^3 + 1.432P^2 + 0.5652s + 0.2087)}$$

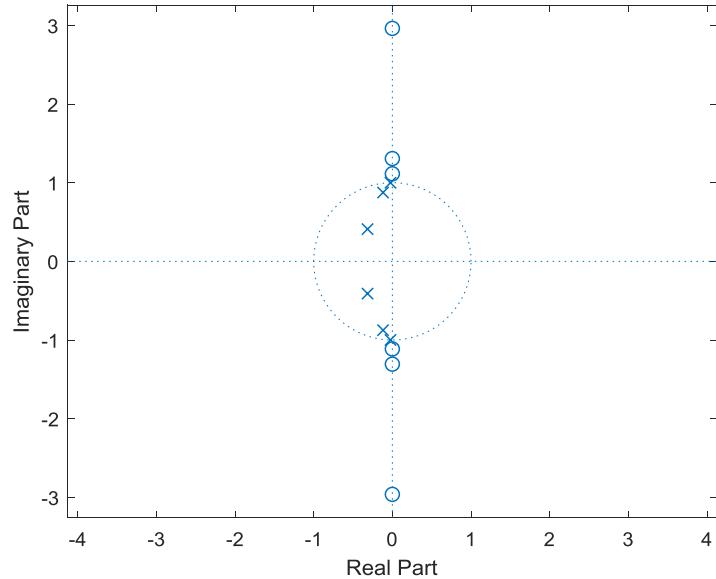


Figure 6-23. Répartition des pôles de l'approximation elliptique du passe-bas normalisé.

On remarque bien (Figure 6-24) une ondulation de la courbe de gain dans la bande passante et dans la bande atténuée. La courbe de délai de groupe est comparable à celle de Chebychev II (à spécifications inchangées).

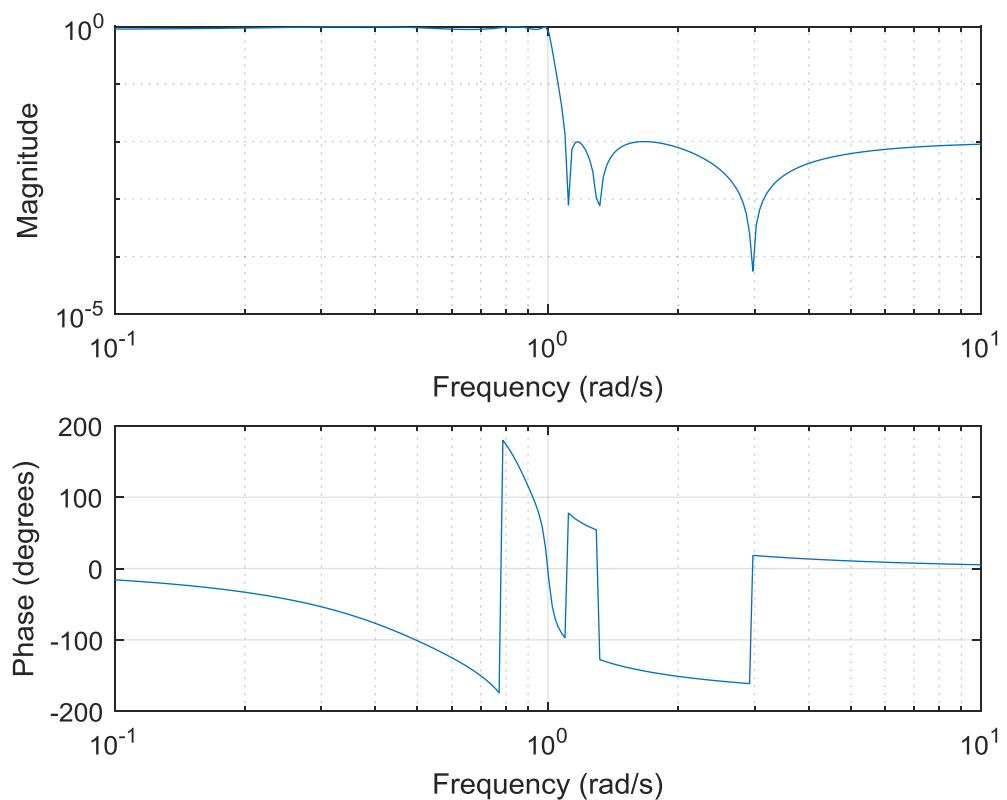


Figure 6-24. Réponse en fréquence de l'approximation elliptique du passe-bas normalisé.

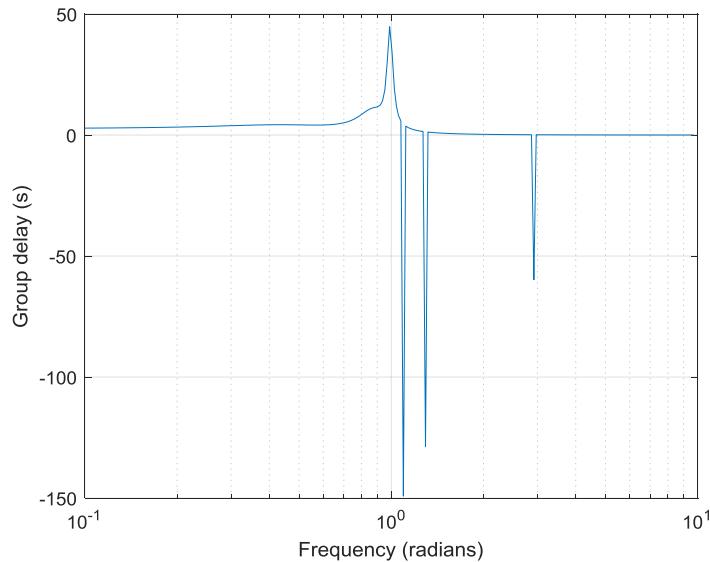


Figure 6-25. Gain et délai de groupe de l'approximation elliptique d'un passe-bas normalisé.

### La dé-normalisation

Cette dernière étape permet de passer de la fonction de transfert normalisée  $H(P)$  à la fonction de transfert  $H(p)$ . La dé-normalisation consiste donc en l'application d'une fonction  $f$  permettant de transformer la variable normalisée  $P=j\Omega$  en la variable  $p=j\omega$ :  $P = f(p)$ . La fonction  $f$  dépend du type de filtre spécifié.

Le Tableau 4 résume les fonctions de dé-normalisation pour les 4 types de filtres.

Tableau 4 : Fonctions de dé-normalisation filtre passe-bas normalisé en fonction du type de filtre à concevoir.

Passe-bas	Passe-haut	Passe-bande	Coupe-bande
$P = \frac{p}{\omega_p}$	$P = \frac{\omega_p}{p}$	$P = \frac{p^2 + \omega_0^2}{Bp}$ <p>où    <math>B = \omega_{p+} - \omega_{p-}</math></p>	$P = \left[ \frac{p^2 + \omega_0^2}{Bp} \right]^{-1} = \frac{Bp}{p^2 + \omega_0^2}$ <p>et    <math>\omega_0 = \sqrt{\omega_{p+} \cdot \omega_{p-}}</math></p>

# Chapitre 7 - Décomposition en sous-bandes et Transformée en Ondelettes

Introduite en 1976 par J. Crochière, la décomposition d'un signal en sous-bandes a été utilisée pour la première fois dans une application de codage de la parole. Depuis, cette technique s'est beaucoup évoluée avec l'introduction de la Transformée en Ondelettes qui lui a fourni un support théorique solide et des applications dans différents secteurs.

Son principe consiste en la décomposition de la bande fréquentielle d'un signal en plusieurs sous-signaux dont chacun correspond à une sous-bande fréquentielle de largeur plus ou moins grande.

Cette décomposition donne lieu à une répartition du contenu informationnel et énergétique du signal selon sa composition fréquentielle. Le principal avantage de cette technique, dans le domaine de la compression des signaux (images ou sons, elle est à la base, entre autres, du codage mp3 et de JPEG2000), par exemple, est de concentrer l'énergie du signal suivant certaines composantes fréquentielles.

Un système de codage/décodage en sous-bandes (Figure 7-1) peut être divisé en deux sous-systèmes distincts:

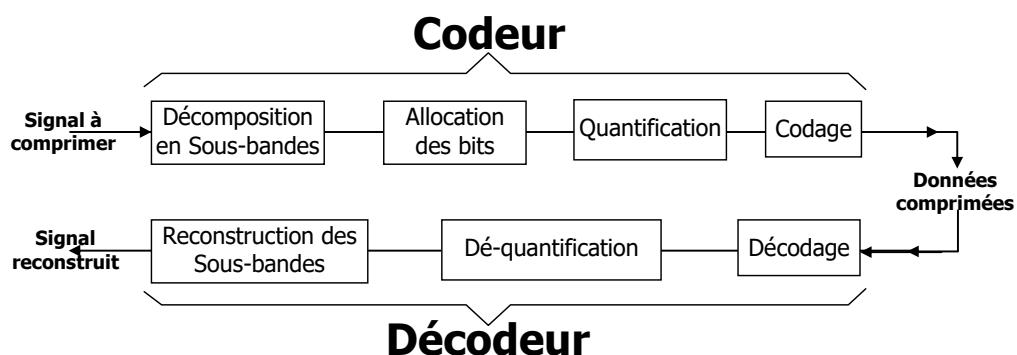


Figure 7-1. Système de codage/décodage en sous-bandes

- Le sous-système de compression formé par la partie "décomposition" contenant les bancs de filtres d'analyse suivie de la partie "codage" contenant les quantificateurs et les codeurs appropriés pour chaque sous-bande.
- Le sous-système de décompression formé par la partie "décodage" contenant les décodeurs correspondants suivie de la partie "reconstruction" contenant les bancs de filtres de synthèse associés à ceux d'analyse.

Un système de décomposition/reconstruction peut introduire trois distorsions distinctes:

- Repliement spectral,
- Distorsion de phase,
- Distorsion d'amplitude.

Les premières méthodes de décomposition en sous-bandes étaient considérées comme des procédés empiriques d'analyse du signal, les systèmes de décomposition/reconstruction

n'introduisant pas d'erreurs n'existaient pas:

*Le sous-échantillonnage, nécessaire pour réduire l'augmentation du volume de données causée par la décomposition, s'accompagnait toujours par une distorsion de repliement spectral (aliasing) qui induit une erreur de reconstruction.*

Dans ce qui suit, nous présentons le principe de décomposition en sous-bandes et les diverses solutions qui ont été proposées, pour aborder ensuite, sa relation avec la transformée en ondelettes.

## 7.1 Principe de la décomposition en sous-bandes

Le système de décomposition en sous-bandes comporte des filtres d'analyse qui décomposent le signal d'entrée en deux ou plusieurs sous-signaux.

Considérons le cas élémentaire d'un système de décomposition en deux sous-signaux: C'est l'élément de base. Il est formé d'un filtre passe-bas Ha et d'un filtre passe-haut Ga suivis chacun par un sous-échantillonneur de rapport 2 à 1.

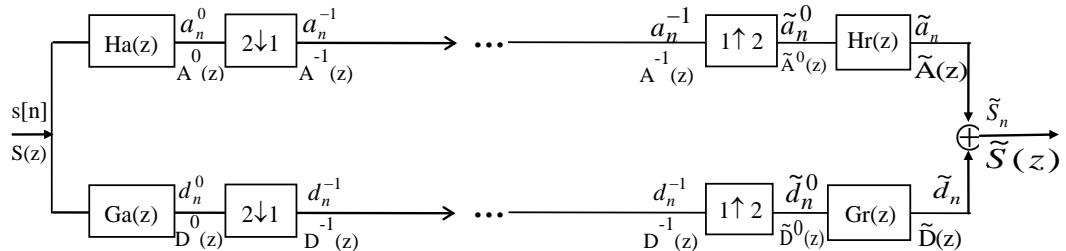


Figure 7-2. Système de décomposition/reconstruction à 2 bandes

La mise en cascade de plusieurs de ces éléments de base forme une structure arborescente symétrique ou asymétrique qui effectue une décomposition en B sous-bandes (B entier quelconque).

Étudions la décomposition d'un signal mono-dimensionnel en 2 sous-bandes, suivie de sa reconstruction pour établir les conditions que doivent satisfaire les filtres pour une reconstruction exacte, en l'absence de toute quantification et codage (système schématisé par la Figure 7-2):

Le signal d'entrée subit les opérations suivantes :

- Le signal discret d'entrée  $s[n]$  ( $n=0,\dots,N-1$ ) de transformée en  $z$ ,  $S(z)$ , est décomposé par les filtres d'analyse passe-bas  $Ha(z)$  et passe-haut  $Ga(z)$  en deux signaux  $a^0[n]$  et  $d^0[n]$ . Ils sont notés ainsi puisque, comme on va le constater plus loin,  $a$  représente une approximation du signal  $s$  alors que  $d$  représente les détails de ce signal.
- Ces signaux sont ensuite sous-échantillonnés par un facteur 2 pour donner, respectivement,  $a^{-1}[n]$  et  $d^{-1}[n]$ . L'indice supérieur (-1) de  $a$  et  $d$  représente la résolution du signal (Sa signification est donnée ultérieurement).
- À la reconstruction, ces signaux sont sur-échantillonnés par le même facteur en insérant un échantillon nul entre chaque couple d'échantillons.
- Une interpolation est ensuite effectuée pour donner aux échantillons ainsi insérés des valeurs proches de celles de leurs voisins. Cette opération est assurée par deux filtres dits de reconstruction

qui sont  $Hr(z)$  et  $Gr(z)$  :

- $Hr(z)$  filtre passe-bas s'appliquant sur le signal sur-échantillonné  $\tilde{a}^0[n]$ .
- $Gr(z)$  filtre passe-haut s'appliquant sur le signal sur-échantillonné  $\tilde{d}^0[n]$ . Les signaux résultants sont notés  $\tilde{a}[n]$  et  $\tilde{d}[n]$ .
- Le signal reconstruit  $\tilde{s}[n]$  est obtenu par addition de ces deux signaux:  $\tilde{s}[n] = 2[\tilde{a}[n] + \tilde{d}[n]]$ .

