

# Table des matières

<b>I Codage de canal, analyse et transmission du signal</b>	<b>5</b>
<b>1 Analyse et traitement du signal 1D</b>	<b>7</b>
1.1 Définition et représentation d'un signal . . . . .	7
1.2 Représentation et classification des signaux . . . . .	8
1.2.1 Signaux analogiques ou numériques . . . . .	8
1.2.2 Signaux périodiques et apériodiques . . . . .	9
1.2.3 Signaux déterministes ou stochastiques . . . . .	9
1.2.4 Signaux d'énergie ou de puissance . . . . .	9
1.2.5 Exemples de signaux analogiques . . . . .	10
1.3 Analyse spectrale : Représentation fréquentielle . . . . .	14
1.4 Système . . . . .	14
1.4.1 Système linéaire . . . . .	15
1.4.2 Système permanent . . . . .	16
1.4.3 Réponse impulsionnelle . . . . .	16
1.4.4 Fonction de transfert . . . . .	17
1.4.5 Réponse fréquentielle . . . . .	17
1.4.6 Exemple . . . . .	18
1.5 Filtre linéaire . . . . .	21
1.5.1 Filtre idéal . . . . .	22
1.5.2 Filtre non-idéal . . . . .	25
1.6 Numérisation . . . . .	26
1.6.1 Échantillonnage . . . . .	27
1.6.2 Quantification . . . . .	34
1.6.3 Codage . . . . .	37
1.6.4 Synthétisons... . . . . .	38
1.7 Transformée de FOURIER discrète . . . . .	39
1.7.1 Discréttisation de la transformée de FOURIER . . . . .	39
1.7.2 DFT - IDFT . . . . .	41
<b>2 Transmission du signal</b>	<b>45</b>
2.1 Introduction . . . . .	45
2.1.1 La modulation . . . . .	45
2.1.2 Schéma complet d'une transmission analogique . . . . .	48
2.2 Généralités . . . . .	48
2.2.1 La porteuse . . . . .	49
2.2.2 Le signal modulé . . . . .	49
2.3 Modulation d'amplitude . . . . .	49
2.3.1 Modulation AM . . . . .	50
2.3.2 Modulation DSB-SC . . . . .	57
2.3.3 Modulation BLU . . . . .	59

2.3.4	Modulation en quadrature . . . . .	64
2.4	Modulation angulaire . . . . .	67
2.4.1	Modulation de phase (PM) . . . . .	69
2.4.2	Modulation de fréquence (FM) . . . . .	69
2.4.3	Modulation FM à bande étroite (NBFM) . . . . .	71
2.4.4	Modulation FM à large bande . . . . .	76
2.5	Multiplexage en fréquence (FDM) . . . . .	81
<b>II</b>	<b>Codage de source</b>	<b>85</b>
<b>3</b>	<b>Introduction à la théorie de l'information, du codage et de la compression</b>	<b>87</b>
3.1	Introduction . . . . .	87
3.1.1	Modèle mathématique d'une source . . . . .	87
3.1.2	Source discrète sans mémoire . . . . .	88
3.2	Mesure de l'information . . . . .	88
3.2.1	Quantité d'information . . . . .	89
3.2.2	Entropie d'une source . . . . .	90
3.2.3	Entropie jointe entre deux sources . . . . .	91
3.2.4	Quantité d'information mutuelle . . . . .	93
3.2.5	Entropie conditionnelle . . . . .	94
3.2.6	Lien entre entropie conditionnelle et information mutuelle . . . . .	95
3.2.7	Exemple : Canal de communication binaire symétrique . . . . .	96
3.3	Codage de sources discrètes . . . . .	98
3.3.1	Codage avec mots de longueur fixe . . . . .	99
3.3.2	Codage par blocs : Extension de la source . . . . .	100
3.3.3	Codage avec mots de longueur variable . . . . .	101
3.3.3.1	Codages avec et sans préfixe . . . . .	102
3.3.3.2	Arbre d'un codage . . . . .	102
3.3.3.3	Longueur moyenne des mots de code . . . . .	103
3.3.3.4	Inégalité de KRAFT . . . . .	104
3.3.3.5	Premier théorème de SHANNON : Théorème de codage de source	105
3.4	Introduction à la compression de données . . . . .	107
3.4.1	Codage de HUFFMAN . . . . .	108
3.4.2	Codage arithmétique . . . . .	111
3.4.3	Codage par dictionnaire : méthode de LEMPEL-ZIV . . . . .	114
3.4.4	Codage par répétition : méthode RLE (RLC) . . . . .	118
3.5	Exercices . . . . .	120
3.6	Annexes . . . . .	121
3.6.1	Maximisation de l'entropie d'une source de $K$ symboles . . . . .	121
3.6.2	Expansion binaire d'un nombre réel . . . . .	122
3.6.3	Table ASCII . . . . .	122
3.7	Références . . . . .	122
<b>III</b>	<b>Traitement d'images</b>	<b>125</b>
<b>4</b>	<b>Analyse et traitement d'images</b>	<b>127</b>
4.1	Introduction . . . . .	127
4.1.1	Type d'images . . . . .	127
4.1.2	Image numérique . . . . .	129

4.1.3	Traitement d'images . . . . .	131
4.2	Traitements linéaires . . . . .	132
4.2.1	Transformée de FOURIER discrète et convolution discrète . . . . .	132
4.2.2	Traitement global . . . . .	134
4.2.3	Traitement local : masques de convolution . . . . .	139
4.3	Traitements non-linéaires . . . . .	141
4.3.1	Images binaires . . . . .	141
4.3.1.1	Rappels sur la théorie des ensembles . . . . .	141
4.3.1.2	Transformations morphologiques élémentaires . . . . .	143
4.3.1.3	Transformations morphologiques complexes . . . . .	147
4.3.2	Images en niveaux de gris . . . . .	148
4.3.2.1	De la notion d'ensemble à celle de fonction . . . . .	149
4.3.2.2	Transformations morphologiques élémentaires . . . . .	149
4.3.2.3	Filtrage non-linéaire . . . . .	150
4.4	Traitement spécifique : Rehaussement et restauration . . . . .	155
4.4.1	Définitions . . . . .	155
4.4.2	Amélioration du contraste par transformation de l'histogramme . . . . .	157
4.4.2.1	Transformation linéaire . . . . .	157
4.4.2.2	Transformation linéaire avec saturation . . . . .	157
4.4.2.3	Transformation linéaire par morceaux . . . . .	159
4.4.2.4	Transformation non-linéaire . . . . .	159
4.4.2.5	Autre fonction possible... Le négatif d'une photo . . . . .	159
4.4.3	Égalisation de l'histogramme . . . . .	161
4.5	Traitement spécifique : Détection de contours . . . . .	165
4.5.1	Opérateurs linéaires basés sur le calcul des dérivées . . . . .	165
4.5.1.1	Opérateur de dérivée première et Gradient . . . . .	166
4.5.1.2	Opérateur de dérivée seconde et Laplacien . . . . .	170
4.5.1.3	Calcul pratique des dérivées et masques de convolution . . . . .	170
4.5.2	Opérateurs non-linéaires basés sur la morphologie mathématique . . . . .	172
4.6	Traitement spécifique : Segmentation et seuillage . . . . .	177
4.6.1	Définition . . . . .	177
4.6.2	Seuillage simple . . . . .	178
4.6.3	Seuillage multiple . . . . .	178
4.6.4	Seuillage automatique . . . . .	179
4.7	Bibliographie . . . . .	181
<b>IV</b>	<b>Analyse de FOURIER</b>	<b>183</b>
<b>5</b>	<b>Séries de FOURIER</b>	<b>185</b>
5.1	Introduction . . . . .	185
5.1.1	Fonction continue par morceaux . . . . .	187
5.1.2	Fonctions périodiques . . . . .	188
5.2	Polynômes de FOURIER (ou trigonométriques) . . . . .	189
5.2.1	Définition . . . . .	189
5.2.2	Intégrales intéressantes . . . . .	189
5.2.3	Calcul des coefficients $a_0$ , $a_n$ et $b_n$ . . . . .	190
5.3	Théorème de FOURIER . . . . .	191
5.3.1	Exemple : Fonction carrée . . . . .	191
5.3.2	Définition et énoncé du théorème . . . . .	194

5.4	Mise en forme des séries de FOURIER . . . . .	195
5.4.1	Forme complexe ou exponentielle . . . . .	195
5.4.2	Forme trigonométrique . . . . .	197
5.4.3	Cas particuliers : fonctions paires, impaires et demi-onde . . . . .	197
5.5	Formule de PARSEVAL . . . . .	198
5.6	Exemple : Droite de FOURIER . . . . .	199
5.6.1	Phénomène de GIBBS . . . . .	200
5.7	Exercices . . . . .	200
5.8	Références . . . . .	201
<b>6</b>	<b>Transformée de FOURIER 1D</b>	<b>203</b>
6.1	Transformée de FOURIER 1D . . . . .	203
6.1.1	Définitions . . . . .	203
6.1.2	Propriétés . . . . .	204
6.1.3	Exemples . . . . .	206
6.2	La fonction Delta de DIRAC . . . . .	209
6.2.1	Définition . . . . .	209
6.2.2	Transformée de FOURIER . . . . .	209
6.2.3	Applications directes . . . . .	210
6.3	Retour à la transformée de FOURIER . . . . .	213
6.4	Quelques signaux fondamentaux . . . . .	213
6.4.1	Définitions . . . . .	213
6.4.2	Paires de transformées de FOURIER . . . . .	214
<b>7</b>	<b>Transformée de FOURIER 2D</b>	<b>215</b>
7.1	Transformée de FOURIER 2D . . . . .	215
7.1.1	Définition . . . . .	215
7.1.2	Propriétés . . . . .	216
7.1.3	Exemples . . . . .	218
7.2	La fonction Delta de DIRAC . . . . .	220
7.2.1	Définition . . . . .	220
7.2.2	Transformée de FOURIER . . . . .	221
7.2.3	Applications directes . . . . .	221

## Première partie

# Codage de canal, analyse et transmission du signal



# Chapitre 1

## Analyse et traitement du signal 1D

Ce chapitre a pour but de définir, ou de rappeler pour certains, la notion de signal. En particulier, il étudie les signaux à une dimension (1D) et les outils principaux permettant de les étudier et de les manipuler.

### 1.1 Définition et représentation d'un signal

Les signaux tels que nous les percevons dans la nature sont *analogiques*, c'est-à-dire qu'il n'est pas possible de déceler une discontinuité. Par exemple, l'aiguille d'un compteur kilométrique évolue d'une position vers une autre sans discontinuité. Elle fournit une *information* chiffrée qui évolue d'une manière continue au cours du temps.