Ces opérations se traduisent par les équations du système suivantes (elles sont détaillées pour la partie basses-fréquences; les détails de celles des hautes fréquences sont obtenus en remplaçant  $a$  par  $d$ ,  $H$  par  $G$ , et  $A$  par  $D$ ):

Éq 7-1

$$A^0(z) = S(z).Ha(z) \quad \text{et} \quad a^{-1}[n] = a^0[2n]$$

avec, par définition,

$$\text{Éq 7-2} \quad A^0(z) = \sum_{n=0}^{N-1} a^0[n].z^{-n} = \sum_{p=0}^{N/2-1} a^0[2p].z^{-2p} + \sum_{p=0}^{N/2-1} a^0[2p+1].z^{-(2p+1)}$$

où le premier terme est la transformée en  $z$  des échantillons d'indice pair et le second celle des échantillons d'indice impair. D'où, l'on peut écrire  $A^0(-z)$  comme suit:

Éq 7-3

$$A^0(-z) = \sum_{p=0}^{N/2-1} a^0[2p].z^{-2p} - \sum_{p=0}^{N/2-1} a^0[2p+1].z^{-(2p+1)}$$

et par suite:

Éq 7-4

$$A^0(z) + A^0(-z) = 2 \sum_{p=0}^{N/2-1} a^0_{2p}.z^{-2p} = 2 \sum_{p=0}^{N/2-1} a_p^{-1}.(z^2)^{-p} = 2.A^{-1}(z^2)$$

Sans quantification ni codage, le signal de reconstruction sur-échantillonné  $\tilde{a}^0$  est donné par :

Éq 7-5

$$\begin{cases} \tilde{a}^0[2n] = a^{-1}[n] \\ \tilde{a}^0[2n+1] = 0 \end{cases}$$

ce qui donne:

Éq 7-6

$$\begin{aligned} \tilde{A}^0(z) &= \sum_{p=0}^{N/2-1} \tilde{a}^0[2p].z^{-2p} + \sum_{p=0}^{N/2-1} \tilde{a}^0[2p+1].z^{-(2p+1)} \\ &= \sum_{p=0}^{N/2-1} \tilde{a}^0[2p].z^{-2p} = \sum_{p=0}^{N/2-1} a^{-1}[p].(z^2)^{-p} = A^{-1}(z^2) \end{aligned}$$

Après filtrage par  $Hr(z)$ , on a:

Éq 7-7

$$\tilde{A}(z) = \tilde{A}^0(z).Hr(z) = A^{-1}(z^2).Hr(z)$$

et d'après l'Éq 7-4, nous obtenons:

Éq 7-8

$$\tilde{A}(z) = \frac{1}{2}[A^0(z) + A^0(-z)].Hr(z)$$

qui, d'après l'Éq 7-1, s'écrit:

$$\text{Eq 7-9} \quad \tilde{A}(z) = \frac{1}{2}[S(z).Ha(z) + S(-z).Ha(-z)].Hr(z)$$

De même, on trouve:

$$\text{Eq 7-10} \quad \tilde{D}(z) = \frac{1}{2}[S(z).Ga(z) + S(-z).Ga(-z)].Gr(z)$$

d'où:

$$\text{Eq 7-11} \quad \tilde{S}(z) = 2.[\tilde{A}(z) + \tilde{D}(z)] = [Ha(z)Hr(z) + Ga(z)Gr(z)].S(z) \\ + [Ha(-z)Hr(z) + Ga(-z)Gr(z)].S(-z)$$

Lorsque les réponses spectrales des filtres Ha et Ga utilisés se recouvrent (ce qui est généralement le cas), le sous-échantillonnage introduit un repliement spectral (ou "aliasing").

Selon l'Eq 7-11, le signal reconstruit est composé de deux termes:

- Le premier correspond à une version pondérée du signal original,
- Le second correspond au repliement spectral. Il est nul lorsque Ha et Ga ne se recouvrent pas c.à.d. Ha est nul lorsque Ga ne l'est pas et vice-versa.

À partir de l'Eq 7-11, on peut énoncer les conditions de reconstruction parfaite qui se traduit par:  $\tilde{S}(z) = S(z)$ ; d'où les conditions de reconstruction parfaite suivantes:

$$\text{Eq 7-12} \quad \begin{aligned} *Ha(z).Hr(z) + Ga(z).Gr(z) &= 1 \\ *Ha(-z).Hr(z) + Ga(-z).Gr(z) &= 0 \end{aligned}$$

Plusieurs solutions peuvent être proposées pour résoudre ce système de deux équations à quatre inconnus. Elles définissent des familles de filtres qui permettent une décomposition/reconstruction plus ou moins parfaite. Parmi les plus connues de ces solutions, on cite:

- Solution d'Esteban - Galand: les QMF
- Solution de Smith-Barnwell: les CQF
- Filtres à Noyau Court (SKF)

## 7.2 Solution d'Esteban - Galand: les QMF

Esteban et Galand, ont été les premiers, en 1977, à énoncer les conditions que doivent satisfaire les fonctions de transfert en z des filtres, pour obtenir une reconstruction exacte.

Leur solution a donné lieu à une classe de filtres FIR qui a été largement utilisée en codage sous-bandes mono-dimensionnel. Ce sont les "Filtres Miroirs en Quadrature": QMF's (Quadrature Mirror Filters). Ils sont nommés ainsi parce qu'ils sont formés d'une paire de filtres qui ont des gains symétriques (Figure 7-3) par rapport au quart de la fréquence de Shannon-Nyquist (effet miroir) et parce que leurs phases sont décalées de  $\pi/2$  radians (quadrature) – voir Remarque ci-dessous.

La solution proposée donne les relations suivantes entre les transformées en z des 4 filtres (ainsi que les relations entre leurs réponses impulsionales):

$$\text{Eq 7-13} \quad \begin{cases} Ga(z) = Ha(-z) \Leftrightarrow ga[n] = (-1)^n \cdot ha[n] \\ Hr(z) = Ha(z) \Leftrightarrow hr[n] = ha[n] \\ Gr(z) = -Ga(z) \Leftrightarrow gr[n] = -ga[n] \end{cases}$$

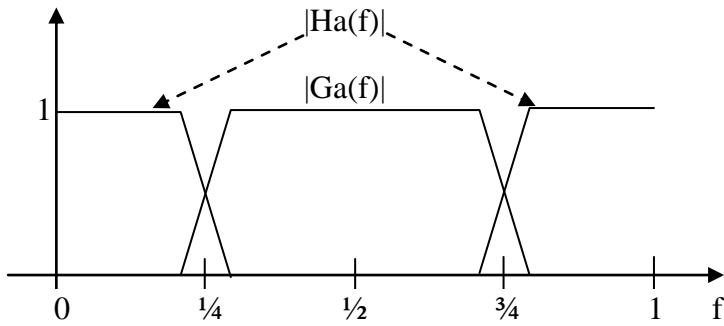


Figure 7-3. Représentation schématique des QMF en fréquence

Ceci fait disparaître le terme dû à l'aliasing du signal de sortie qui devient:

$$\text{Eq 7-14} \quad \tilde{S}(z) = [Ha^2(z) - Ga^2(z)].S(z)$$

D'où, une fonction de transfert globale du système égale à:  $T(z) = Ha^2(z) - Ga^2(z)$

**Remarque:** Dans le domaine fréquentiel, la relation entre les filtres d'analyse se traduit par :

$$Ga(\omega) = Ha(\omega - \pi) = Ha^*(\pi - \omega) \Rightarrow \begin{cases} |Ga(\omega)| = |Ha(\omega - \pi)| \\ Arg(Ga(\omega)) = Arg(Ha(\pi - \omega)) + \frac{\pi}{2} \end{cases}. \text{ Donc, Phase en quadrature}$$

Étant donné que  $Ha(z)$  et  $Ga(z)$  sont à phase linéaire, ce système n'introduit pas une distorsion de phase. Le problème restant est celui d'éliminer la distorsion d'amplitude en faisant  $|T(z)|=1$  pour tout  $z$ , afin d'obtenir la reconstruction exacte.

Le seul filtre à phase linéaire et à réponse impulsionnelle finie qui satisfait cette condition est de longueur 2. Les transformées en z des filtres correspondants sont:

$$\text{Eq 7-15} \quad Ha(z) = \frac{1}{2}(1+z^{-1}) \quad \text{et} \quad Ga(z) = \frac{1}{2}(1-z^{-1})$$

La capacité de séparation des bandes spectrales de ces deux filtres est très faible, ce qui réduit l'efficacité du codage des sous-bandes et limite ainsi leur utilisation dans le domaine de la compression.

Par conséquent, on constate qu'avec les hypothèses posées par la solution d'Esteban-Galand pour les filtres QMF, il est impossible d'avoir un système de décomposition /reconstruction en sous-bandes qui élimine à la fois la distorsion d'amplitude et la distorsion de phase.

Ainsi, la solution adoptée tolère une faible distorsion d'amplitude. Leur pouvoir séparateur entre les bandes spectrales augmente avec leur taille mais aux dépens d'une augmentation du coût de calcul.

L'utilisation des QMF a amélioré nettement la qualité du système lorsqu'on la compare à celle des systèmes antérieurs qui étaient basés sur les bandes de transition étroites et les paramètres de réjection des filtres de décomposition et de reconstruction pour minimiser les effets d'aliasing.

### 7.3 Solution de Smith-Barnwell: les CQF

Plus tard, en 1986, Smith et Barnwell ont proposé une solution analytique plus générale permettant l'élaboration des filtres permettant, cette fois, une reconstruction plus exacte dite parfaite; l'idée de base est la suivante: pour éliminer les trois types de distorsion (aliasing, Chaouki DIAB

d'amplitude, et de phase) simultanément, on doit admettre un simple retard R sur le signal de sortie par rapport à celui de l'entrée:

$$\text{Éq 7-16} \quad \tilde{s}[n+R] = s[n] \quad \text{pour } n=0, \dots, N-1$$

La solution qui en découle est la suivante:

$$\text{Éq 7-17} \quad \begin{cases} Ga(z) = -Ha(-z^{-1}) \cdot z^{-R} \Leftrightarrow ga[n] = (-1)^n \cdot ha[R-n] \\ Hr(z) = Ga(-z) \Leftrightarrow hr[n] = (-1)^n \cdot ga[n] \\ Gr(z) = -Ha(-z) \Leftrightarrow gr[n] = -(-1)^n \cdot ha[n] \end{cases}$$

Cette solution définit une autre classe des filtres appelés Filtres Conjugués en Quadrature (ou CQF: Conjugate Quadrature Filters). R est égal à L-1, L étant la longueur de chacun de ces filtres.

$H_a$  n'est pas nécessairement à phase linéaire; ce qui serait le cas des trois autres filtres. Ceci induit un inconvénient pour la compression qui souffre du fait que les réponses impulsionales de tels filtres ne sont pas symétriques.

Pour résoudre ce problème, une solution dérivée a été proposée pour élaborer des filtres à phase linéaire qui vérifient les relations de l'Éq 7-17. Ces filtres ne sont réalisables que pour de courtes longueurs ( $L \leq 4$ ), d'où leur nom de **Filtres à Noyau Court (SKF: Short Kernel Filters)**. Leur principal avantage est leur faible coût de calcul alors que leur inconvénient majeur pour la compression est leur faible pouvoir de séparation entre les sous-bandes spectrales.

## 7.4 Relation avec la représentation temps-fréquence

La théorie de la transformée en ondelettes a vu le jour avec, en particulier, les travaux de Y. Meyer et de I. Daubechies. Cette théorie est venue donner un aspect plus global de la représentation des signaux en sous-bandes. Celle-ci est alors appelée représentation temps-fréquence.

L'idée de représenter un signal à la fois en fonction du temps et de la fréquence vient du fait qu'une représentation en fréquence seulement, décrit le signal par projection sur une base de fonctions sinusoïdales vibrant sans amortissement sur tout l'axe du temps. Ces fonctions ne sont pas localisées dans le temps, même si le signal a une durée courte.

Une représentation en temps-fréquence, remplace ces sinusoïdes par d'autres fonctions élémentaires qui vibrent, comme des sinusoïdes, sur une certaine plage de temps mais s'amortissent très rapidement à l'extérieur de cette plage.

La fréquence de vibration est "locale": elle varie en fonction de l'instant temporel considéré. Ainsi ces fonctions élémentaires dépendent de deux paramètres  $j$  et  $n$  où  $j$  est lié à la fréquence et  $n$  au temps.

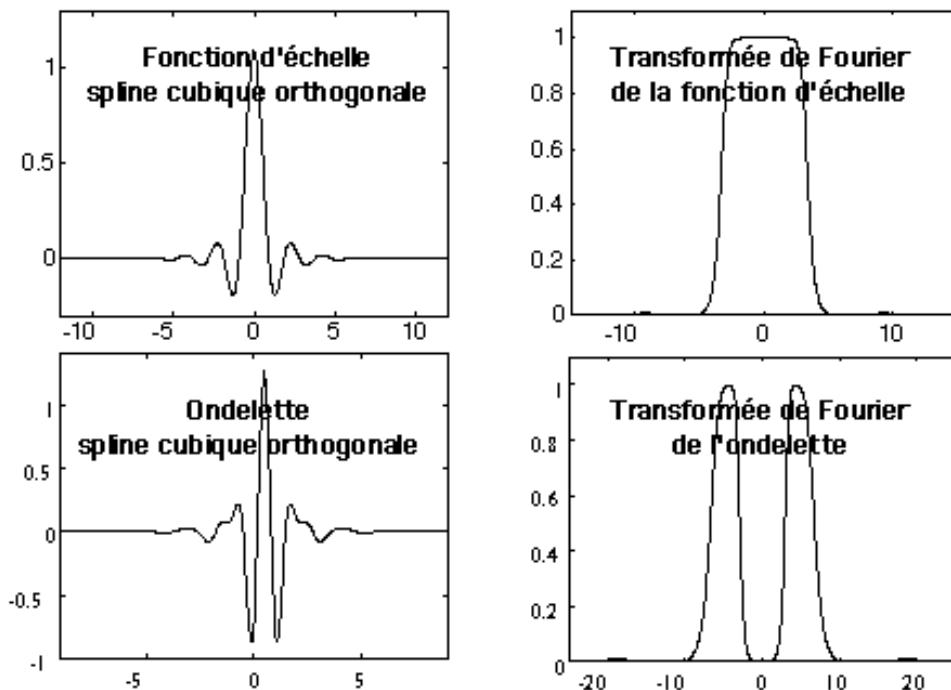
Les coefficients  $D_n^j$  que l'on affecte à chaque fonction élémentaire  $\psi_n^j$  lorsqu'on décompose un signal quelconque sur la base constituée par les fonctions  $\psi$ , donnent une information directe sur les propriétés temporelles et fréquentielles du signal.