Un autre exemple est le son qui arrive à nos oreilles en continu. Il s'agit d'une onde qui se propage dans l'air et qui est perceptible grâce au détecteur de pression qu'est le tympan de l'oreille. Il est possible de transformer une onde sonore en une tension électrique à l'aide d'un microphone. Cette tension électrique évoluera donc de la même manière (nous dirons proportionnellement) que l'onde sonore étudiée. Comme l'aiguille du compteur kilométrique, cette tension fournit une information chiffrée qui évolue de manière continue au cours du temps.

Nous en arrivons donc à la définition que nous adopterons pour un signal :

**Définition [Signal].** *En toute généralité, un signal constitue un flot d'informations qui évolue au cours du temps.*

Bien sûr, cette définition est un peu simpliste et nécessitera un raffinement dans la suite. Une première remarque s'impose néanmoins. Les signaux cités en exemple ci-dessus sont dit "à une dimension (1D)". Cela signifie que l'information fournie ne dépend que d'*un seul paramètre* qui est ici le temps. La notion de temps n'est pas toujours justifiée. En effet, on pourrait par exemple étudier la force mécanique s'exerçant sur une poutre d'un bâtiment sur toute sa longueur. L'information est alors ici la valeur de la force qui varie selon la position étudiée sur la poutre. Cette information est également un signal, mais plus un signal temporel. On pourrait dire qu'il s'agit d'un signal de "position". Cette remarque se justifiera plainement pour les signaux bidimensionnels (2D), ou images, dont l'information dépend de deux paramètres qui sont les coordonnées ( $x, y$ ) d'un pixel de l'image. Néanmoins, dans ce chapitre, nous nous contenterons de considérer les signaux temporels.

Qu'en est-il de l'information binaire (les 1 et les 0 transmis sur un réseau par exemple) ou textuelle (les lettres constituant un fichier texte classique par exemple) que l'on rencontre

en permanence en informatique ? Et bien, ce type d'information constitue également un signal d'informations, et dans la plupart des cas temporel. La grosse différence par rapport aux exemples cités plus hauts est que les informations arrivent de manière discrète et non continue. Par exemple, un réseau présentant un débit de 1 kbits/seconde, véhicule des signaux temporels dont chaque nouvelle information (ici le bit) arrive toutes les 0,001 secondes. En toute généralité, on parlera de signaux numériques et non plus analogiques.

Il est à présent grand temps de réaliser une classification des signaux.

## 1.2 Représentation et classification des signaux

On peut distinguer plusieurs classes de signaux :

- *analogiques* ou *numériques* ; tels que ceux que nous avons déjà introduits ci-dessus,
- *périodiques* ou *apériodiques*,
- *déterministes* ou *stochastiques* (aléatoires),
- d'*énergie* ou de *puissance*.

Passons ces différents types de signaux en revue.

### 1.2.1 Signaux analogiques ou numériques

Un signal *analogique* est une fonction  $g(t)$  continue du temps  $t$ . Par contre, un signal *numérique* est un signal temporel discontinu ; on le notera  $g[n]$  où  $n$  est l'indice entier d'un élément pris dans l'ensemble d'instants  $\{t_0, t_1, \dots\}$ . On parle encore de *signaux à temps discrets*.

Néanmoins, il est nécessaire de faire la distinction entre la nature de l'information (signal d'information), analogique ou numérique, et sa représentation sur un média ou un canal de communication. La figure 1.1 aide à clarifier la distinction entre un signal d'information et représentation. Un même signal d'information peut avoir plusieurs représentations.

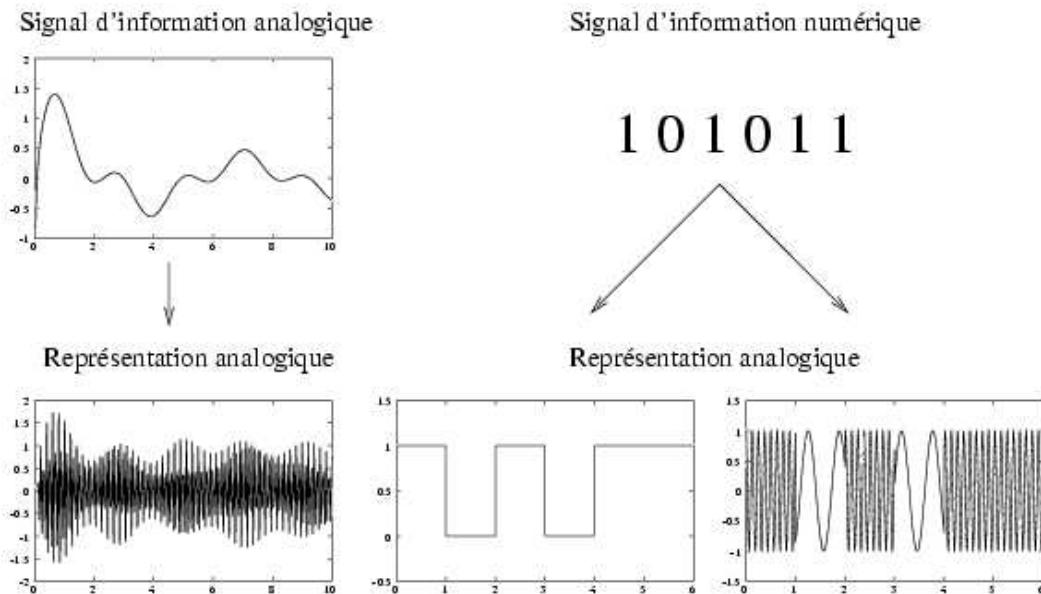


FIGURE 1.1 – Représentation d'un signal analogique ou numérique.

### 1.2.2 Signaux périodiques et apériodiques

Un signal  $g(t)$  est *périodique* s'il satisfait la relation suivante :

$$g(t) = g(t + T_0) \quad \forall t \quad (1.1)$$

où  $t$  est la variable de temps et  $T_0$  est une constante. La plus petite valeur de  $T_0$  pour laquelle cette relation est vérifiée est appelée *période fondamentale* de  $g(t)$ . Cette relation signifie que le signal se répète toutes les  $T_0$  secondes.

S'il n'existe pas de constante pour laquelle la relation (1.1) est respectée, on dit que le signal  $g(t)$  est *apériodique* ou encore non-périodique.

### 1.2.3 Signaux déterministes ou stochastiques

Un signal *déterministe* a une évolution connue et prévisible, contrairement aux signaux *aléatoires* ou *stochastiques*. En d'autres termes, un signal est déterministe s'il est possible de connaître sa valeur à tout instant  $t$ , comme par exemple

$$g(t) = \sin(2\pi t)$$

Il est en effet possible de connaître la valeur du signal à tout instant en remplaçant  $t$  par sa valeur dans l'expression de la fonction  $g$ .

### 1.2.4 Signaux d'énergie ou de puissance

La plupart du temps, les signaux étudiés sont à un moment ou un autre représentés sous la forme d'un signal électrique, par exemple lors de sa transmission sur un câble réseau. Le signal représente alors une tension ou une intensité électrique. Considérons une tension électrique  $v(t)$  qui, à travers une résistance  $R$ , produit un courant  $i(t)$ . La puissance instantanée dissipée dans cette résistance est définie par

$$p(t) = \frac{|v(t)|^2}{R} \quad (1.2)$$

ou encore

$$p(t) = R |i(t)|^2 \quad (1.3)$$

Quelle qu'en soit l'expression, la puissance instantanée est une fonction quadratique du signal. À travers une résistance de 1 [Ohm], les expressions sont égales, que le signal représente une tension ou un courant. Il est dès lors coutume de normaliser l'expression pour une résistance de 1 Ohm. On obtient alors

**Définition [Puissance instantanée].** La puissance instantanée d'un signal  $g(t)$  est définie par

$$p(t) = |g(t)|^2 \quad (1.4)$$

Cette expression est valable pour n'importe quel signal analogique. Mais qu'en est-il de l'énergie, notion intimement liée à celle de puissance ? Tout d'abord, un petit rappel... L'énergie est le produit de la puissance et de l'intervalle de temps considéré :  $E = P.\Delta t$ . Dans notre cas, la puissance est une fonction du temps et par convention, un signal évolue pour  $t$  qui va de  $-\infty$  à  $+\infty$ . Dès lors,

**Définition [Énergie].** L'énergie totale d'un signal  $g(t)$  est définie par

$$E = \int_{-\infty}^{+\infty} |g(t)|^2 dt \quad (1.5)$$

Cette intégrale peut exister (c'est-à-dire donner une valeur finie) ou non (c'est-à-dire donner une valeur infinie) selon le signal  $g$  considéré. Donc, certains signaux ont une énergie finie ; on les appelle *signaux d'énergie*.

Pour les signaux qui possède une énergie infinie, on utilise alors la notion de puissance moyenne qui est la moyenne temporelle de l'énergie.

**Définition [Puissance moyenne].** La puissance moyenne d'un signal  $f(t)$  est définie par

$$P = \lim_{T \rightarrow +\infty} \frac{\int_{-T}^{+T} |g(t)|^2 dt}{2T} \quad (1.6)$$

Dans le cas d'un signal périodique de période  $T_0$ , l'expression de la puissance moyenne devient

$$P = \frac{1}{T_0} \int_0^{T_0} |g(t)|^2 dt \quad (1.7)$$

Synthétisons à présent. Les définitions de l'énergie et de la puissance moyenne amènent à distinguer deux types de signaux :

- les signaux à énergie finie, pour lesquels  $0 < E < +\infty$ . Un signal physiquement réalisable est à énergie finie.
- les signaux à puissance finie. Dans ce cas, la puissance est bornée, à savoir  $0 < P < +\infty$ .

Ces deux contraintes sont mutuellement exclusives. En particulier, un signal à énergie finie a une puissance moyenne nulle alors qu'un signal à puissance finie possède une énergie infinie.

### 1.2.5 Exemples de signaux analogiques

#### Exemple 1

Considérons le signal suivant

$$g(t) = A \cos(2\pi f_0 t)$$

où  $A$  et  $f_0$  sont des constantes représentant respectivement l'amplitude et la fréquence du signal ; ici cosinusoidé. La figure 1.2(a) illustre un tel signal pour  $A = 2$  et  $f_0 = 2 [Hz]$ . Tout d'abord, il s'agit bien d'un signal analogique vu que la fonction sinus est bien une fonction continue du temps. Ensuite, nous n'avons aucune contrainte sur sa durée, on peut donc conclure qu'il va de  $-\infty$  à  $+\infty$ . Il est donc possible que  $g(t)$  soit périodique. En effet si nous posons

$$T_0 = \frac{1}{f_0}$$

nous pouvons écrire

$$g(t) = A \cos(2\pi f_0 t) = A \cos(2\pi f_0(t + T_0)) = g(t + T_0)$$

et cela quelque soit la valeur de  $t$ . Le signal  $g(t)$  est donc périodique de période  $T_0$ . Bien sûr,  $2T_0, 3T_0, \dots$  sont également des périodes du signal mais  $T_0$  constitue la période fondamentale.

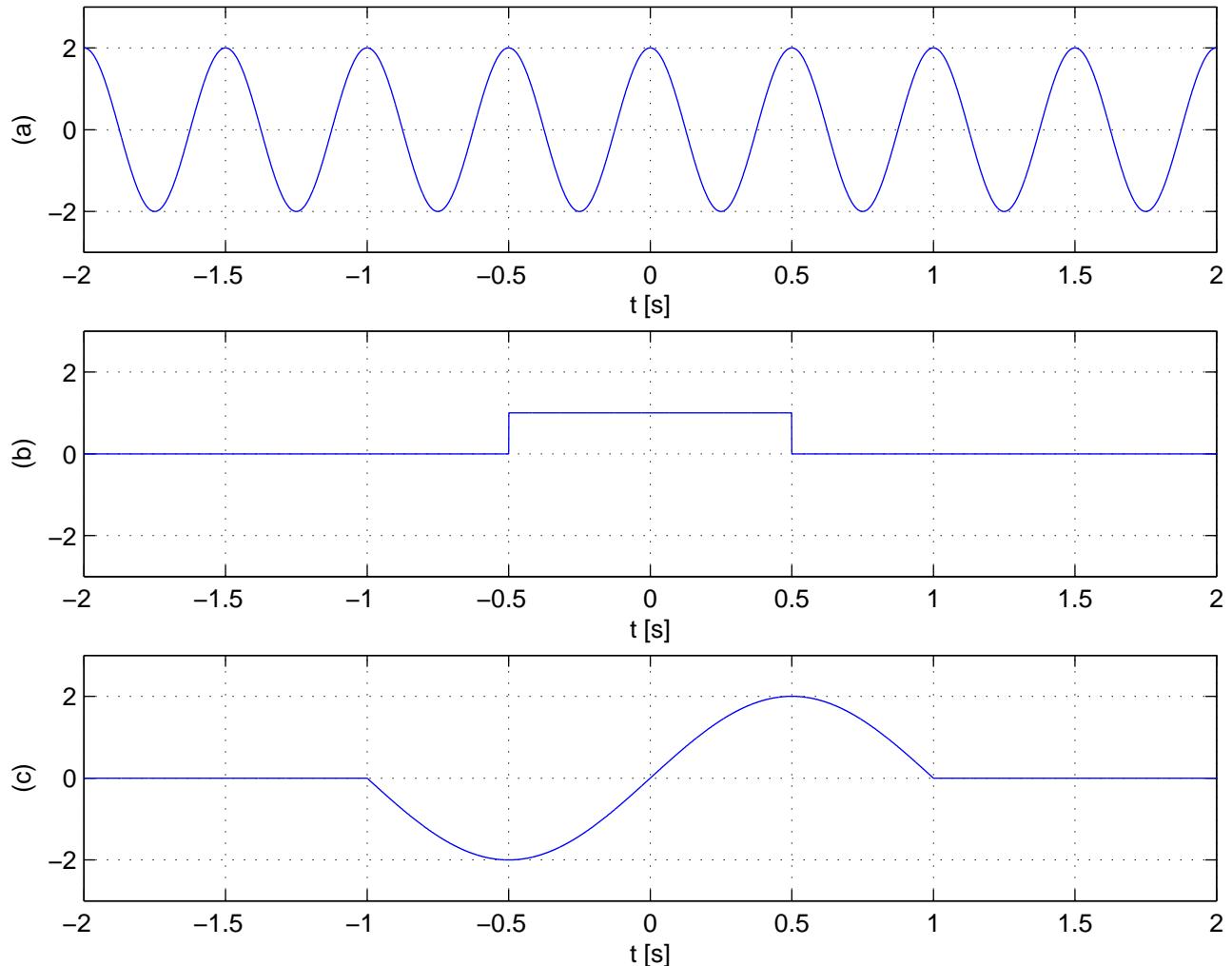


FIGURE 1.2 – Exemples de signaux analogiques.

Calculons à présent la puissance instantanée, l'énergie et la puissance de ce signal. La puissance instantanée est donnée par

$$p(t) = |g(t)|^2 = A^2 \cos^2(2\pi f_0 t)$$

tandis que l'énergie est fournie par

$$E = \int_{-\infty}^{+\infty} p(t) dt = \int_{-\infty}^{+\infty} A^2 \cos^2(2\pi f_0 t) dt = +\infty$$

Cette intégrale diverge indiquant que l'énergie du signal est infinie.

Calculons à présent sa puissance :

$$\begin{aligned}
 P &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} p(t) dt \\
 &= \lim_{T \rightarrow +\infty} \frac{A^2}{2T} \int_{-T}^{+T} \cos^2(2\pi f_0 t) dt \\
 &= \lim_{T \rightarrow +\infty} \frac{A^2}{4T} \int_{-T}^{+T} (1 + \cos(4\pi f_0 t)) dt \\
 &= \lim_{T \rightarrow +\infty} \frac{A^2}{4T} \left[ t + \frac{\sin(4\pi f_0 t)}{4\pi f_0} \right]_{-T}^T \\
 &= \lim_{T \rightarrow +\infty} \frac{A^2}{4T} \left( 2T + 2 \frac{\sin(4\pi f_0 T)}{4\pi f_0} \right) \\
 &= \lim_{T \rightarrow +\infty} \left( \frac{A^2}{2} + \frac{A^2}{2} \frac{\sin(4\pi f_0 T)}{4\pi f_0 T} \right) \\
 &= \frac{A^2}{2}
 \end{aligned}$$

vu que la fonction *sinc* tend vers 0 à l'infini. La puissance du signal  $g(t)$  est donc finie et il s'agit donc d'un signal de puissance. Bien sûr, un tel signal n'est pas réalisable physiquement car il aurait fallu qu'il démarre en  $-\infty$  et disposer d'une énergie infinie, ce qui n'est faisable pour nous, pauvres mortels... Néanmoins, ce genre de signal est couramment utilisé pour la modélisation et l'étude des systèmes de télécommunication. Nous y reviendrons.

## Exemple 2

Considérons le signal suivant

$$g(t) = \begin{cases} 1 & \text{si } -\frac{1}{2} < t < +\frac{1}{2} \\ 0 & \text{sinon} \end{cases}$$

Il s'agit d'une impulsion rectangulaire d'amplitude 1 et durée égale à 1 seconde ; celle-ci est représentée à la figure 1.2(b). Nous la noterons également

$$g(t) = \text{rect}(t)$$

Le signal  $g(t)$  est bien analogique étant donné que la fonction  $g(t)$  est définie pour tout  $t$ . Il n'est pas périodique vu qu'il ne se répète pas dans le temps. sa puissance instantanée est fournie par

$$p(t) = \text{rect}^2(t)$$

tandis que son énergie est donnée par

$$\begin{aligned}
 E &= \int_{-\infty}^{+\infty} \text{rect}^2(t) dt \\
 &= \int_{-\frac{1}{2}}^{+\frac{1}{2}} 1 dt \\
 &= 1
 \end{aligned}$$

Son énergie est donc finie ; il s'agit donc d'un signal d'énergie. Vérifions tout de même que sa puissance est nulle :

$$\begin{aligned} P &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} \text{rect}^2(t) dt \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \\ &= 0 \end{aligned}$$

ouf...

### Exemple 3

Considérons le signal suivant

$$g(t) = \begin{cases} A \sin(2\pi f_0 t) & \text{si } -\frac{T_0}{2} < t < +\frac{T_0}{2} \\ 0 & \text{sinon} \end{cases}$$

où  $f_0 = 1/T_0$ . Il est représenté à la figure 1.2(c) pour  $A = 2$  et  $T_0 = 2 [s]$ . On pourrait croire qu'il s'agit d'une simple sinusoïde, dans le genre de l'exemple 1. Néanmoins, il n'en est rien. Il s'agit d'une impulsion sinusoïdale ; impulsion car ce signal ne dure qu'un laps de temps fini et enfin sinusoïdale car il a la forme d'une sinusoïde.

Il s'agit toujours d'un signal analogique car la fonction  $g(t)$  est connue en tout  $t$ . Ce signal n'est pas périodique car il ne se répète pas, contrairement à une sinusoïde qui s'étend de  $-\infty$  à  $+\infty$ . Sa puissance instantanée est donnée par

$$p(t) = A^2 \sin^2(2\pi f_0 t) \text{rect}\left(\frac{t}{T_0}\right)$$

tandis que son énergie est donnée par

$$\begin{aligned} E &= \int_{-\infty}^{+\infty} |g(t)|^2 dt \\ &= A^2 \int_{-\frac{T_0}{2}}^{+\frac{T_0}{2}} \sin^2(2\pi f_0 t) dt \\ &= \frac{A^2 T_0}{2} \end{aligned}$$

qui est donc finie ! Il s'agit donc d'un signal d'énergie et non de puissance comme une simple sinusoïde ! Nous laisserons au lecteur motivé le soin de montrer que la puissance de  $g(t)$  est bien égale à 0.

### 1.3 Analyse spectrale : Représentation fréquentielle

Les signaux *déterministes* ont une évolution temporelle connue de leur valeur, et comme on l'a vu, on peut leur donner une représentation temporelle à l'aide d'une fonction  $g(t)$ . Il existe un outil permettant de représenter ces signaux dans le domaine fréquentiel. Nous allons donc introduire ici une autre représentation du signal  $g(t)$ , donc il ne s'agit pas d'une modification du signal mais bien d'un autre moyen de l'interpréter et de l'analyser. Cet outil porte le nom de transformée de FOURIER.

Soit un signal déterministe  $g(t)$ , la transformée de FOURIER de  $g(t)$  est définie par

$$G(f) = \int_{-\infty}^{+\infty} g(t) e^{-j2\pi t f} dt \quad (1.8)$$

où  $f$  est la variable exprimant la fréquence exprimée en Hertz, noté [Hz]. La fonction  $G(f)$  constitue la représentation fréquentielle du signal temporel  $g(t)$ . Il est possible de retrouver le signal temporel de départ en utilisant la transformation de FOURIER inverse :

$$g(t) = \int_{-\infty}^{+\infty} G(f) e^{j2\pi t f} df \quad (1.9)$$

Il est loin d'être évident, d'un premier abord en observant sa définition, de donner à la transformée de FOURIER une interprétation physique claire. Néanmoins, elle en a une très intéressante et très utile comme nous le verrons dans la suite, celle de donner les composantes fréquentielles du signal. Afin que le lecteur puisse mieux intégrer cette notion, nous le renvoyons à la section consacrée exclusivement à la transformée de FOURIER.