La transformée en ondelettes est une forme de cette représentation temps-fréquence.

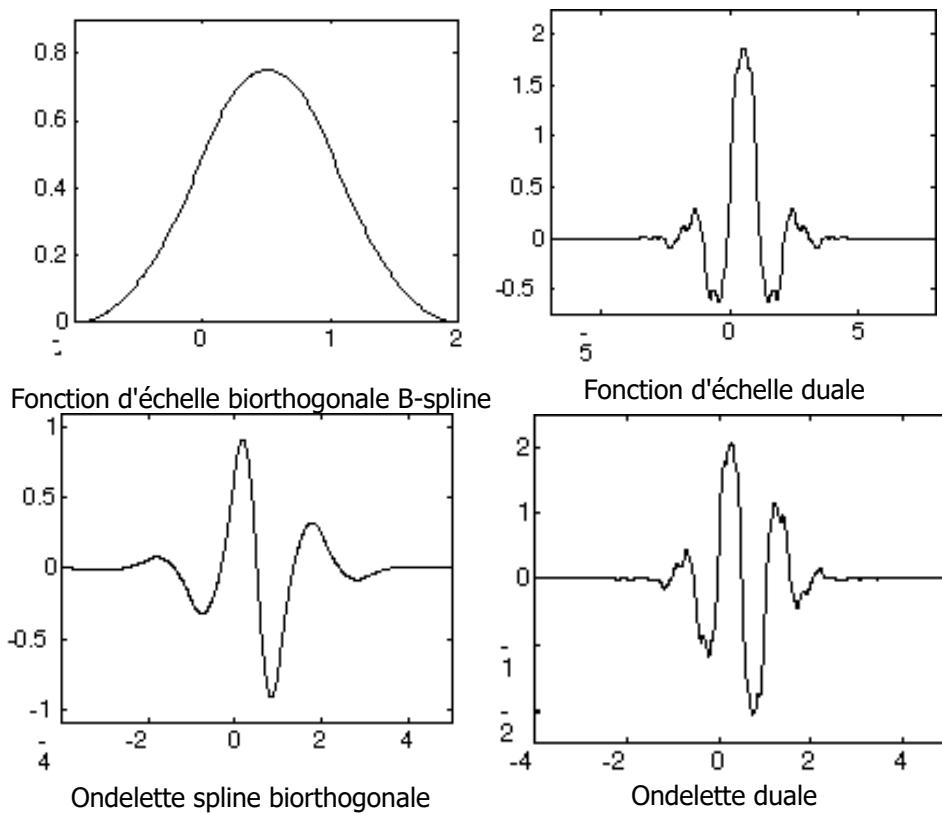
Notons que cette représentation s'applique aux images; l'espace des pixels remplace le temps. Dans ce cas on parle d'une représentation espace-fréquence.

## 7.5 Représentation par Transformée en Ondelettes

Les fonctions  $\psi_n^j$  constituent une famille de fonctions déduites à partir d'une fonction particulière  $\psi$  appelée ondelette; elle est une fonction oscillante de hautes fréquences.



a )



b )

Figure 7-4. Exemples de formes d'ondelettes et de fonctions d'échelle correspondantes: a) orthogonales b) bi-orthogonales

On définit également une famille des fonctions  $\phi_n^j$  complémentaire à celle des ondelettes.

Ces deux familles utilisent une notion très importante dans la transformation en ondelettes: c'est la notion de résolution.

**Notion de résolution:** Intuitivement, la résolution  $r$  donne la taille minimale des détails que l'on trouvera dans une approximation du signal original à cette résolution.

**Principe:** La décomposition en ondelettes consiste à calculer l'approximation du signal original à différentes résolutions, puis à calculer les signaux de détails perdus lors du passage d'une résolution donnée à une résolution inférieure.

Pour conserver une quantité de données constante, on démontre que la séquence des résolutions doit varier exponentiellement:  $(r^j)_{j \in \mathbb{Z}}$  ( $r > 1$ ). Pour simplifier le calcul, on choisit généralement  $r = 2$ .

On montre que pour tout  $j \in \mathbb{Z}$ , l'ensemble des fonctions  $\{\phi_n^j\}_{n \in \mathbb{Z}}$  forme une base orthonormée de l'espace  $V_j \subset L^2(\mathbb{R})$  contenant tous les signaux de résolution  $2^j$ ,  $L^2(\mathbb{R})$  étant l'ensemble des signaux réels à énergie finie.

Plus  $j$  est négatif, moins les détails sont visibles: les hautes fréquences disparaissent. Les fonctions  $\{\phi_n^j\}_{n \in \mathbb{Z}}$  sont obtenues à partir d'une fonction unique  $\phi(x)$  dite "fonction d'échelle" par une dilatation par  $2^j$  et une translation sur une grille de points séparés par des intervalles proportionnels à  $2^{-j}$ :

$$\text{Eq 7-18} \quad \phi_n^j(x) = \sqrt{2^j} \cdot \phi(2^j(x - 2^{-j}n)) = \sqrt{2^j} \cdot \phi(2^jx - n)$$

Ainsi, l'approximation d'un signal  $s(x)$  à la résolution  $2^j$ , est la projection orthogonale de  $s(x)$  sur  $V_j$ . Elle est caractérisée par l'ensemble d'échantillons donnés par:

$$\text{Eq 7-19} \quad S_n^j = \langle s(x), \phi_n^j(x) \rangle = [s(x) * \phi^j(-x)](2^{-j}n)$$

où  $\langle \cdot, \cdot \rangle$  symbolise le produit scalaire dans  $L^2(\mathbb{R})$  et  $*$  le produit de convolution. C'est aussi, donc, le produit de convolution de  $s(x)$  par la fonction de  $\phi(x)$  dilatée par  $2^j$ , au point  $2^{-j}n$ , avec  $\phi^j(x) = \sqrt{2^j} \phi(2^jx)$ .

Le passage d'une résolution  $2^{j+1}$  à une résolution  $2^j$  s'accompagne d'une perte de détails dont la représentation doit se faire dans l'espace vectoriel  $O_j$  orthogonal à  $V_j$ .

L'ensemble  $\{\psi_n^j\}_{n \in \mathbb{Z}}$  forme une base orthonormée de l'espace  $O_j$ . Ainsi l'ensemble des échantillons  $D_n^j$  représente les détails du signal  $s(x)$  à la résolution  $2^j$ . Les fonctions  $\{\psi_n^j\}_{n \in \mathbb{Z}}$  découlent de l'ondelette  $\psi(x)$  par une dilatation par  $2^j$  et une translation sur une grille dont l'intervalle entre les points est proportionnel à  $2^{-j}$ .

**Relation entre  $\phi(x)$  et  $\psi(x)$ :** Le lien entre la fonction d'échelle  $\phi(x)$  et l'ondelette correspondante  $\psi(x)$  est établi par la relation:

$$\text{Éq 7-20} \quad \Psi(\omega) = H\left(\frac{\omega}{2} + \pi\right) \cdot \Phi\left(\frac{\omega}{2}\right) e^{-i\omega/2}$$

où  $\Psi(\omega)$  et  $\Phi(\omega)$  sont les transformées de Fourier de  $\psi(x)$  et  $\phi(x)$  respectivement et  $H(\omega)$  est celle d'un filtre  $h$  de réponse impulsionnelle:

$$\text{Éq 7-21} \quad h(n) = \frac{1}{\sqrt{2}} \langle \phi_0^{-1}, \phi_n^0 \rangle$$

**Filtres d'implantation:**  $h$  est un filtre passe-bas ayant les deux propriétés suivantes:

- a)  $H(\omega)$  est une fonction différentiable et de période  $2\pi$  avec  $|H(0)| = 1$
- b)  $|H(\omega)|^2 + |H(\omega+\pi)|^2 = 1$ .

Si on considère  $S^0$  comme l'approximation à la résolution  $j=0$ , du signal continu original  $s(x)$ , obtenue par un échantillonnage approprié, alors son approximation  $S^{-1}$  à la résolution  $j=-1$ , sera obtenue par une convolution de  $S^0$  avec  $h(-n)$  et en ne retenant qu'un échantillon sur 2.

Plus généralement,  $S^j$  peut être obtenu à partir de  $S^{j+1}$  en le convoluant avec  $h(-n)$  et en ne retenant qu'un échantillon sur 2.

Les détails  $D^j$  sont obtenus aussi à partir de  $S^{j+1}$  par une convolution avec un filtre  $g(-n)$ , et en ne retenant qu'un échantillon sur 2. Ce filtre passe-haut a pour réponse impulsionnelle:

$$\text{Éq 7-22} \quad g(n) = \frac{1}{\sqrt{2}} \langle \psi_0^{-1}, \psi_n^0 \rangle$$

$S^{j+1}$  est ainsi décomposé en deux sous-signaux  $S^j$  et  $D^j$  où  $S^j$  est son filtré passe-bas et  $D^j$  son filtré passe-haut, tous deux sous-échantillonés par un facteur 2. On notera la conservation de la quantité de données.

L'ensemble des coefficients  $\{S^j, (D^j)_{j \leq j \leq -1}\}$  définit la représentation en ondelettes de  $S^0$  à la résolution  $2^j$  ( $j < 0$ ).

### Lien avec les sous-bandes:

Ainsi la représentation en ondelettes d'un signal  $S^0$  à une résolution -1 consiste à le décomposer en deux sous-signaux  $S^{-1}$  et  $D^{-1}$  qui sont les résultats de deux filtrages passe-bas et passe-haut, respectivement, du signal original suivis d'un sous-échantillonnage par un facteur 2. Ce n'est autre qu'une décomposition en deux sous-bandes avec des filtres d'analyse générés à partir d'une ondelette.

Cette décomposition en ondelettes se distingue de celle en sous-bandes classique par le fait que les Ha et Ga obtenus à partir des ondelettes sont tels que  $S^{-1}$  et  $D^{-1}$  sont, respectivement, les projections orthogonales de  $S^0$  dans les deux sous-espaces  $V_{-1}$  et  $O_{-1}$  orthogonaux, ayant tous deux, une base de fonctions orthonormées obtenues par translations et dilatations de la fonction d'échelle  $\phi$  ou de l'ondelette  $\psi$  correspondante.

C'est un avantage important pour l'efficacité de codage et de compression, non satisfait par les filtres classiques tels les QMF et les CQF. Les filtres d'analyse et de synthèse calculés à partir des Chaouki DIAB

ondelettes vérifient les relations suivantes:

- 1)  $ha(n) = \frac{1}{\sqrt{2}} \langle \phi_0^{-1}, \phi_n^0 \rangle$
- 2)  $|Ha(\omega)|^2 + |Ha(\omega+\pi)|^2 = 1$ . C'est la condition nécessaire et suffisante pour avoir les deux bases orthonormées de  $V_j$  et  $O_j$ .
- 3)  $Ga(\omega) = e^{-i\omega} Ha^*(\omega+\pi)$  où  $Ha^*$  est le conjugué de  $Ha$ .  $\Rightarrow ga(n) = (-1)^{1-n} \cdot ha(1-n)$
- 4)  $hr(n) = ha(-n) \Rightarrow Hr(\omega) = Ha^*(\omega)$
- 5)  $gr(n) = ga(-n) \Rightarrow Gr(\omega) = Ga^*(\omega)$
- 6) La fonction de transfert est alors:  $T(\omega) = Ha(\omega)Hr(\omega) + Ga(\omega)Gr(\omega)$   
 $= Ha(\omega)Ha^*(\omega) + Ga(\omega)Ga^*(\omega) = |Ha(\omega)|^2 + |Ga(\omega)|^2$   
or,  $|Ga(\omega)|^2 = |Ha^*(\omega+\pi)|^2 = |Ha(\omega+\pi)|^2 \Rightarrow T(\omega) = |Ha(\omega)|^2 + |Ha(\omega+\pi)|^2 = 1$

La reconstruction est donc parfaite [ $T(\omega) = 1$ ] avec des filtres  $H$  et  $G$  non symétriques et à réponse impulsionnelle, théoriquement infinie mais pratiquement finie puisqu'elle s'amortit rapidement.

## 7.6 Cas d'images

On peut facilement généraliser la décomposition en ondelettes à des signaux de dimension supérieure à 2. Nous allons nous limiter ici au cas de l'image. À la fonction d'échelle mono-dimensionnelle  $\phi(x)$  se substitue une fonction d'échelle bi-dimensionnelle  $\phi(x,y)$  qui peut être séparable:  $\phi(x,y) = \phi(x)\phi(y)$ .

L'espace des détails  $O_j$  à la résolution  $j$ , a alors pour fonctions de base:

Eq 7-23  $\{\phi_n^j(x), \psi_m^j(y), \psi_n^j(x), \phi_m^j(y), \psi_n^j(x), \psi_m^j(y)\}_{(m,n) \in \mathbb{Z}^2}$

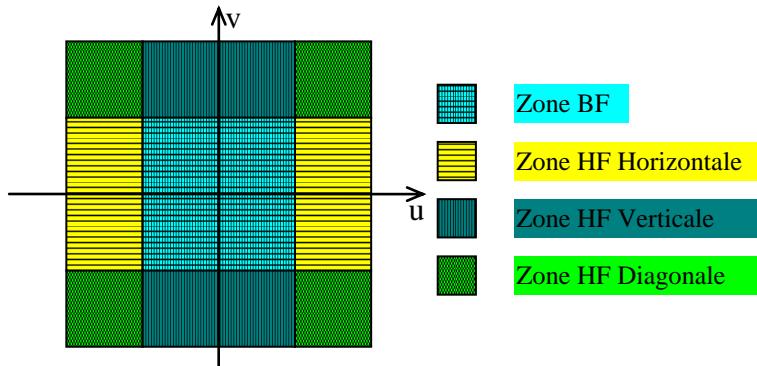


Figure 7-5. Partition du plan fréquentiel pour une décomposition 2D en 4 sous-bandes

La "différence" entre la sous-image  $S^{j+1}$  et la sous-image  $S^j$  est donnée par trois sous-images de détails  $D^{jH}, D^{jV}, D^{jD}$  qui sont, respectivement, les détails suivant les directions horizontale, verticale et diagonale.