### 1.4 Système

On peut considérer un système comme une “boîte noire” prenant en entrée un signal  $x(t)$  et fournissant en sortie un signal  $y(t)$  obtenu en faisant subir au signal  $x(t)$  une série d'opérations de traitement. Nous dirons que le système a réalisé un *traitement* sur le signal  $x(t)$  pour fournir en sortie le signal  $y(t)$ . Ceci est illustré à la figure 1.3. Un système, tel que nous le considérerons dans ce cours, sera donc une entité matérielle ou logicielle faisant subir à un signal un certain nombre de modifications.

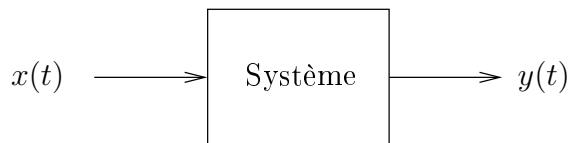


FIGURE 1.3 – Illustration d'un système.

Par entité matérielle, on entend par exemple un circuit électronique, un canal de communication comme un câble réseau ou un canal hertzien qui sont des systèmes physiques que l'on peut modéliser mathématiquement afin d'en prédire leur comportement. Par entité logicielle, on entend par exemple un programme informatique dont le but est de traiter un signal dans un but précis. Il est donc possible de créer ses propres systèmes de traitement du signal tandis que certains systèmes existent déjà bel et bien physiquement et, dans ce cas, on ne peut que

les modéliser.

Comme pour les signaux, il est possible de donner une classification des systèmes. Deux catégories importantes sont les systèmes *linéaires* et les systèmes *non-linéaires*. Pour la première catégorie, nous allons voir que l'analyse de FOURIER est un outil très important.

On notera

$$y(t) = O(x(t))$$

le fait que le système dont l'opérateur est  $O$  transforme en le signal  $x(t)$  en le signal  $y(t)$ . On appellera *opérateur* l'opération mathématique qui exprime la relation existant entre la sortie  $y(t)$  et l'entrée  $x(t)$  du système.

### Exemple

Considérons le système défini par la relation suivante

$$y(t) = 3x(t)$$

Ce système multiplie simplement le signal d'entrée par un facteur 3. On peut donc le considérer comme un simple amplificateur.

#### 1.4.1 Système linéaire

Un système  $O$  est linéaire à la condition suivante. Si

$$y_1(t) = O(x_1(t))$$

et

$$y_2(t) = O(x_2(t))$$

alors

$$ay_1(t) + by_2(t) = O(ax_1(t) + bx_2(t)) \quad (1.10)$$

### Exemple

Reprenons le système de l'exemple précédent. Dès lors, nous avons

$$y_1(t) = O(x_1(t)) = 3x_1(t)$$

$$y_2(t) = O(x_2(t)) = 3x_2(t)$$

et donc

$$\begin{aligned} ay_1(t) + by_2(t) &= a3x_1(t) + b3x_2(t) \\ &= 3(ax_1(t) + bx_2(t)) \\ &= O(ax_1(t) + bx_2(t)) \end{aligned}$$

Ce système est donc linéaire.

## Exemple

Considérons à présent le système caractérisé par la relation suivante :

$$y(t) = x^2(t)$$

Il élève donc le signal d'entrée au carré. Calculons le membre de gauche de (1.10) :

$$ay_1(t) + by_2(t) = ax_1^2(t) + bx_2^2(t)$$

Le membre de droite est quant à lui donné par

$$O(ax_1(t) + bx_2(t)) = a^2x_1^2(t) + b^2x_2^2(t) + 2abx_1(t)x_2(t)$$

qui est différent du membre de gauche. Ce système est donc non-linéaire.

## 1.4.2 Système permanent

Un système est *permanent* ou *invariant dans le temps* si le fait de décaler l'entrée revient à décaler le signal de sortie d'une valeur de temps. En d'autres mots, un système est permanent si ses caractéristiques ne changent pas au cours du temps. En termes mathématiques, cela s'exprime ainsi : si

$$y(t) = O(x(t))$$

alors

$$y(t - \tau) = O(x(t - \tau))$$

où  $\tau$  est un intervalle de temps constant.

Un canal téléphonique est raisonnablement permanent pour la durée de la conversion. Par contre, l'invariance est une hypothèse fausse pour des communications mobiles.

## 1.4.3 Réponse impulsionale

La théorie des systèmes linéaires nous apprend (nous l'admettrons dans le cadre de ce cours) qu'un système linéaire permanent peut être caractérisé entièrement par une fonction que nous noterons  $h(t)$  et appellerons *réponse impulsionale*. Cette fonction intervient dans la relation suivante, liant l'entrée  $x(t)$  à la sortie  $y(t)$  :

$$y(t) = \int_{-\infty}^{+\infty} h(t - \tau) x(\tau) d\tau \quad (1.11)$$

Nous reconnaissions ici le produit de convolution. Dès lors, le signal  $y(t)$  à la sortie d'un système linéaire permanent correspond au produit de convolution entre le signal  $x(t)$  à l'entrée du système et la réponse impulsionale  $h(t)$  du système, ce que nous noterons également

$$y(t) = (h \otimes x)(t) \quad (1.12)$$

Justifions à présent le terme "réponse impulsionale". Considérons un système de réponse impulsionale  $h(t)$  et appliquons-lui à son entrée une impulsion de DIRAC :

$$x(t) = \delta(t)$$

La sortie est alors donnée par

$$\begin{aligned} y(t) &= \int_{-\infty}^{+\infty} h(t - \tau) \delta(\tau) d\tau \\ &= h(t) \end{aligned}$$

Dès lors, la réponse impulsionale correspond à la sortie, ou la *réponse*, du système lorsqu'on lui applique à son entrée une *impulsion* de DIRAC.

### 1.4.4 Fonction de transfert

Nous venons de voir qu'un système linéaire peut entièrement être défini par sa réponse impulsionnelle  $h(t)$ . Nous allons voir à présent qu'il est également possible de lui donner une interprétation fréquentielle. Il suffit pour cela de se rappeler de la propriété de la transformée de FOURIER portant sur le produit de convolution dans le domaine temporel. En effet, en prenant la transformée de FOURIER de la relation (1.12), il vient

$$Y(f) = H(f) X(f) \quad (1.13)$$

où  $X(f)$  et  $Y(f)$  sont respectivement les transformées de FOURIER de  $x(t)$  et  $y(t)$ . La fonction  $H(f)$  est la transformée de FOURIER de la réponse impulsionnelle  $h(t)$ , et porte le nom de *fonction de transfert* du système. Dès lors, faire passer un signal  $x(t)$  au travers d'un système linéaire revient à multiplier sa transformée de FOURIER par la fonction de transfert du système. Ceci est intéressant dans le sens où une multiplication est beaucoup plus simple qu'un produit de convolution, et cela d'autant plus lorsque l'on est conscient de la richesse des informations présentes dans la représentation fréquentielle d'un signal.

Comme pour les signaux, on a donc une représentation fréquentielle des systèmes linéaires. Ceci est illustré à la figure 1.4.

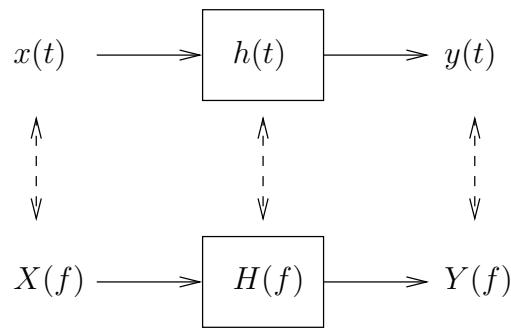


FIGURE 1.4 – Représentations temporelle et fréquentielle des systèmes linéaires.

### 1.4.5 Réponse fréquentielle

Nous allons à présent étudier le comportement d'un système linéaire vis-à-vis d'un signal important de la théorie du signal, il s'agit de la (co)sinusoïde. Elle est particulière dans le sens où son spectre ne comporte qu'une seule harmonique, située à la fréquence de la (co)sinusoïde. Il s'agit donc d'un signal élémentaire, car tout signal peut se décomposer en une somme d'harmoniques.

Considérons donc le signal suivant

$$x(t) = e^{j2\pi f_0 t}$$

qui représente un (co)sinusoïde de fréquence  $f_0$  et appliquons-le à l'entrée d'un système linéaire

de réponse impulsionnelle  $h(t)$ . La sortie est alors donnée par

$$\begin{aligned} y(t) &= \int_{-\infty}^{+\infty} h(\tau) x(t - \tau) d\tau \\ &= \int_{-\infty}^{+\infty} h(\tau) e^{j2\pi f_0(t-\tau)} d\tau \\ &= e^{j2\pi f_0 t} \int_{-\infty}^{+\infty} h(\tau) e^{-j2\pi f_0 \tau} d\tau \\ &= e^{j2\pi f_0 t} H(f_0) \end{aligned}$$

On remarque donc que la sortie du système est toujours une (co)sinusoïde de même fréquence mais multipliée par la facteur

$$H(f_0) = \|H(f_0)\| e^{j\theta(f_0)}$$

où  $\|H(f)\|$  et  $\theta(f)$  sont respectivement le module et l'argument du nombre complexe  $H(f)$ . Dès lors, nous avons

$$y(t) = \|H(f_0)\| e^{j(2\pi f_0 t + \theta(f_0))}$$

Le système a donc pour effet de modifier l'amplitude de la (co)sinusoïde par le facteur  $\|H(f_0)\|$ , appelé *gain en fréquence* du système, et d'incrémenter la phase de la (co)sinusoïde d'un terme  $\theta(f_0)$ , appelé *déphasage du système* à la fréquence  $f_0$ . Bien sûr, le gain et le déphasage d'un système n'ont pas, en général, la même valeur pour toutes les fréquences. Par contre, il est important de remarquer qu'un système linéaire ne modifie pas la fréquence d'une (co)sinusoïde, mais uniquement son amplitude et sa phase.

#### 1.4.6 Exemple

Afin d'illustrer ces notions, nous allons considérer un exemple bien précis. Considérons le circuit électrique de la figure 1.5. Il s'agit d'un circuit  $RC$  à l'entrée duquel on applique une tension  $x(t)$  et à la sortie duquel on observe une tension  $y(t)$ . Nous allons mettre ce système en équation afin d'obtenir sa fonction de transfert.

L'application de la tension  $x(t)$  à l'entrée du circuit fait naître un courant électrique  $i(t)$  traversant la résistance  $R$  et la capacité  $C$ . La physique nous apprend que la relation entre le courant  $i(t)$  traversant la capacité  $C$  et la tension  $y(t)$  aux bornes de cette capacité est donnée par

$$i(t) = C \frac{dy}{dt}(t)$$

En appliquant la loi des mailles à la maille de gauche du circuit nous avons

$$x(t) = Ri(t) + y(t)$$

En remplaçant  $i(t)$  par sa valeur dans cette expression, nous obtenons la relation liant l'entrée  $x(t)$  et la sortie  $y(t)$  :

$$x(t) = RC \frac{dy}{dt}(t) + y(t) \quad (1.14)$$

Il s'agit d'une équation différentielle linéaire. Si on connaît le signal  $x(t)$ , il est possible de résoudre cette équation pour obtenir la sortie  $y(t)$  du système. Cette équation définit donc l'opérateur  $O$  du système. Néanmoins, nous sommes encore loin de la réponse impulsionnelle ou de la fonction de transfert.

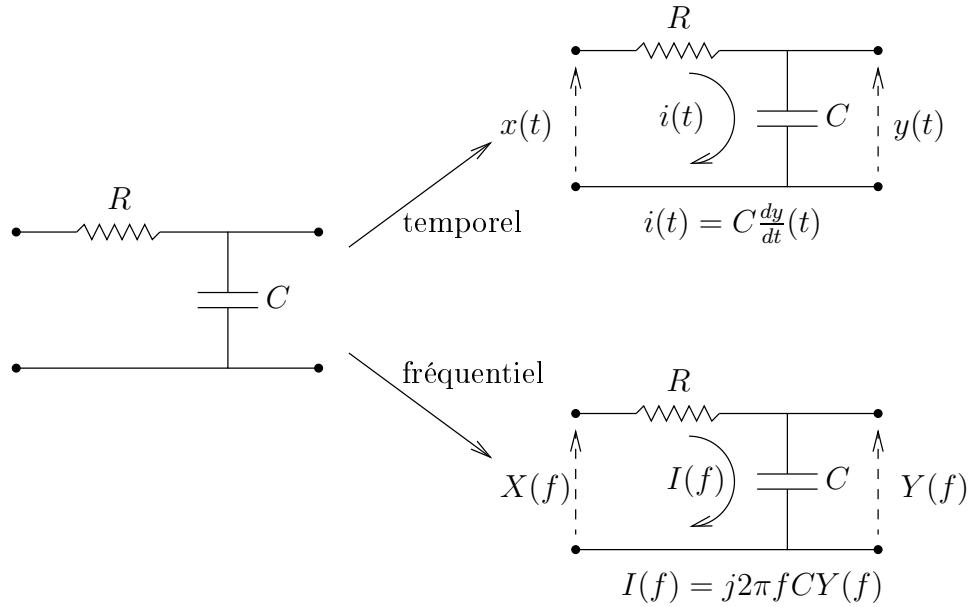


FIGURE 1.5 – Exemple de système linéaire : un circuit électrique et ses représentations temporelle et fréquentielle.

Tentons à présent de dégager la fonction de transfert du système. Pour cela, nous prenons la transformée de FOURIER de la relation (1.14) :

$$X(f) = j2\pi RCf Y(f) + Y(f)$$

En se rappelant de la définition (1.13) de la fonction de transfert d'un système, il vient

$$H(f) = \frac{Y(f)}{X(f)} = \frac{1}{1 + j2\pi RCf} \quad (1.15)$$

En calculant la transformée de FOURIER inverse la fonction de transfert, nous obtenons la réponse impulsionnelle du système :

$$h(t) = \frac{1}{RC} e^{-\frac{t}{RC}} u(t) \quad (1.16)$$

Nous savons donc à présent comment réagit notre système lorsqu'on applique à son entrée une impulsion de DIRAC :

$$\delta(t) \longrightarrow \frac{1}{RC} e^{-\frac{t}{RC}} u(t)$$

Étudions à présent son comportement lorsqu'on lui applique le signal suivant

$$x(t) = A \cos(2\pi f_0 t)$$

On pourraît réaliser la convolution entre  $h(t)$  et  $x(t)$ , mais je ne le conseillerais même pas au lecteur motivé. Le passage par la transformée de FOURIER serait moins douloureux mais néanmoins fastidieux... Nous allons plutôt nous souvenir qu'un système linéaire transforme une cosinusoïde en une autre cosinusoïde dont l'amplitude et la phase ont été modifiées. Par facilité nous mettons la fonction de transfert sous la forme module-argument :

$$H(f) = ||H(f)|| e^{j\theta(f)} = \frac{1}{\sqrt{1 + 4\pi^2 R^2 C^2 f^2}} e^{j \arctan(-2\pi R C f)}$$

Le gain  $\|H(f)\|$  et le déphasage  $\theta(f)$  introduit par ce système (pour  $R = 1 [Ohm]$  et  $C = 0,005 [F]$ ) sont représentés à la figure 1.6.

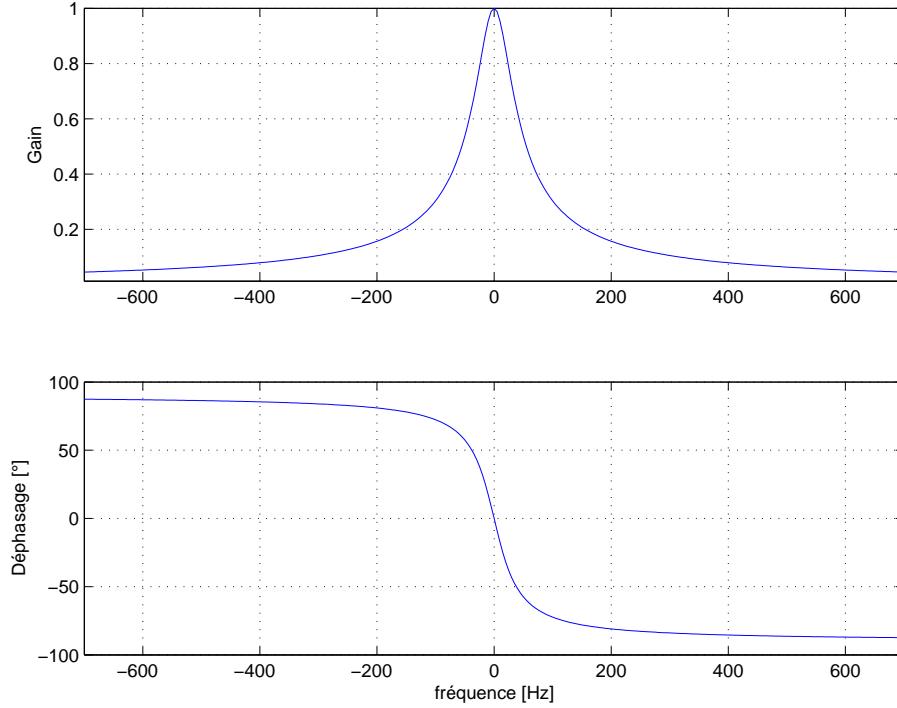


FIGURE 1.6 – Gain  $\|H(f)\|$  et déphasage  $\theta(f)$  du système.

Nous pouvons à présent calculer le signal  $y(t)$  à la sortie du système :

$$y(t) = \frac{A}{\sqrt{1 + 4\pi^2 R^2 C^2 f_0^2}} \cos(2\pi f_0 t + \arctan(-2\pi R C f_0))$$

Nous remarquons donc que le signal subit une atténuation d'amplitude d'autant plus importante que sa fréquence  $f_0$  est élevée. Ce circuit laisse donc passer les harmoniques de fréquence faible et atténue fortement les harmoniques de fréquence élevée. Nous dirons que ce système est un *filtre passe-bas*.

La figure 1.7a illustre le comportement du système ( $R = 1 [Ohm]$  et  $C = 0,005 [F]$ ) pour un signal cosinusoïdal ( $A = 1$  et  $f_0 = 40 [Hz]$ ). À cette fréquence, le signal de sortie a une amplitude égale à

$$\frac{1}{\sqrt{1 + 4\pi^2 R^2 C^2 f_0^2}} = 0,62$$

On observe également le déphasage du signal de sortie.

La figure 1.7b illustre le comportement du même système pour un signal d'entrée  $x(t)$  rectangulaire périodique de fréquence fondamentale  $f_0 = 10 [Hz]$ . Pour rappel, ce signal est constitué d'une infinité d'harmoniques. Chacune de ces harmoniques subit, au travers du système, une atténuation d'amplitude et un déphasage différents. Le signal de sortie  $y(t)$  correspond à la somme de ces harmoniques modifiées. L'observation du signal  $y(t)$  nous montre à nouveau que les hautes fréquences d'un signal (qui ont été fortement atténueres par le système étudié ici) correspondent aux variations rapides du signal, dans ce cas-ci aux transitions brutes du signal de -1 à +1 ou de +1 à -1.