Les sous-images  $S^j, D^{jH}, D^{jV}, D^{jD}$  sont obtenues à partir de  $S^{j+1}$ , par une double convolution avec les filtres  $h(-n)$  et  $g(-n)$  suivie d'un sous-échantillonnage par un facteur 2, appliquée aux lignes de  $S^{j+1}$  puis aux colonnes.

Pour une résolution  $J < 0$ , la représentation en ondelettes de l'image  $S^0$  est l'ensemble des sous-images suivantes:  $\{S^J, (D^{jH}, D^{jV}, D^{jD})_{J \leq j \leq -1}\}$

Le total des pixels de ces sous-images est égal au nombre de pixels de l'image originale  $S^0$ , puisque une sous-image de résolution  $2^{-j}$  a un nombre de pixels égal au nombre de pixels de l'image originale divisé par  $2^{2j}$ .

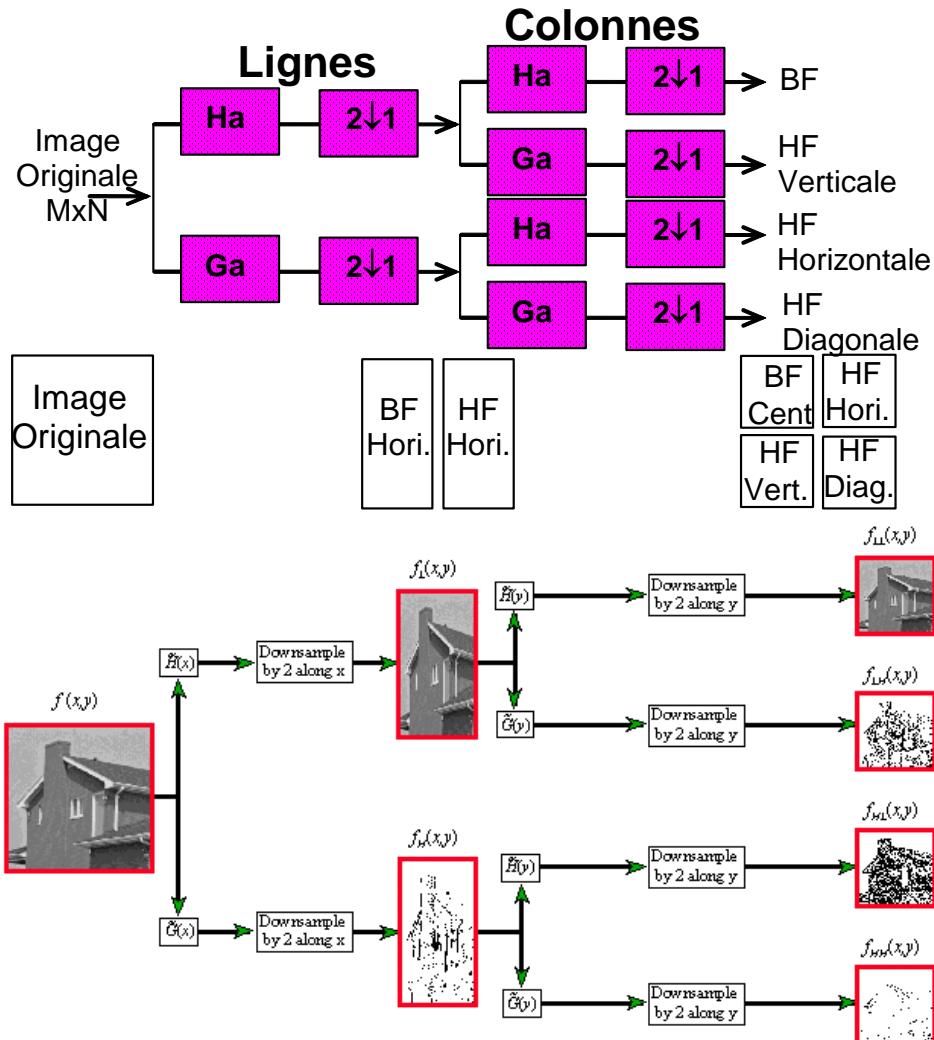


Figure 7-6. Schéma-bloc de décomposition 2D en 4 sous-bandes

Ainsi, la représentation en ondelettes à une résolution  $2^{-1}$  d'une image produit quatre sous-images dont l'une est l'approximation de l'originale à cette résolution et les 3 autres représentent les détails directionnels perdus lors de passage de la résolution  $2^0$  à  $2^{-1}$ .

La même représentation est obtenue en utilisant la méthode de décomposition en sous-bandes qui consiste à appliquer les filtres  $h_a$  et  $g_a$  sur les  $M$  lignes de l'image suivie d'un sous-échantillonnage par un facteur 2, ce qui donne deux sous-images de  $M$  lignes et de  $N/2$  colonnes chacune ( $N$  étant le nombre de colonnes de l'image originale). Ensuite, on applique les mêmes filtres sur les  $N$  colonnes de ces deux sous-images, ce qui donne finalement 4 sous-images de  $M/2 \times N/2$  pixels chacune.

La première, qui correspond à la sous-bande de basses-fréquences, est l'approximation de l'image à la résolution  $2^{-1}$ , et les trois autres correspondent aux sous-bandes de hautes fréquences selon l'axe horizontal et selon l'axe vertical et selon les deux axes (diagonal) respectivement comme le montre la Figure 7-5.

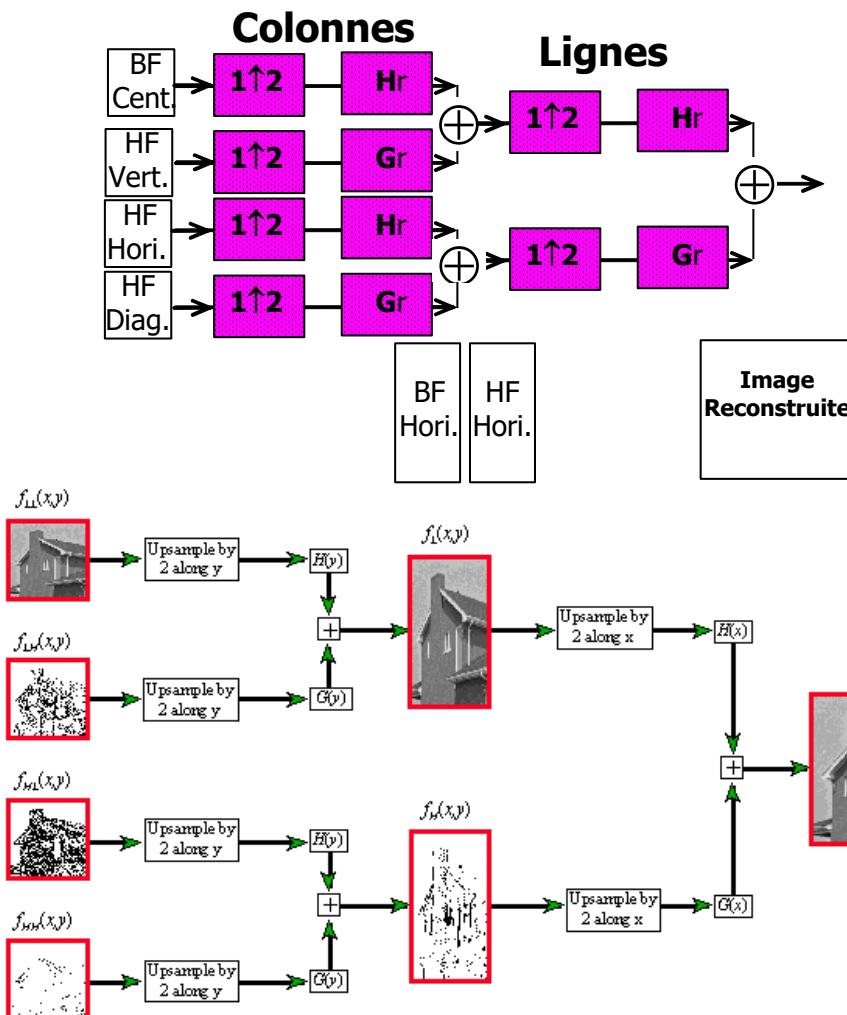


Figure 7-7. Reconstruction 2D à partir de 4 sous-bandes

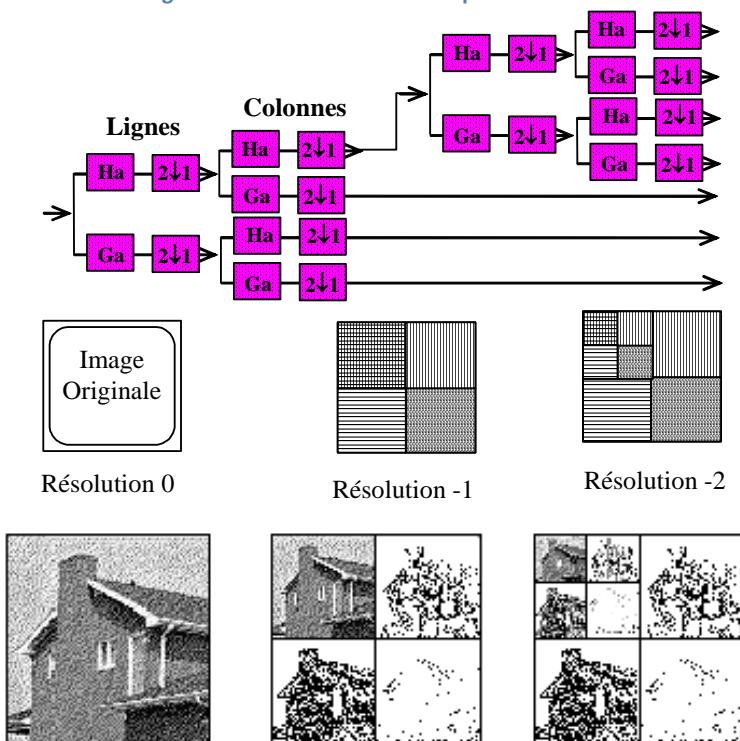


Figure 7-8. Décomposition 2D pyramidale à la résolution  $2^{-2}$  (7 sous-bandes)

Le système de décomposition correspondant est schématisé sur la Figure 7-6. Le système de reconstruction correspondant est donné par la Figure 7-7.

On peut aller plus loin, en décomposant la sous-bande de basses fréquences en 4 sous-bandes de résolution inférieure (Figure 7-8 et Figure 7-9) pour obtenir une décomposition pyramidale.

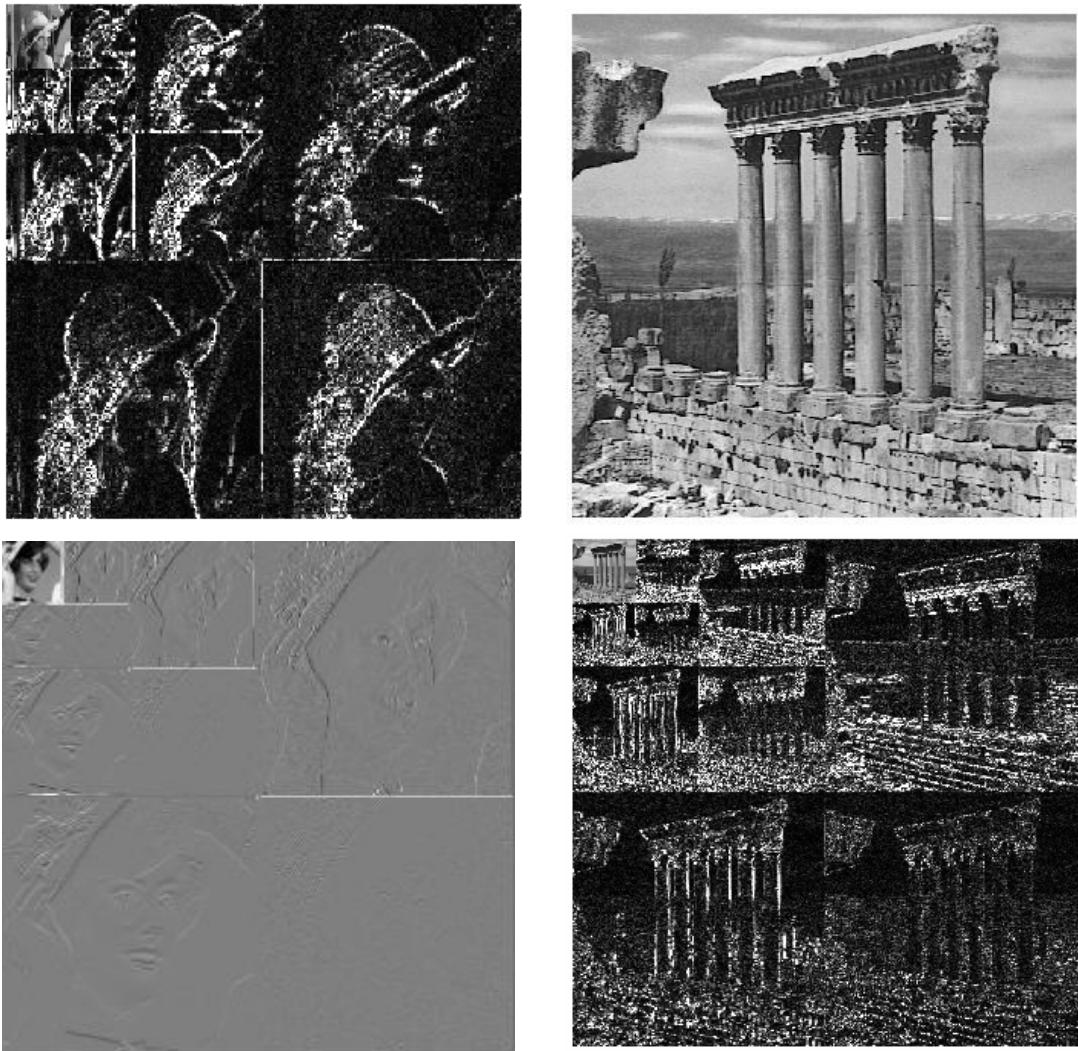


Figure 7-9. Exemples d'images décomposées pyramidalement jusqu'à la résolution  $2^3$

## Application de la DWT

Dans le standard JPEG2000, la DWT est opérée sur des "tiles" de chacune des composantes; Deux transformées en ondelettes réversible et non réversible sont utilisées. La transformée réversible associe aux valeurs entières des pixels de l'image des coefficients entiers, en revanche la transformée non réversible leur associe des coefficients réels à virgule flottante.