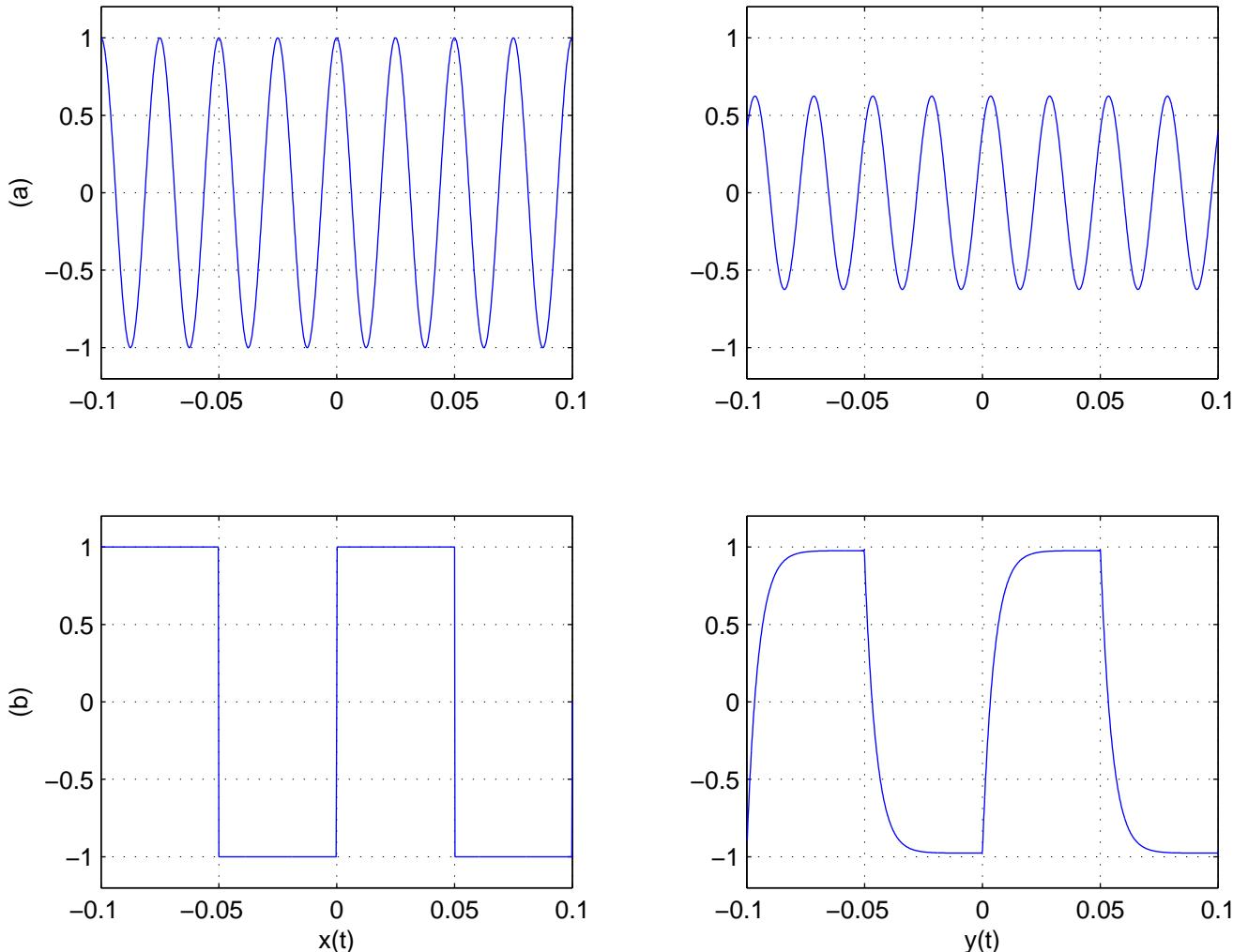


FIGURE 1.7 – Illustration du comportement du système pour  $R = 1$  [Ohm] et  $C = 0,005$  [F] :  
(a) Signal cosinusoïdal ( $A = 1$  et  $f_0 = 40$  [Hz]). (b) Signal rectangulaire périodique de fréquence fondamentale  $f_0 = 10$  [Hz].

## 1.5 Filtre linéaire

Un *filtre* est un système qui réalise une opération de traitement d'un signal. L'action d'un filtre peut consister à retenir, supprimer, modifier des composantes indésirables d'un signal et à en laisser passer librement les éléments utiles. Ces composantes sont bien sûr les harmoniques du signal.

Un filtre est *linéaire* si son action peut se mettre sous la forme du produit de convolution (1.12). Tout filtre linéaire est donc caractérisée par sa réponse impulsionale ou sa fonction de transfert. Un exemple classique de filtre linéaire est celui du circuit électrique  $RC$  du paragraphe précédent. Nous avons déjà dit à son sujet qu'il s'agit d'un filtre passe-bas car il laisse passer les basses fréquences et atténue (mais ne supprime pas !) les hautes fréquences. Il s'agit également d'un filtre tout à fait réalisable physiquement (il suffit en effet de construire le circuit électrique élémentaire de la figure 1.5!).

Nous allons à présent voir qu'il existe d'autres types de filtres caractérisés par la bande des fréquences qu'ils laissent passer, il s'agit des filtres passe-haut et passe-bande. Nous distinguons la notion de *filtre idéal* et de *filtre non-idéal*.

### 1.5.1 Filtre idéal

Un filtre idéal est caractérisé par le fait que sa fonction de transfert ne peut prendre que les valeurs 1 ou 0, en fonction de la fréquence, selon que l'on souhaite garder ou supprimer telle ou telle harmonique. Dès lors, un filtre idéal laisse passer sans aucune modification (d'amplitude ou de phase) certaines harmoniques, et supprime complètement les autres. La bande de fréquence dans laquelle un filtre idéal est non nul est appelée *bande passante* du filtre.

#### Filtre passe-bas idéal

Un *filtre passe-bas* idéal est un filtre qui laisse passer toutes les harmoniques en-dessous d'une fréquence que nous noterons  $W$ , et qui supprime toutes les autres. Sa fonction de transfert est donnée par

$$H(f) = \begin{cases} 1 & \text{si } -W \leq f \leq +W \\ 0 & \text{sinon} \end{cases} \quad (1.17)$$

qui peut encore s'écrire

$$H(f) = \text{rect}\left(\frac{f}{2W}\right)$$

où  $W$  est la *bande passante* du filtre. La valeur particulière de la fréquence  $f = W$  est également appelée *fréquence de coupure* du filtre, car c'est à partir de cette fréquence que les harmoniques de fréquence supérieure sont "coupées". La fonction de transfert du filtre passe-ideal est représentée à la figure 1.8a.

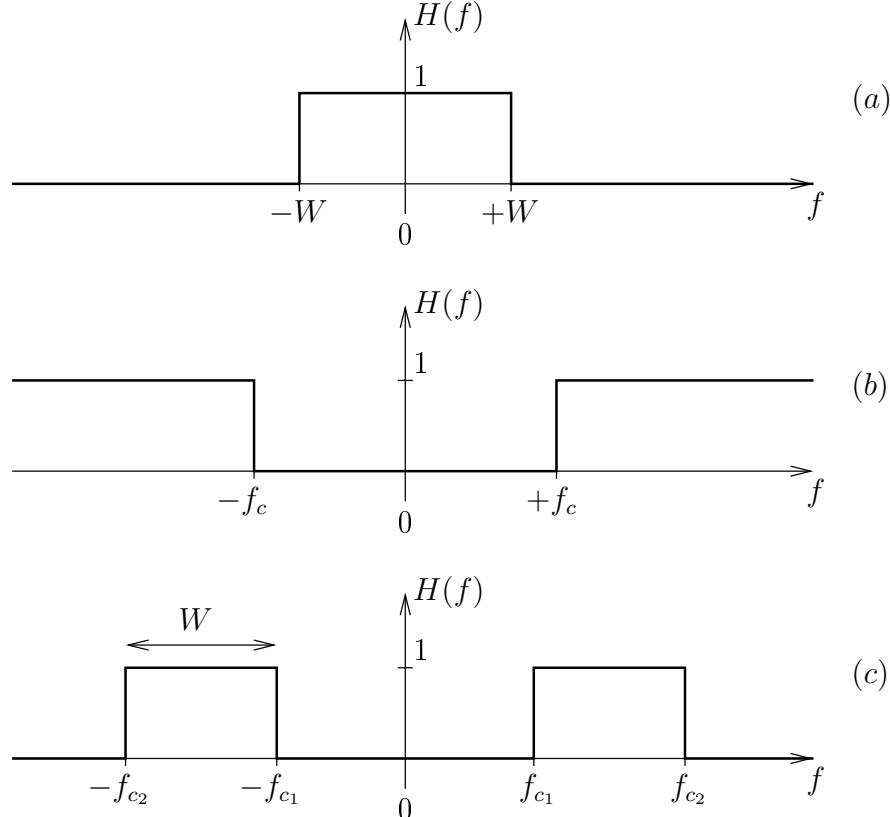


FIGURE 1.8 – Fonction de transfert de filtres idéaux : (a) Filtre passe-bas. (b) Filtre passe-haut. (c) Filtre passe-bande.

La réponse impulsionnelle du filtre s'obtient en calculant la transformée de FOURIER inverse de cette dernière relation :

$$h(t) = 2W \operatorname{sinc}(2Wt)$$

Ici, problème... Rappelons-nous que la réponse impulsionnelle  $h(t)$  d'un système correspond au signal que l'on observerait à la sortie du système si on lui appliquait à son entrée une impulsion de DIRAC  $\delta(t)$ . Or, l'impulsion de Dirac apparaît à l'instant  $t = 0$  tandis que le signal  $\operatorname{sinc}(2Wt)$  à débuté en  $t = -\infty$ ... Cela signifie que l'*effet*, c'est-à-dire la sortie  $h(t)$ , apparaît avant la *cause*  $\delta(t)$ , ce qui est impossible ! C'est pour cela que l'on qualifie ce filtre de *non-causal*, et qu'il est impossible d'en réaliser physiquement (avec un circuit électrique par exemple). Alors, pourquoi en parler ? Parce qu'il constitue un outil très pratique pour la modélisation de systèmes élaborés comme ceux rencontrés en télécommunications. Ensuite, comme on va le voir très rapidement, un filtre idéal peut être implémenté à l'aide d'un simple programme informatique à condition de réaliser un traitement numérique du signal et non plus analogique comme nous l'avons fait jusque maintenant. Encore un peu de patience !

### Filtre passe-haut idéal

Un *filtre passe-haut idéal* réalise l'opération complémentaire d'un filtre passe-bas idéal, c'est-à-dire qu'il laisse passer toutes les harmoniques situées au-dessus d'une certaine fréquence  $f_c$ , appelée *fréquence de coupure*, et qui supprime toutes les autres. La fonction de transfert d'un tel filtre est donnée par

$$H(f) = \begin{cases} 0 & \text{si } -f_c \leq f \leq +f_c \\ 1 & \text{sinon} \end{cases} \quad (1.18)$$

qui peut encore s'écrire

$$H(f) = 1 - \operatorname{rect}\left(\frac{f}{2f_c}\right)$$

Celle-ci est représentée à la figure 1.8b.

Pour un tel filtre, on ne peut pas vraiment parler de bande passante car celle-ci est infinie. Le paramètre important est la fréquence de coupure. Pour la même raison que celle évoquée pour le filtre passe-bas, un filtre passe-haut idéal ne peut être réalisé physiquement.

### Filtre passe-bande idéal

Un *filtre passe-bande idéal* est caractérisé par le fait qu'il laisse passer toutes les harmoniques dans une bande de fréquence délimitée par deux fréquences  $f_{c_1}$  et  $f_{c_2}$  ( $f_{c_1} < f_{c_2}$ ), appelées respectivement *fréquence de coupure inférieure* et *fréquence de coupure supérieure*, et supprime les autres. Donc, un tel filtre supprime les basses et les hautes fréquences et conservent les fréquences "moyennes". Sa fonction de transfert est donnée par

$$H(f) = \begin{cases} 1 & \text{si } f_{c_1} \leq |f| \leq +f_{c_2} \\ 0 & \text{sinon} \end{cases} \quad (1.19)$$

et est représentée à la figure 1.8c. La bande passante de ce filtre correspond à la longueur de la bande de fréquence dans laquelle les harmoniques sont conservées, elle est donc égale à

$$W = f_{c_2} - f_{c_1}$$

### Exemple

Nous allons à présent illustrer le comportement de ces différents filtres idéaux sur un signal rectangulaire périodique de fréquence fondamentale  $f_0 = 10 [Hz]$ . Ce signal, ainsi que son spectre, sont représentés à la figure 1.9a.

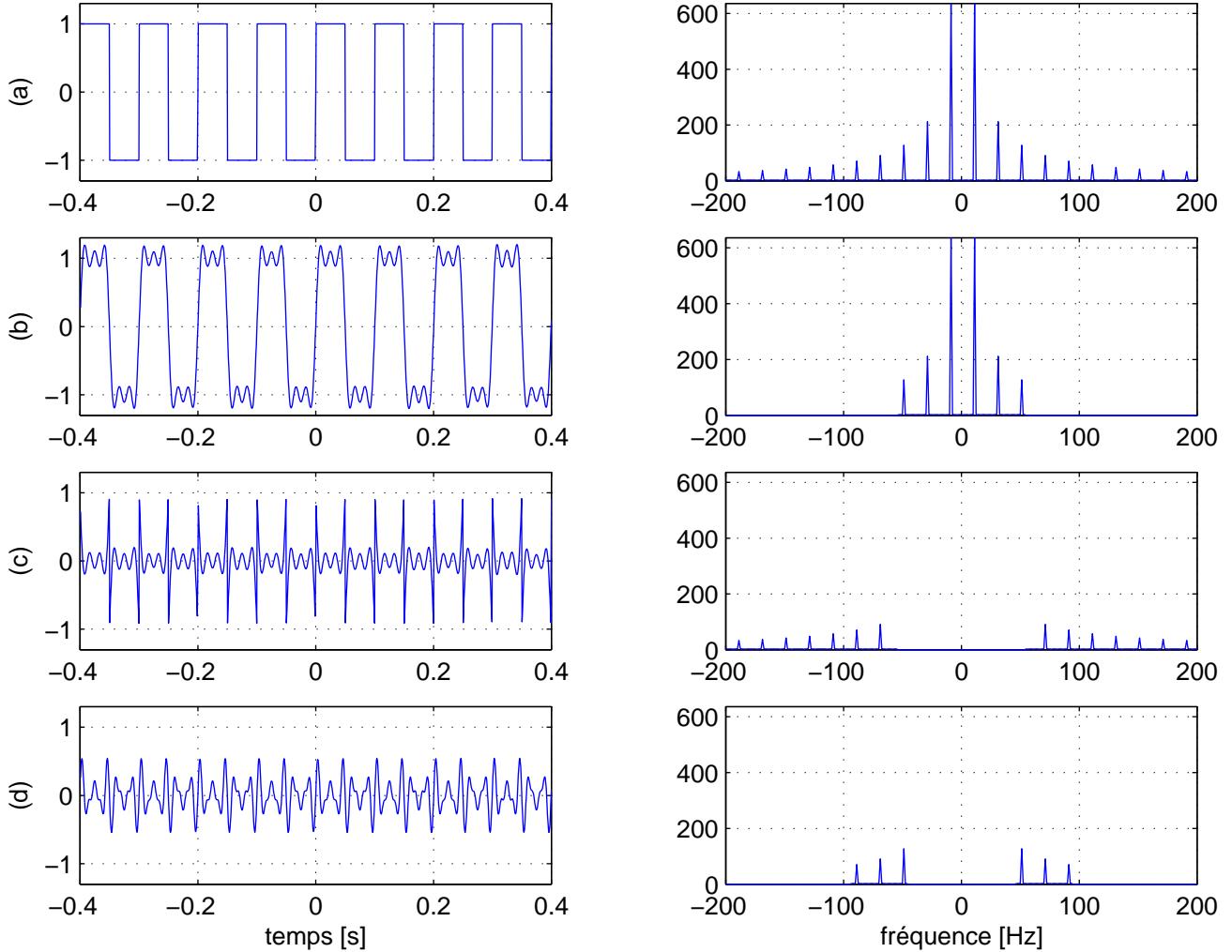


FIGURE 1.9 – Illustration du filtre idéal : (a) Signal avant filtrage. (b) Filtrage passe-bas de fréquence de coupure  $f_c = W = 55 [Hz]$ . (c) Filtrage passe-haut de fréquence de coupure  $f_c = 55 [Hz]$ . (d) Filtrage passe-bande de fréquences de coupure  $f_{c1} = 45 [Hz]$  et  $f_{c2} = 95 [Hz]$ .

La figure 1.9b montre le signal (et son spectre) obtenu par le passage du signal de départ au travers un filtre passe-bas idéal de bande passante (ou fréquence de coupure dans ce cas-ci) égale à  $55 [Hz]$ . Dans le spectre, toutes les harmoniques de fréquence supérieure à  $55 [Hz]$  ont complètement disparu. Dans le graphe temporel, les transitions abruptes (de  $-1$  vers  $+1$ , et de  $+1$  vers  $-1$ ) ont laissé la place à des transitions à pente plus douce, illustrant la disparition des hautes fréquences. Sur les paliers ( $\pm 1$ ), on voit à présent apparaître des ondulations. Ce phénomène est connu sous le nom de *phénomène de GIBBS*, celui-ci apparaît lorsque la fonction de transfert du filtre utilisé présente des transitions brutes, ce qui est nettement le cas pour les filtres idéaux.

Mais, n'a-t-on pas dit qu'il était impossible d'implémenter un filtre passe-bas idéal ? Alors, comment la figure 1.9 peut-elle nous montrer des exemples de signaux filtrés par des filtres

idéaux ? La réponse est dans le fait que ces simulations ont été réalisées de manière numérique et que le problème de non-causalité de ces filtres “disparaît” en quelque sorte (mais rassurez-vous, d’autres problèmes bien piquants apparaissent alors !)... ce qui nous donne la possibilité de construire un filtre passe-bas idéal à l’aide d’un programme simplicime sous Matlab (ou autre comme Java, je dis ça mais je ne dis rien...). Nous y reviendrons...

La figure 1.9c montre le signal (et son spectre) obtenu par le passage du signal de départ au travers un filtre passe-haut idéal de fréquence de coupure égale à  $f_c = 55 \text{ [Hz]}$ . Il s’agit donc du filtre complémentaire du filtre passe-bas précédent. On observe que les harmoniques de fréquence inférieure à  $55 \text{ [Hz]}$  ont complètement disparu. Le graphe temporel nous montre que les hautes fréquences sont toujours bien présentes. Cela se traduit par la présence de “pics” aux places des transitions abruptes (de -1 vers +1, et de +1 vers -1) du signal de départ. Le phénomène de GIBBS est toujours bien visible.

Enfin, la figure 1.9d montre le signal (et son spectre) obtenu par le passage du signal de départ au travers un filtre passe-bande idéal de fréquence de coupure inférieure égale à  $f_{c_1} = 45 \text{ [Hz]}$  et de fréquence de coupure supérieure égale à  $f_{c_2} = 95 \text{ [Hz]}$ . La bande passante de ce filtre est donc égale à  $W = 50 \text{ [Hz]}$ . Bien qu’il soit beaucoup plus difficile de donner une interprétation au signal obtenu, le filtre passe-bande n’en est pas moins important, il constitue un élément fondamental des systèmes de télécommunications.

### 1.5.2 Filtre non-idéal

Nous venons de voir que les filtres linéaires idéaux ont des caractéristiques remarquables. Ils permettent de sélectionner précisément les harmoniques que l’on souhaite conserver et celles que l’on désire supprimer, et cela sans aucune modification de l’amplitude et de la phase des harmoniques. Leur gros défaut (et pas des moindres !) est qu’ils ne sont pas réalisables en pratique, à l’aide de composants électriques par exemple.

Dans la pratique, on utilise donc des filtres bien réels qui tentent au maximum de se rapprocher du filtre idéal souhaité. Ces filtres sont qualifiés de *filtres non-idéaux*, dans le sens où leur fonction de transfert ne vaut plus 1 ou 0 comme dans le cas des filtres idéaux. Celle-ci est maintenant une véritable fonction  $H(f)$  à valeurs complexes (en général), ce qui provoque un déphasage mais également une atténuation des harmoniques. De plus, les filtres idéaux ne suppriment pas complètement les harmoniques indésirables, ils ne font que les atténuer fortement.

Un exemple de filtre linéaire non idéal est le circuit de la figure 1.5 que nous avons déjà étudié. Il s’agit d’un filtre linéaire *non-idéal* passe-bas. Pour ce genre de filtre, il est plus difficile de définir la notion de bande passante. Il n’y a pas en effet de transition nette entre les fréquences que l’on souhaite conserver et celles que l’on souhaite supprimer.

## 1.6 Numérisation

Jusqu'à présent, nous avons essentiellement parlé des signaux analogiques. Il est à présent grand temps d'aborder une autre grande catégorie de signal : les signaux numériques. Ceux-ci ne sont plus caractérisés par une fonction mathématique connue en tout temps. Ils sont constitués d'un ensemble (ou plutôt *séquence*) discret de valeurs numériques. Ces valeurs numériques peuvent provenir directement de l'*échantillonnage* d'un signal analogique, ou alors d'un *codage* quelconque. Lorsque les valeurs numériques sont binaires (c'est-à-dire 1 ou 0), on parle de signal numérique binaire.

On peut donc représenter un signal numérique de la manière suivante :

$$\{x_k\} = \dots, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, x_2, x_3, \dots \quad (1.20)$$

où  $x_k$  ( $k$  étant un nombre entier quelconque appelé *indice*) est un nombre réel ou complexe. Dans le cas où les  $x_k$  valent 0 ou 1, on parle de signal numérique *binaire*. En général, on parle de *séquence* numérique car les valeurs  $x_k$  ont un ordre bien précis spécifié par la valeur de l'indice. Cet ordre est très souvent lié à notion de *temps*, mais il n'en est pas toujours ainsi.

Par exemple, les valeurs  $x_k$  peuvent correspondre aux valeurs d'un signal analogique  $g(t)$  "observé" à des instants bien précis :

$$x_k = g(kT_s) \quad k = -\infty, \dots, +\infty$$

Dès lors, le signal numérique correspondant est

$$\{x_k\} = \dots, g(-3T_s), g(-2T_s), g(-T_s), g(0), g(T_s), g(2T_s), g(3T_s), \dots$$

La quantité  $g(kT_s)$  est appelé *échantillon* du signal  $g(t)$  à l'instant  $t = kT_s$ , et l'action de "collecter" les différents échantillons du signal est appelée *échantillonnage* du signal  $g(t)$ . Par abus de langage, nous dirons que  $x_k$  (pour une valeur de  $k$  donnée) est un échantillon du signal numérique  $\{x_k\}$ . Dans cet exemple, le signal analogique  $g(t)$  est "observé" à intervalle de temps régulier, ici toutes les  $T_s$  secondes. Cette grandeur,  $T_s$ , est appelée *période d'échantillonnage* du signal analogique. Par la même occasion, on définit la *fréquence d'échantillonnage*

$$f_s = \frac{1}{T_s} \quad (1.21)$$

exprimée en [Hz] ou en [echantillon/sec].