La transformée réversible est implantée avec une technique dite de "Lifting" au lieu du filtrage par convolution. Le nombre de niveaux de décomposition est un paramètre de la transformée. Une valeur typique de ce paramètre est 6 (pour une image suffisamment large).

Les filtres d'analyse utilisés par JPEG2000 partie 1 sont :

- soit les filtres de Daubechies 9/7 pour un codage irréversible avec pertes
- soit les filtres de Daubechies 5/3 pour un codage réversible sans pertes.

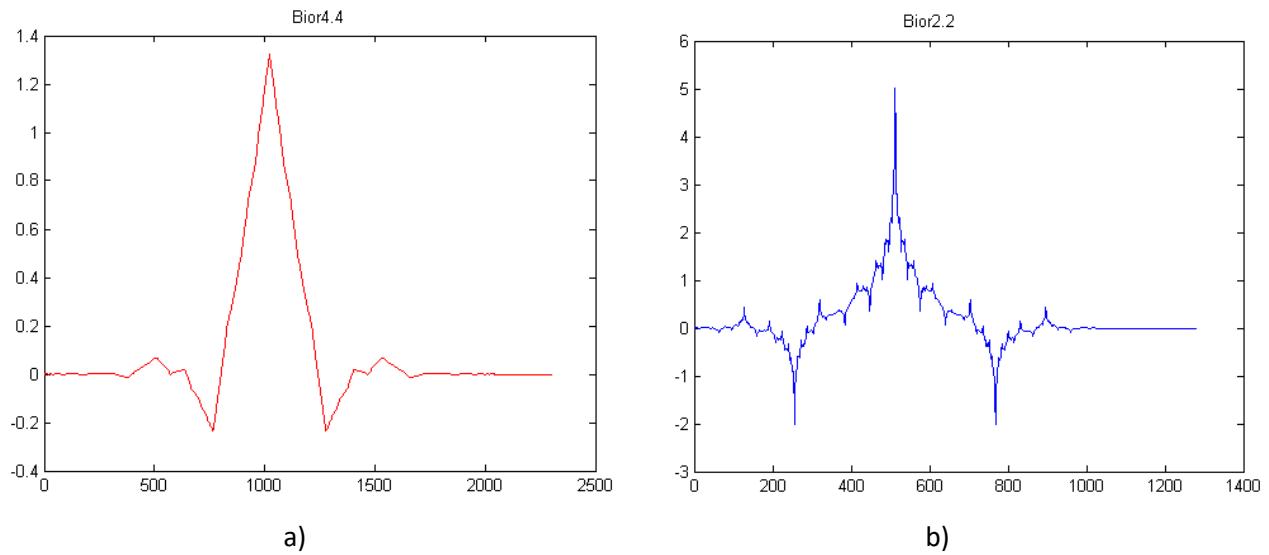


Figure 7-10 Formes d'ondelettes biorthogonales de Daubechies

Tableau 5 : Coefficients des filtres d'analyse et de synthèse conçus à partir des ondelettes de Daubechies 9/7 et 5/3

Daubechies 9/7:		<i>Les coefficients des filtres d'analyse ...</i>		
n	Passe-bas, $h_a(n)$	Passe-haut, $g_a(n)$		
0	+0.602949018236360	+1.115087052457000		
$\pm 1$	+0.266864118442875	-0.591271763114250		
$\pm 2$	-0.078223266528990	-0.057543526228500		
$\pm 3$	-0.016864118442875	+0.091271763114250		
$\pm 4$	+0.026748757410810			
<i>... et ceux des filtres de synthèse</i>				
n	Passe-bas, $h_r(n)$	Passe-haut, $g_r(n)$		
0	+1.115087052457000	+0.602949018236360		
$\pm 1$	+0.591271763114250	-0.266864118442875		
$\pm 2$	-0.057543526228500	-0.078223266528990		
$\pm 3$	-0.091271763114250	+0.016864118442875		
$\pm 4$		+0.026748757410810		
Daubechies 5/3: <i>Les coefficients des filtres d'analyse</i>		<i>Les coefficients des filtres de synthèse</i>		
n	Passe-bas, $h_a(n)$	Passe-haut, $g_a(n)$	Passe-bas, $h_r(n)$	Passe-haut, $g_r(n)$
0	6/8	1	1	6/8
$\pm 1$	2/8	-1/2	1/2	-2/8
$\pm 2$	-1/8			-1/8

Les premiers sont constitués d'un filtre passe bas à 9 coefficients et d'un filtre passe haut à 7 coefficients tous deux à coefficients irrationnels. Les seconds sont constitués d'un filtre passe bas à 5 coefficients et d'un filtre passe haut à 3 coefficients tous deux à coefficients rationnels.

Les filtres de Daubechies 9/7 sont générés à partir d'une ondelette biorthogonale de Daubechies (connue en Matlab par 'bior4.4') représentée sur la Figure 7-10 a). Les coefficients de filtres d'analyse et de synthèse sont donnés par le Tableau 5.

Pour le cas de compression sans perte, la DWT est implantée avec les filtres 5/3 où les résultats sont arrondis aux valeurs entières. En Matlab, l'ondelette correspondante est nommée 'bior2.2'; elle est représentée sur la Figure 7-10 b), et les coefficients de filtres d'analyse et de synthèse sont donnés dans le Tableau 5.

### Décodage progressif

La construction de l'image à partir des sous-bandes peut se faire progressivement, d'autant

plus que le caractère "multi-résolution" de ces sous-images s'apprête parfaitement à ce décodage progressif. L'affichage de l'image décodée peut s'effectuer en 2 modes, comme le montre l'illustration de la Figure 7-11.

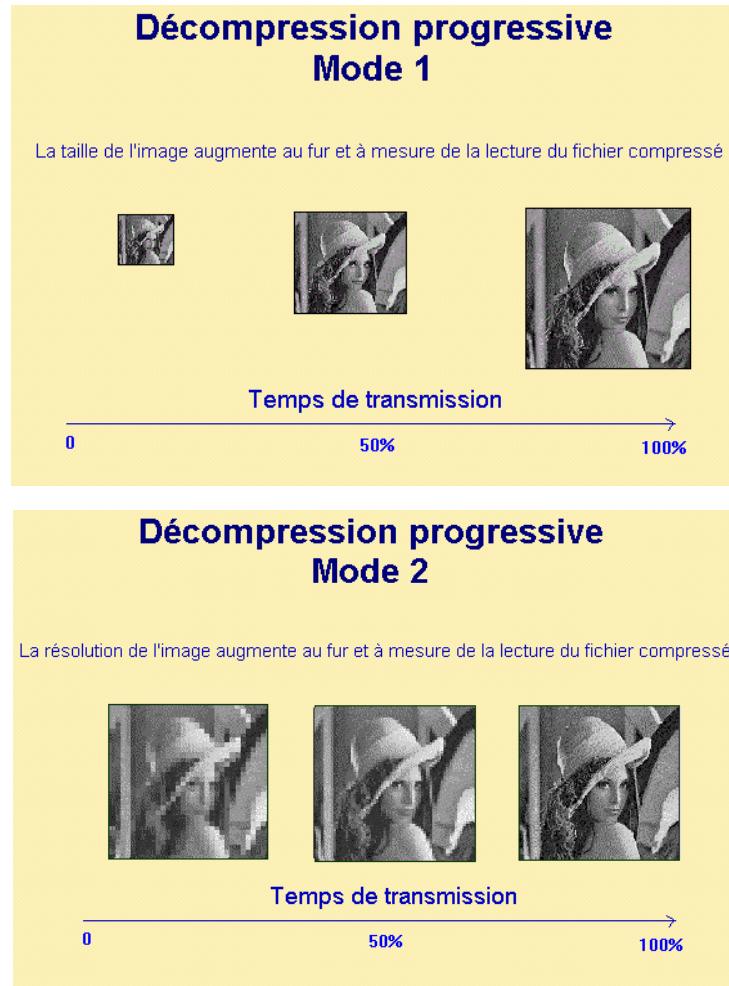


Figure 7-11. Modes de Décodage progressif

### Problème de bords:

Les filtres à reconstruction exacte sont conçus pour des signaux infinis tels que la parole. Or, pour un signal fini de M échantillons, le filtrage numérique par convolution produit un signal de  $M + L - 1$  échantillons, L étant la taille du filtre. Cette augmentation, due au régime transitoire de l'opération de filtrage, ne permet pas, lors de la décomposition d'une image en sous-bandes, de conserver la même quantité de pixels que dans l'image originale. Or, ceci réduit l'efficacité de la compression puisque la décomposition d'une image de  $M \times N$  pixels en 4 sous-bandes par des filtres

de taille L produit 4 sous-images de  $\frac{M+L}{2} \times \frac{N+L}{2}$  pixels chacune.

La solution adoptée pour conserver la même quantité d'échantillons qu'au départ, consiste à tronquer les signaux filtrés en négligeant les échantillons qui correspondent au régime transitoire. Ceci induit une perte d'information qui se traduit par une erreur de reconstruction. On montre que cette erreur sera concentrée sur les bords de l'image reconstruite. Cette erreur s'imprègne dans l'image proportionnellement à la taille du filtre, mais aussi, au degré de résolution atteint par l'image décomposée: Plus le nombre de sous-images tronquées est grand plus l'erreur est importante. La Figure 7-12 illustre ce phénomène dans le cas d'une décomposition en 4 sous-bandes.

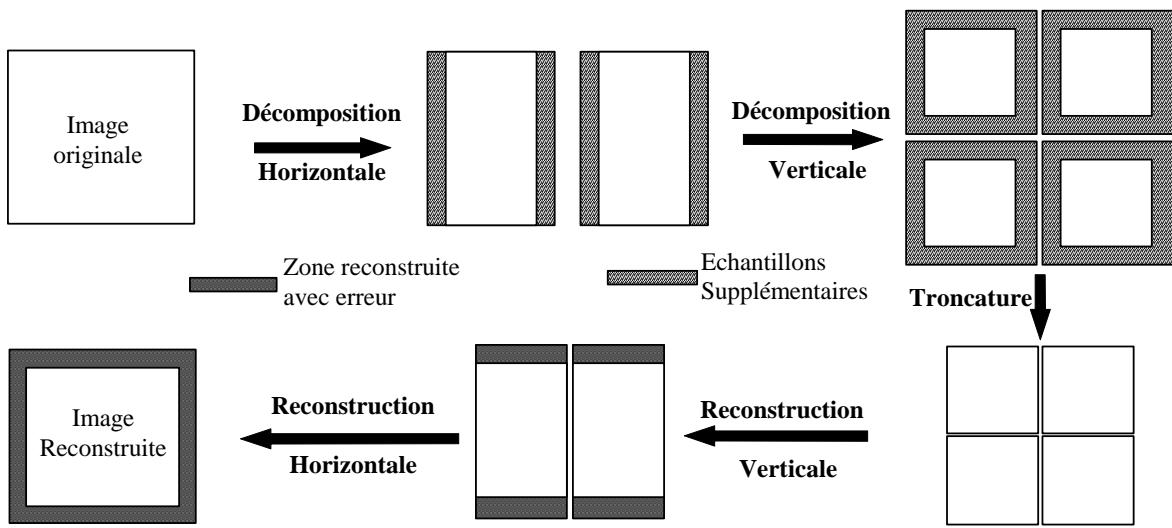


Figure 7-12. Illustration des erreurs de reconstruction sur les bords

Cette erreur de bords réduit considérablement la qualité de l'image reconstruite, surtout lorsque l'on décompose l'image à une faible résolution.

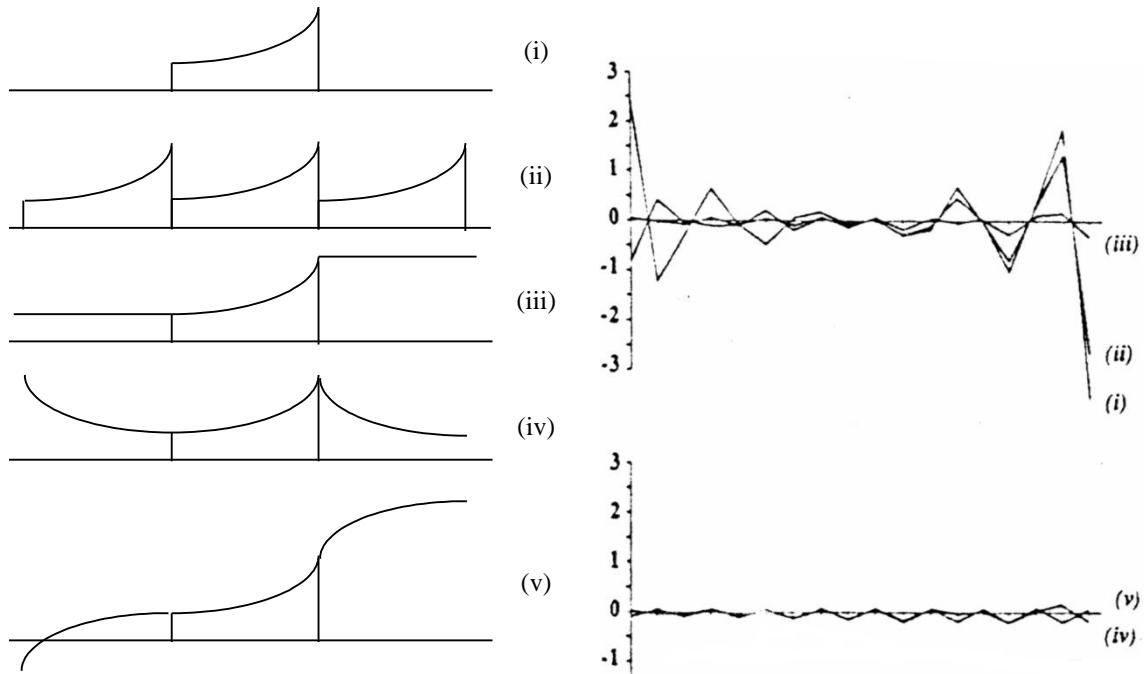


Figure 7-13. Exemples d'extensions proposées et erreurs de reconstruction associées

Plusieurs remèdes sont proposés pour réduire l'effet de ce phénomène. Ces remèdes consistent à prolonger de part et d'autre le signal de façon à assurer sa continuité ou sa périodicité; Bien qu'elles n'éliminent pas complètement l'effet de bords, certaines de ces solutions réduisent d'une manière satisfaisante l'erreur sur les bords.

Parmi les extensions proposées, on en présente quelques-unes sur la Figure 7-13, ainsi qu'un exemple de l'erreur résiduelle associée à chacune d'elles lorsqu'elles sont testées sur un signal fini en utilisant des filtres à reconstruction parfaite. On remarque que l'extension qui assure la continuité du signal avec une certaine symétrie par rapport au premier et au dernier échantillon, réduit au mieux l'erreur sur les bords.

# Chapitre 8 - Applications au Traitement d'images

## 8.1 Introduction

Le traitement numérique d'images comporte :

- les techniques d'amélioration de la qualité visuelle d'images dégradées ou déformées
- les techniques d'analyse d'images telles que l'extraction des contours, la segmentation et l'étude des textures.
- les techniques de compression
- les techniques de synthèse d'images.

Dans ce chapitre, nous présentons quelques méthodes de traitements de base telles que le zoom numérique (réduction et agrandissement), la réduction du bruit ou la détection de contours.

## 8.2 Exemples de traitement d'images

### Zoom: Réduction et Agrandissement

Pour réduire les dimensions d'une image d'un facteur  $r:1$  dans les deux directions, on doit remplacer chaque  $r \times r$  pixels de l'image originale par un seul pixel dont la valeur est la moyenne de ces  $r \times r$  pixels.

Pour agrandir les dimensions d'une image d'un facteur  $1:r$  dans les deux directions, il faut répéter chaque pixel  $r$  fois dans une direction, puis dans l'image résultante, répéter chaque pixel  $r$  fois dans l'autre direction. Les valeurs de ces pixels ajoutés peuvent avoir soit la même valeur que le pixel répété dans chaque direction, soit une valeur interpolée entre les pixels de l'image originale.

Une illustration des effets de la réduction et de l'agrandissement est présentée sur la Figure 8-1.

### Filtrage Linéaire

Comme dans le cas du filtrage mono-dimensionnel, le filtrage linéaire dans le cas d'image consiste à remplacer la valeur de chaque pixel dans l'image d'entrée par une nouvelle valeur dans l'image filtrée.

Cette valeur filtrée résulte d'un produit de convolution de la matrice caractéristique du filtre (généralement appelée noyau ou masque, équivalente à la réponse impulsionnelle d'un filtre linéaire mono-dimensionnel) avec l'image à filtrer, lorsque le masque est centré sur le pixel en question.

Le masque doit être glissé sur tous pixels de l'image d'entrée l'un après l'autre. Pour chaque pixel, le produit de convolution consiste à multiplier la valeur de l'élément du masque par la valeur du pixel correspondant de l'image d'entrée. La somme des produits donne la valeur de sortie du filtre pour le pixel courant sur lequel le masque est centré. Ceci se traduit par l'expression de convolution bidimensionnelle suivante :

$$A_f[i, j] = h * A[i, j] = \sum_{k=-m}^m \sum_{l=-n}^n h[k, l] \cdot A[i - k, j - l] \text{ pour } i=0, \dots, M-1 \text{ et } j=0, \dots, N-1$$

où  $h$  est le noyau d'un filtre de taille  $(2m+1) \times (2n+1)$ ,  $A$  est la matrice de l'image à filtrer de taille  $M \times N$

et  $A_f$  est la matrice résultante de l'image filtrée. Les dimensions du noyau sont généralement choisies impaires afin de faciliter le repérage du centre du masque.

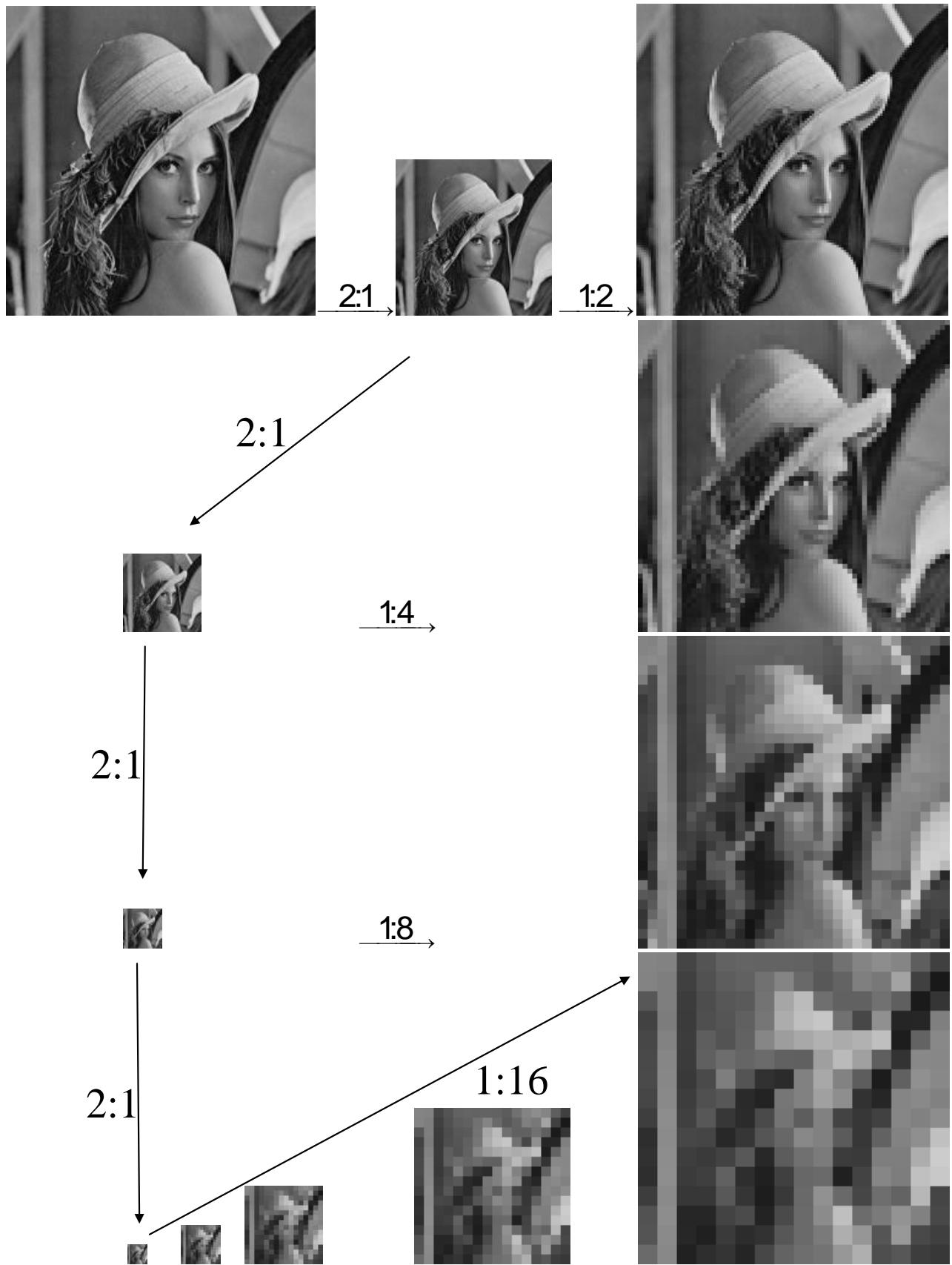


Figure 8-1. Illustration des effets des zooms numériques

### Réduction du bruit ou Lissage spatial:

Les régions plus ou moins uniformes dans une image se caractérisent par leur intensité moyenne.

Les fluctuations autour de cette intensité moyenne peuvent provenir soit du dispositif d'acquisition (caméra, amplificateur, quantification,...), soit de la scène elle-même (poussières, rayures,...), soit de perturbations externes.

Ces fluctuations sont généralement désignées sous le terme **bruit d'image** (Figure 8-2).

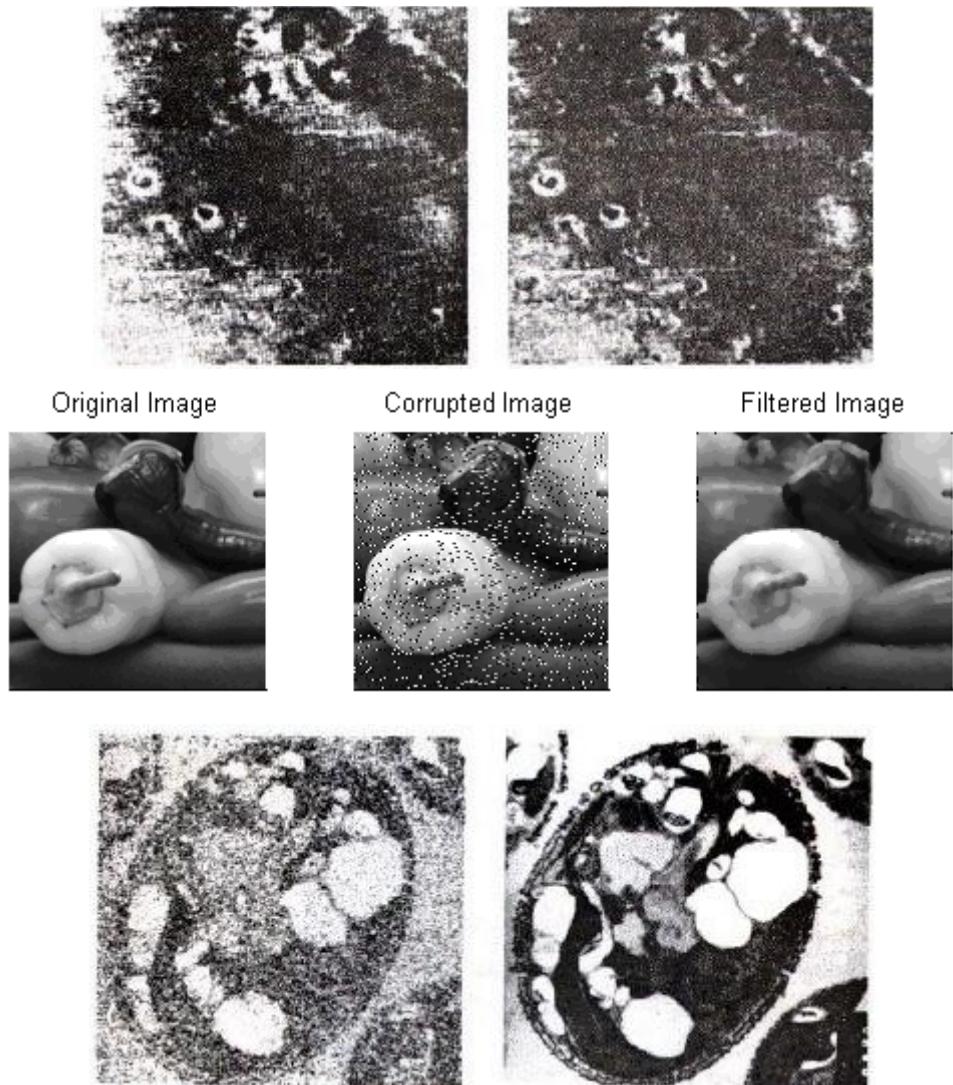


Figure 8-2. Images bruitées et filtrées

L'échelle spatiale des fluctuations est relativement faible par rapport aux dimensions des régions. Le bruit est donc plutôt de type haute fréquence. Dans la plupart des cas, le bruit d'image est considéré aléatoire, centré et additif.

La diminution du bruit se ramène donc à un problème de traitement du signal qui consiste à retrouver par filtrage les niveaux d'intensité nominaux de chaque région de l'image et ceci en réduisant l'amplitude des variations d'intensité dans chaque région, tout en conservant les transitions entre régions adjacentes. Cette préoccupation de conserver les transitions apparaît surtout dans les différentes approches de segmentation.

Les méthodes les plus simples et les plus faciles à implanter sont fondées sur le filtrage linéaire stationnaire (invariant par translation).

Un lissage de l'image par des filtres linéaires passe-bas a l'inconvénient de ne pas faire la distinction entre les transitions et les composantes du bruit, car elles contiennent toutes les deux de l'énergie en hautes fréquences.

En éliminant les composantes de hautes fréquences, les filtres linéaires élargissent les zones de contours. La recherche d'un filtre qui permet de supprimer le bruit en conservant les discontinuités du signal a conduit aux techniques de filtrage non linéaires.

Nous présentons d'abord le filtre moyenneur, qui est le filtre linéaire le plus connu pour lisser les images, avant d'examiner les différentes techniques de filtrage non linéaire réductrices du bruit.

### Le Filtre moyenneur

Un filtre moyenneur est caractérisé par un noyau dont les valeurs sont égales et dont la somme est égale à 1. Ainsi la valeur de chaque élément du noyau vaut  $\frac{1}{(2m+1)(2n+1)}$ .

Ceci signifie que la sortie du filtre moyenneur n'est autre que la valeur moyenne des valeurs des pixels du voisinage englobé par le masque du filtre.

Donc, un filtre moyenneur est défini par les dimensions de son noyau  $(2m+1) \times (2n+1)$ .

#### EXEMPLE 1 : LISSAGE PAR UN FILTRE MOYENNEUR

Considérons le filtre moyenneur de taille 3x3 ( $m=n=1$ ), et prenons une partie de l'image de 5x5 pixels qu'on note ainsi:

P1	P2	P3	P4	P5
P6	P7	P8	P9	P10
P11	P12	P13	P14	P15
P16	P17	P18	P19	P20
P21	P22	P23	P24	P25

Le filtre moyenneur 3x3 a le masque suivant:  $\frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

Calculons par exemple la valeur qui remplace celle du pixel P12 dans l'image lissée:

$$P12' = \frac{1}{9} (P6 + P7 + P8 + P11 + P12 + P13 + P16 + P17 + P18).$$

pour lequel on a centré le masque sur le pixel P12 et on a multiplié chacune des valeurs des pixels masqués par  $1/9$ .

En glissant la fenêtre sur chacun des pixels de l'image, on peut calculer les nouvelles valeurs de ces pixels pour obtenir l'image lissée par ce moyenneur 3x3, sachant que pour les pixels-frontières où certains pixels voisins ne sont pas définis, deux solutions peuvent être adoptées:

- soit on affecte la valeur du pixel le plus proche à chacun des pixels non définis,
- soit on les considère nuls (égaux à 0).

L'inconvénient de ce filtre est d'introduire un flou au niveau des zones de transition. Il est utilisé généralement dans les zones pour lesquelles la variance est faible.

### Filtres non-linéaires:

#### *Le filtre sigma:*

Ce filtre permet d'éviter le "moyennage" des zones de transition en ne faisant intervenir dans la moyenne que les points du voisinage dont la valeur est proche du point central. Si de plus le nombre de ces points est trop faible, on considère le point central comme bruité, et il est remplacé par la moyenne de ses voisins.

#### *Le V-filtre:*

Ce filtre partage le voisinage du pixel central en quatre quadrants. Il affecte au point central la valeur de la moyenne dans le quadrant où la variance est la plus faible.

#### *Le filtre de Nagao:*

Il repose sur le même principe que le V-filtre mais le voisinage considéré est une fenêtre, sur laquelle sont définis huit domaines de 7 pixels et un domaine de 9 pixels (Figure 8-3). Les domaines D1, D2 et D3 se déduisent de D0 par une rotation de  $\pi/2$ . Les domaines D5, D6 et D7 se déduisent de D4 par une rotation de  $\pi/2$ . La valeur du point central est la moyenne du domaine de variance minimale.

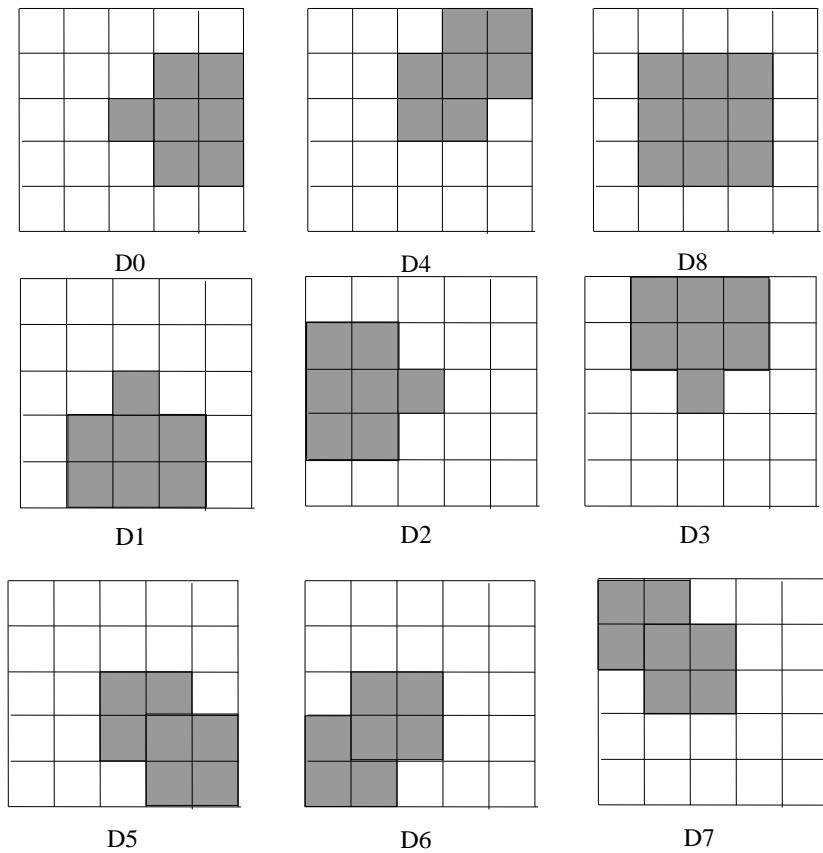


Figure 8-3. Masques du Filtre DE NAGAO

#### *Le filtre médian:*

Ce filtre affecte au point central la valeur de la médiane calculée sur le voisinage. Il ne dégrade pas les zones de transition. Ce filtre fait partie d'une grande famille des filtres qui s'appellent filtres d'ordre. Le paragraphe suivant est consacré à introduire ce type des filtres.

### 8.3 Les Filtres d'Ordre

## Notions sur la statistique d'ordre:

Soit un ensemble  $X$  fini contenant  $N$  variables réelles  $X = \{x_1, x_2, \dots, x_N\}$ .

On appelle  $i^{\text{ème}}$  statistique de  $X$ , l'élément de rang  $i$  obtenu après avoir trier par ordre croissant les valeurs algébriques des  $x_i$ . Cette statistique d'ordre  $i$  est notée  $x_{(i)}$ . Par conséquent,  $x_{(1)}$  est l'élément de  $X$  ayant la plus petite valeur et  $x_{(N)}$  celui ayant la valeur maximale.

On appelle  $X_{(.)} = \begin{pmatrix} x_{(1)} \\ x_{(2)} \\ \vdots \\ x_{(N)} \end{pmatrix}$ , vecteur ordonné de  $X$ , le vecteur tel que  $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(N)}$

Il est le résultat d'une opération de tri par ordre croissant notée par:  $X_{(.)} = OS(X)$

## Définition d'un filtre d'ordre:

Un filtre d'ordre monodimensionnel de longueur  $2n+1$  est un filtre non-linéaire, défini par un vecteur réel  $\mu$  de  $2n+1$  composantes, qui fait correspondre à la séquence d'entrée  $\{x_i, i \in Z\}$  la séquence de sortie  $\{y_k, k \in Z\}$  telle que:  $y_k = \mu \cdot X_{(.)}^k$

avec  $X_{(.)}^k = OS(\{x_{k-n}, x_{k-(n+1)}, \dots, x_k, \dots, x_{k+n}\})$  et  $\mu = (\mu_1 \ \mu_2 \ \dots \ \mu_{2n+1})$

D'où  $y_k = \mu_1 \cdot x_{(1)}^k + \mu_2 \cdot x_{(2)}^k + \dots + \mu_{2n+1} \cdot x_{(2n+1)}^k$

$y_k$  est la valeur filtrée obtenue à l'instant (ou à l'indice)  $k$ ; la fenêtre unidimensionnelle de largeur  $2n+1$  étant centrée autour du point  $k$ . La séquence  $y$  est obtenue en glissant cette fenêtre sur tous les points de la séquence  $x$ .

Ce filtrage peut être généralisé au cas bidimensionnel où les éléments à trier seront les points à l'intérieur de la fenêtre ou du masque utilisé. La valeur de sortie est alors affectée au point où se trouve le centre du masque.

Le filtrage d'ordre peut être présenté comme la cascade d'une opération non linéaire (le tri) et d'une opération linéaire (la combinaison linéaire), comme le montre la Figure 8-4.

Suivant les valeurs des coefficients  $\mu_i$ , le filtre d'ordre correspondant possède des propriétés très différentes.

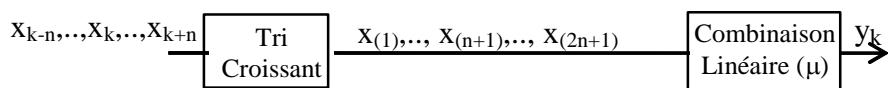


Figure 8-4. Schéma-bloc d'un filtre d'ordre

## Filtres d'ordre usuels:

### Filtre médian:

Pour ce filtre, les coefficients  $\mu_i$  sont tous nuls sauf celui du centre qui est égal à 1:

$$\begin{aligned} \mu_{n+1} &= 1 \\ \mu_i &= 0 \text{ pour } i \neq n+1 \end{aligned}$$

Ce qui correspond à faire passer à la sortie la valeur médiane de la séquence présentée à l'entrée. Le médian d'une séquence de  $N$  valeurs est la valeur pour laquelle il existe  $(N-1)/2$  valeurs qui lui sont inférieures ou égales et  $(N-1)/2$  valeurs qui lui sont supérieures ou égales.

#### Filtre moyenneur:

Le filtre moyenneur est un cas particulier des filtres d'ordre puisqu'il correspond au cas où tous les coefficients  $\mu_i$  sont égaux à  $1/(2n+1)$ . Il n'est pas nécessaire pour ce filtre d'effectuer le tri, et il est de ce fait linéaire.

## 8.4 Détection des contours:

Dans une image, les variations d'intensité représentent des changements de propriétés physiques ou géométriques de la scène ou de l'objet observé, correspondant par exemple à:

- des variations d'illumination, des ombres,
- des changements d'orientation ou de distance de l'observateur,
- des changements du coefficient de réflexion de surface,
- des variations d'absorption des rayons (lumineux, X, ...etc.)

Dans un grand nombre de cas, ces variations d'intensité sont des informations importantes pour les opérations d'analyse d'images. Elles constituent les frontières de régions correspondant à des bords ou parties d'objets de la scène. La localisation de ces frontières ou "détection de contours" constitue une approche de la segmentation d'image. Cette approche est dite "approche frontière" en opposition à l'autre approche dite "région" où l'on cherche à identifier les zones homogènes à l'intérieur d'une image.

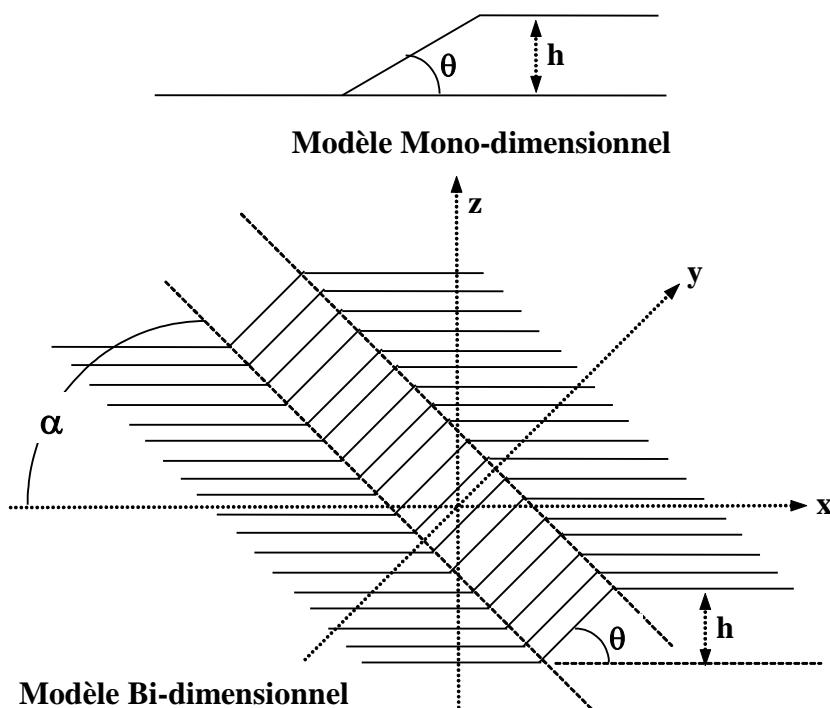


Figure 8-5. modèles théoriques de contour

Intuitivement, dans une image numérique, les contours se situent sur les pixels appartenant à des régions ayant des intensités moyennes différentes; il s'agit de contours de type "saut d'amplitude".

Un contour peut également correspondre à une variation locale d'intensité entre un minimum et un maximum, il s'agit alors de contour "en toit".

Ces deux types ne couvrent pas tous les cas et ne s'appliquent pas en particulier aux frontières séparant des régions de textures différentes.

La Figure 8-5 représente les modèles théoriques de contour dans les cas mon- et bidimensionnels. Ils sont caractérisés par la hauteur  $h$  qui mesure la différence entre les niveaux de gris de part et d'autre de la zone de transition et par l'angle  $\theta$  que fait la rampe de la transition avec l'axe horizontal. Ces deux paramètres ( $h$  et  $\theta$ ) déterminent la largeur de la zone de transition. Un contour existe lorsque  $\theta$  et  $h$  sont supérieurs à des valeurs critiques spécifiées.

Pour le cas bidimensionnel, l'orientation du contour par rapport à l'axe horizontal et sa longueur constituent deux paramètres supplémentaires.

### Approche dérivative

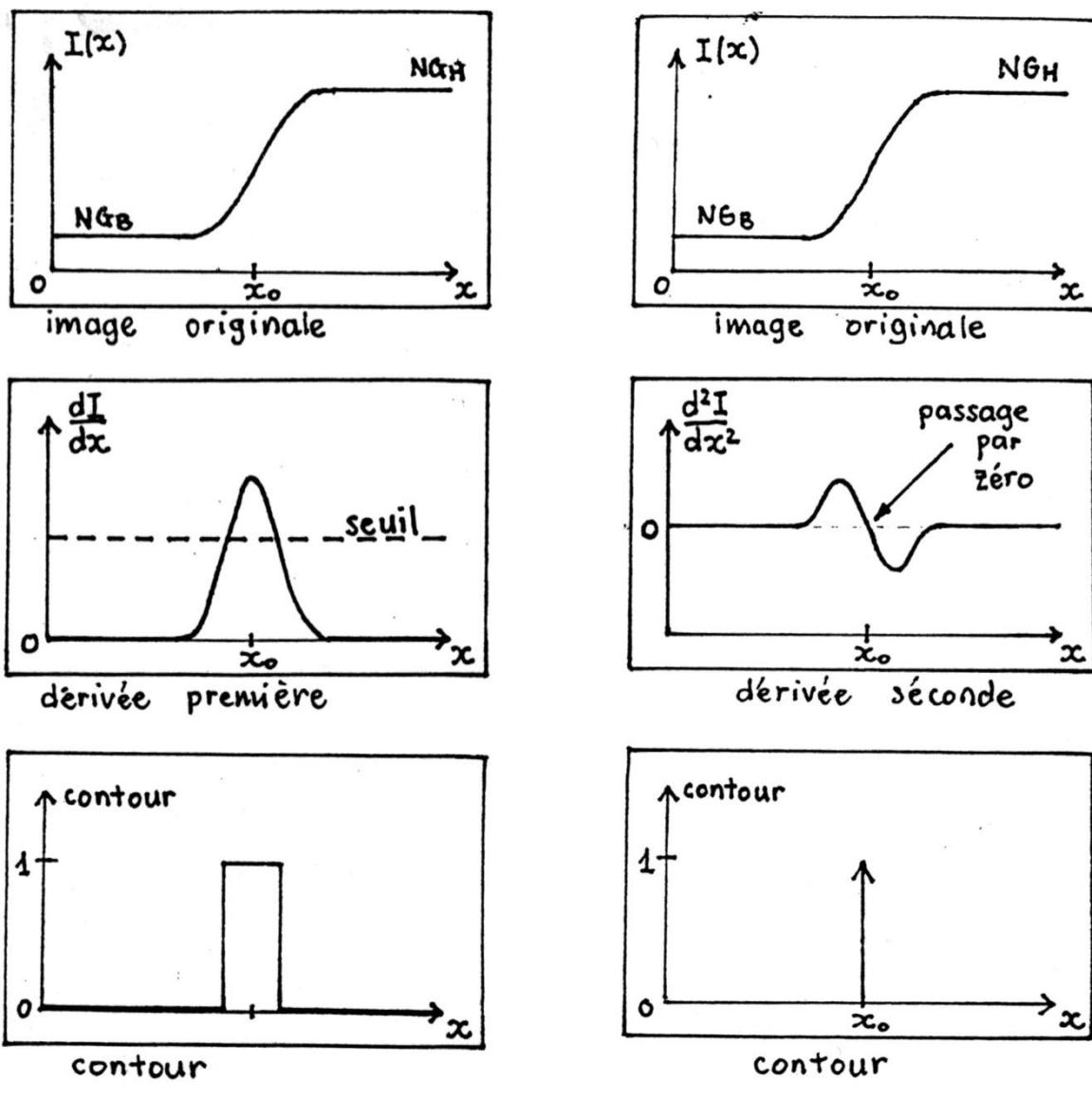


Figure 8-6. Illustration des méthodes dérivatives sur un signal monodimensionnel  $I(x)$

Une manière d'aborder ce problème est de considérer que l'image numérique représente une

fonction scalaire à support borné et dérivable en tout point. Ainsi, les contours sont assimilés aux points de fort gradient et de dérivée seconde nulle. C'est l'approche dérivative. Elle est la plus simple à mettre en œuvre. Il existe deux autres approches dites surfacique et morphologique qui ne seront pas traitées dans le cadre de ce cours.

Le principe général des méthodes dérivatives est illustré sur un signal monodimensionnel  $I(x)$  présentant une transition avec un saut d'amplitude en  $x_0$ . Les allures de la dérivée première (Gradient) et de la dérivée seconde (Laplaciens) sont données sur la Figure 8-6.

Si on considère que la transition du signal est repérée par son point d'inflexion, sa localisation peut se faire par recherche du maximum local de la valeur absolue de la dérivée première ou par recherche du passage à zéro de la dérivée seconde. On peut ainsi définir la zone de transition du signal comme un intervalle comprenant le maximum (ou le minimum) local de la dérivée première ou le passage à zéro de la dérivée seconde.

L'identification d'une zone de transition du signal peut être faite par seuillage de la norme de sa dérivée première. Si le seuil est trop bas, on détecte même des transitions dues au bruit. Le seuillage est utilisé aussi pour la détection du passage à zéro pour la dérivée seconde.

Dans le cas d'images, les deux méthodes dérivatives consistent à approximer le gradient ou le laplacien en chaque point de l'image par des combinaisons linéaires des pixels voisins en utilisant ce qu'on appelle des opérateurs ou masques.

#### **Opérateurs linéaires (de convolution):**

Plusieurs masques de convolution ont été proposés. Parmi lesquels, on cite:

#### **1) Opérateurs dérivateurs du premier ordre (Gradient):**

##### **■ Opérateurs de Prewitt et de Sobel:**

Pour ces opérateurs, les dérivées directionnelles horizontale et verticale s'expriment sous la forme de deux produits de convolution avec deux masques directionnels  $H_L$  et  $H_C$  donnés par:

$$H_L = \begin{pmatrix} 1 & 0 & -1 \\ c & 0 & -c \\ 1 & 0 & -1 \end{pmatrix} \text{ et } H_C = \begin{pmatrix} 1 & c & 1 \\ 0 & 0 & 0 \\ -1 & -c & -1 \end{pmatrix}$$

qui sont les noyaux de convolution de filtres à réponse impulsionnelle finie. Les masques de Prewitt sont définis pour  $c=1$  et ceux de Sobel pour  $c=2$ .

Pour une image  $A$ , les résultats des deux produits de convolution sont deux images  $A_L$  et  $A_C$  telles que:  $A_L[i, j] = H_L * A[i, j]$  et  $A_C[i, j] = H_C * A[i, j]$

Il est intéressant de remarquer que les 2 masques de Sobel et de Prewitt correspondent à la composition de deux convolutions.

Par exemple,  $H_L$  représente la réponse impulsionnelle d'un filtre séparable comprenant un

lissage suivant la direction verticale à l'aide du filtre mono-dimensionnel  $H_{Lv} = \begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$  et une dérivation

suivant la direction horizontale à l'aide du filtre mono-dimensionnel  $H_{Lh} = (1 \ 0 \ -1)$ . D'où, l'on peut

écrire:  $H_L = H_{Lv} \cdot H_{Lh}$  et,  $A_L[i, j] = H_L * A[i, j] = \sum_{k=-m}^m H_{Lv}[k] \cdot \sum_{l=-n}^n H_{Lh}[l] \cdot A[i-k, j-l]$

où  $(2m+1) \times (2n+1)$  sont les dimensions du noyau de  $H_L$ . Il en est de même pour  $H_C$ .

À partir de  $A_L$  et  $A_C$ , on peut calculer la norme et l'orientation du gradient qui, après seuillage, permet de déterminer les points et l'orientation du contour:

$$|\nabla A[i, j]| = \sqrt{A_L^2[i, j] + A_C^2[i, j]} \quad \text{et} \quad \theta[i, j] = \frac{\pi}{2} - \arctg\left(\frac{A_C[i, j]}{A_L[i, j]}\right)$$

Une extraction de contours faite par l'opérateur de Sobel sur les images BUREAU et Coca sont présentées sur la Figure 8-7.

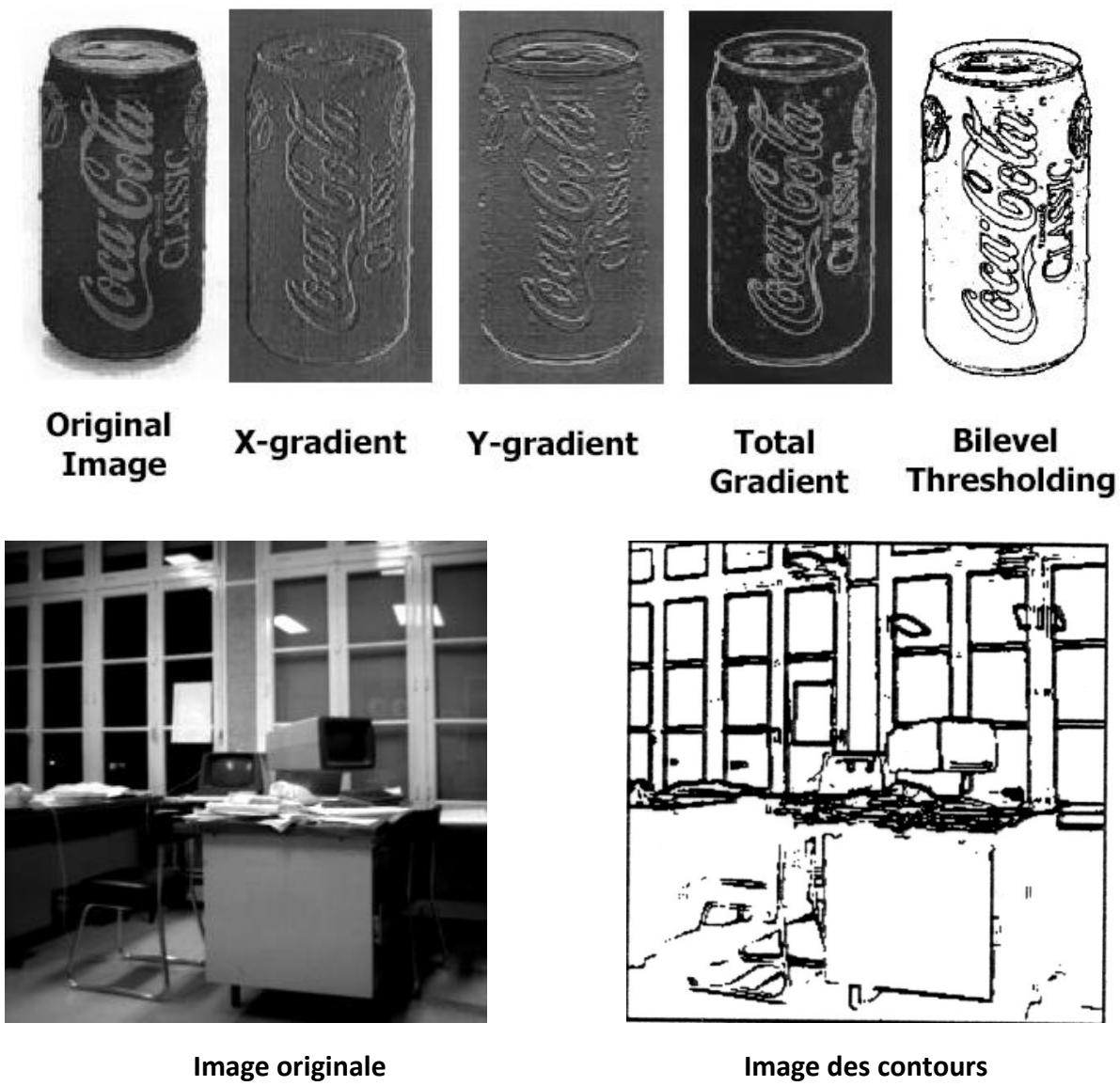


Figure 8-7. Exemples de détection de contours

On observe que les contours sont épais. Le seuil est réglé empiriquement. Cet opérateur est facile à implanter. Il a été utilisé dans les premières réalisations temps réel de détection de contours dans les années 1980. Les processeurs actuels de traitement du signal autorisent d'opérateurs plus performants tels que ceux de Kirsh.

### ■ Opérateurs de Kirsh:

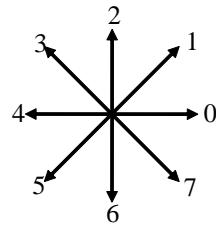
C'est un opérateur à huit masques correspondant chacun à une direction préférentielle et obtenu par rotation de  $\pi/4$  de l'opérateur de base

$H_0 = \begin{pmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{pmatrix}$ . La numérotation des masques est faite selon l'ordre

illustré ci-contre (directions de Freeman en 8-connexité).

$$\text{D'où, } H_1 = \begin{pmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{pmatrix} \text{ et } H_2 = \begin{pmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{pmatrix} \text{ et ainsi de suite... .}$$

Le gradient retenu est celui ayant la plus grande norme:  $|\nabla A|_{\max} = \max_i \{ |H_i * A| ; i = 0, \dots, 7 \}$  dont l'orientation est donné par:  $\frac{\pi}{4} i_{\max}$  où  $i_{\max}$  est l'indice du plus grand gradient.



### ■ Opérateur de Roberts:

C'est l'opérateur le plus simple qui consiste à calculer le gradient dans les deux directions diagonales à l'aide des deux masques  $H_1$  et  $H_2$  de taille 2x2:

$$H_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad H_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

ce qui correspond à :  $Gx[i,j] = A[i,j] - A[i-1,j-1]$  et  $Gy[i,j] = A[i-1,j] - A[i,j-1]$ .

### 2) Opérateurs dérivatifs du second ordre (Laplacien):

La valeur maximale de la dérivée première correspond à un passage par zéro de la dérivée seconde. La localisation de ces passages par zéro donne les points - contours. Le laplacien est approximé, pour une image numérique, par une combinaison linéaire de pixels voisins. L'approximation discrète la plus simple du laplacien, calculée sur un voisinage 3x3, correspond au

$$\text{masque } \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\text{Robinson a proposé un masque légèrement différent: } \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

Cette méthode donne des contours très fins (de l'ordre d'un pixel), mais elle est très sensible au bruit, ce qui apparaît dans la présence des points faux sur le contour détecté. L'exploitation de l'image des points candidats est de ce fait très délicate. De plus, elle n'apporte aucune information sur l'orientation des contours.

Il existe d'autres méthodes de détection qui essayent d'améliorer le comportement des méthodes précédentes, telles que celle de Marr et Hildeth qui ont proposé, pour réduire l'effet d'amplification du bruit à hautes-fréquences du Laplacien, de filtrer l'image par un filtre linéaire dont la réponse impulsionnelle est une gaussienne, et d'appliquer ensuite le Laplacien sur l'image filtrée.

D'autre part, Deriche a proposé, pour améliorer les contours obtenus par les méthodes de gradient, de prolonger les chaînes détectées en suivant les crêtes jusqu'à ce que le module du

gradient soit inférieur à un certain seuil bas. Cette technique de suivi de contours permet d'obtenir des contours plus complets sans ajout de bruit.

#### **Opérateurs non-linéaires (Filtres d'ordre):**

Les filtres d'ordre peuvent être utilisés pour détecter les transitions contenues dans un signal. L'idée d'un tel détecteur est d'estimer, à partir de la différence entre deux statistiques d'ordre symétriques par rapport au centre de la séquence d'entrée, le gradient du signal et en savoir s'il s'agit d'une transition ou non, en comparant cette différence à une valeur-seuil donnée.

Plus précisément, on considère un filtre d'ordre de taille  $N = 2n+1$  et de paramètre  $k$  ( $1 \leq k \leq n$ ). A la séquence d'entrée  $X_i$  ( $i=1, \dots, N$ ) correspond la sortie  $Y$  telle que:  $Y = X_{(n+1+k)} - X_{(n+1-k)}$

C'est la différence entre la  $(n+1+k)$ ème statistique et la  $(n+1-k)$ ème statistique de la séquence d'entrée. Cette différence peut fournir une indication sur la valeur du gradient du signal, qui après seuillage, nous permet de décider s'il s'agit d'une transition au centre de la fenêtre ou pas.

Le choix du paramètre  $k$  est essentiel dans l'opération de la détection puisqu'il détermine les statistiques impliquées dans la différence. Si  $k = n$  alors la différence est effectuée entre le maximum et le minimum des échantillons de l'entrée. Si  $k = 1$ , alors la différence est effectuée entre les statistiques directement supérieure  $X_{(n+2)}$  et directement inférieure  $X_{(n)}$  à la statistique médiane qui est  $X_{(n+1)}$ .

Le principal avantage de ce type de détecteurs est leur pouvoir de lissage, meilleur par rapport aux autres, et leur préservation des transitions ainsi que leur faible sensibilité au bruit.

Rappelons, enfin, que l'efficacité de n'importe quel détecteur dépend du type de bruit présent (pulsionnel, gaussien, blanc, ...etc.) sur l'image traitée et qui modifie le comportement de certains détecteurs.