Insistons sur le fait qu'un signal numérique n'est pas nécessairement lié à un signal analogique. Considérons un texte contenant la phrase suivante :

*...ceci est une phrase...*

Si on remplace chaque caractère par son code ASCII correspondant, nous obtenons le signal numérique suivant

$$\{y_k\} = \dots, 99, 101, 99, 105, 32, 101, 115, 116, 32, 117, 110, 101, 32, 112, 104, 114, 97, 115, 101, \dots$$

qui n'a aucun rapport avec un quelconque signal analogique. On voit donc dans ce dernier exemple, que l'on a remplacé chaque caractère par son *code* ASCII correspondant ; cette opération est appelée *codage*. Dans la continuation de cet exemple, on pourrait, à partir de ce signal

numérique, construire un signal numérique *binaire* en remplaçant par exemple chaque nombre entier par sa valeur binaire. En se limitant aux trois premiers échantillons, nous obtenons

$$\{z_k\} = \dots, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, \dots$$

Mais alors, comment savoir où commencent et finissent les caractères ? Et voilà... Les problèmes du numérique commencent... Ici, il s'agit du problème important de la *synchronisation*, bien connu des programmeurs réseau. Nous y reviendrons.

Dans notre vie de tous les jours, une part importante de l'information que nous traitons est essentiellement visuelle et auditive (bien que certains informaticiens fous pourraient se contenter de 1 et de 0 :-). Ce type d'information est, de par nature, analogique. Par contre, les technologies récentes de l'information sont pour la plupart numérique, et même binaire... Par exemple, un disque dur ou un DVD sont des supports permettant de stocker photos, vidéos, ... sous la forme d'un flot de 1 et de 0. Les réseaux informatiques de part le monde véhiculent quantité de photos ou de vidéos sous la forme élémentaire de 1 ou de 0 transmis séquentiellement. Il a donc fallu mettre au point des techniques permettant de transformer des signaux analogiques en signaux numériques binaires et cela sans (trop) détériorer la qualité de l'information originale. Cette transformation porte le nom de *numérisation*.

Comme nous allons le voir dans cette section, le *processus de numérisation* nécessite plusieurs opérations dont certaines ont déjà été citées plus haut :

1. *échantillonnage* du signal analogique,
2. *quantification* des échantillons, qui se fait, comme nous allons le voir, avec une certaine perte d'information, et,
3. *codage* des échantillons quantifiés.

La quantification est une étape dont nous n'avons pas encore parlé. Dans un des exemples précédents, lorsque nous avons remplacé chaque valeur numérique de code ASCII par sa valeur binaire, cela s'est fait sans aucune quantification. Expliquons-nous. Il existe, dans la version de base du code ASCII, 256 valeurs de codes qui peuvent donc être représentées sur 8 bits au maximum. Dans le cas d'un signal analogique  $g(t)$  échantillonné, les valeurs d'échantillon  $g(kT_s)$  peuvent prendre n'importe quelle valeur dans l'intervalle théorique  $]-\infty, +\infty[$ . Il y a donc une infinité de valeurs possibles pour chaque échantillon, ce qui voudrait dire qu'il faudrait une infinité de bits pour coder un échantillon ! Ceci n'est pas envisageable... L'étape de quantification, qui est réalisée juste entre l'échantillonnage et le codage, consiste à remplacer chaque échantillon  $g(kT_s)$  par sa valeur quantifiée (qui peut être proche ou non de la valeur exacte de l'échantillon) en introduisant ce que l'on appelle une *erreur de quantification* qui se traduit par une perte d'informations. Les valeurs quantifiées existent en nombre fini et dès lors, chaque valeur quantifiée peut être représentée par un nombre fini de bits. Nous reviendrons en détails sur la quantification dans la suite.

Étudions à présent en détails, les différentes étapes du processus de numérisation.

### 1.6.1 Échantillonnage

Soit  $g(t)$  un signal analogique à énergie finie, connu pour tout temps  $t$ . Supposons que l'on prenne un échantillon de ce signal à une cadence régulière, soit un échantillon toutes les  $T_s$  secondes. La séquence ainsi obtenue est notée  $\{g(kT_s)\}$  où  $k$  est un entier et  $T_s$  est la période d'échantillonnage. Pour rappel, son inverse  $f_s = 1/T_s$  est la fréquence d'échantillonnage. Cette forme d'échantillonnage est appelée échantillonnage instantané. L'important théorème qui suit

fournit les conditions nécessaires à l'obtention d'un échantillonnage de "qualité".

**Théorème de SHANNON.** Une fonction  $g(t)$  à énergie finie et à spectre limité, c'est-à-dire dont la transformée de FOURIER est nulle pour  $|f| > W$ , est entièrement déterminée par ses échantillons  $g(kT_s)$  pour autant que la fréquence échantillonnage  $f_s$  soit strictement supérieure au double de la plus haute fréquence du signal :

$$f_s > 2W$$

Avant de démontrer ce théorème important, notons que si les conditions du théorème de SHANNON sont vérifiées, il n'y a aucune perte d'information lors de l'opération d'échantillonnage, alors que l'on a l'impression de "perdre" les valeurs du signal entre les instants d'échantillonnage. Nous verrons qu'il est possible de retrouver ces valeurs "perdues" grâce à l'opération inverse de l'échantillonnage. Mais avant cela, démontrons le théorème...

Mathématiquement, la fonction échantillonnée, que nous noterons  $g_s(t)$ , peut être représentée par un train d'impulsions de DIRAC pondérées par la valeur des échantillons. En d'autres mots,  $g_s(t)$  est obtenue en multipliant la fonction  $g(t)$  par le train d'impulsions de DIRAC suivant :

$$\sum_{k=-\infty}^{+\infty} \delta(t - kT_s) \Rightarrow f_s \sum_{n=-\infty}^{+\infty} \delta(f - n f_s)$$

La situation est représentée à la figure 1.10.

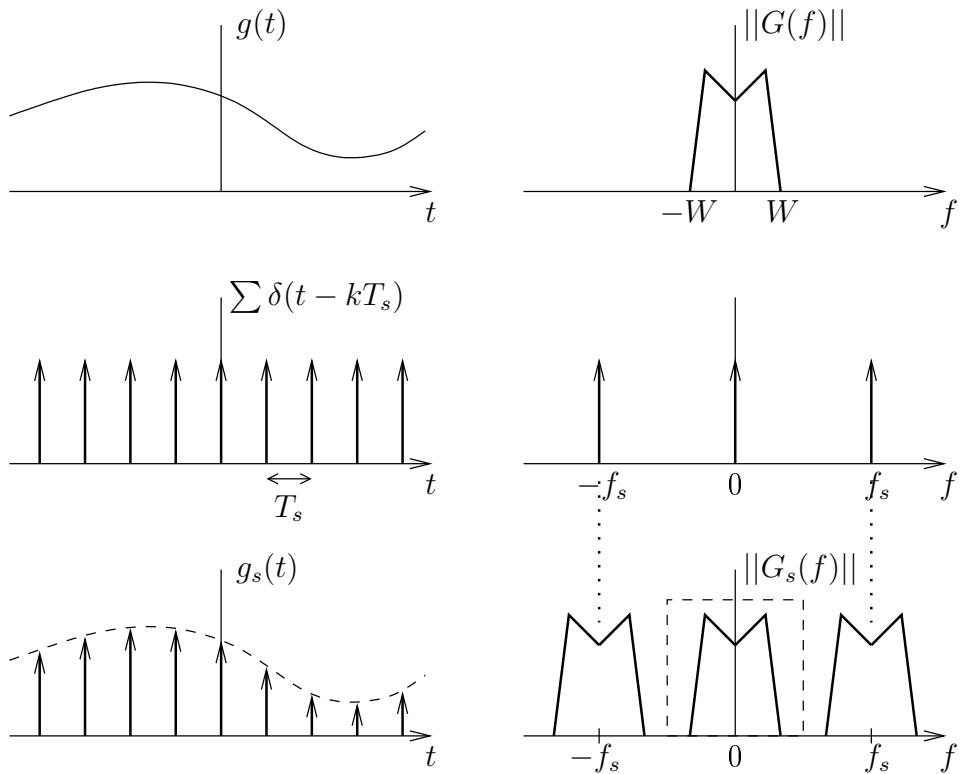


FIGURE 1.10 – Échantillonnage instantané.

La fonction échantillonnée peut donc s'exprimer par

$$\begin{aligned} g_s(t) &= \sum_{k=-\infty}^{+\infty} g(kT_s) \delta(t - kT_s) \\ &= g(t) \sum_{k=-\infty}^{+\infty} \delta(t - kT_s) \end{aligned}$$

La fonction échantillonnée  $g_s(t)$  étant le produit de  $g(t)$  par un train d'impulsions de DIRAC, sa transformée de FOURIER  $G_s(f)$  est le produit de convolution de  $G(f)$  et de la transformée de FOURIER du train d'impulsions de DIRAC qui est elle-même un train d'impulsions de DIRAC :

$$G_s(f) = G(f) \otimes \left\{ f_s \sum_{n=-\infty}^{+\infty} \delta(f - nf_s) \right\}$$

En calculant ce produit de convolution, nous obtenons

$$G_s(f) = f_s \sum_{n=-\infty}^{+\infty} G(f - nf_s) \quad (1.22)$$

*La transformée de FOURIER de la fonction échantillonnée s'obtient donc en répétant la transformée de FOURIER de  $g(t)$ , centrée sur tous les multiples de la fréquence d'échantillonnage  $f_s$ .* Ceci est illustré à la figure 1.10 (en bas à droite). On trouve donc dans le spectre  $G_s(f)$  la “copie” du spectre  $G(f)$  sans altération. Néanmoins, on ne pourra reconstituer  $g(t)$ , c'est-à-dire  $G(f)$ , à partir de  $g_s(t)$ , c'est-à-dire  $G_s(f)$ , que si les répliques de  $G(f)$  ne se recouvrent pas. Cela n'est possible que si la bande passante de  $g(t)$  est limitée à une valeur  $W$  et que la fréquence d'échantillonnage est strictement supérieure à  $2W$ .

L'opération de reconstruction de  $g(t)$  à partir de sa version échantillonnée  $g_s(t)$  est très simple ; il suffit d'utiliser un filtre passe-bas dont la fréquence de coupure est comprise entre  $W$  et  $f_s - W$  et ayant un gain  $T_s$  dans la bande passante.

Remarquons que la contrainte  $f_s > 2W$  est souvent ramenée à  $f_s \geq 2W$ . Il s'agit d'une erreur dans le cas des signaux sinusoïdaux. En effet, il suffit d'échantillonner une sinusoïde de fréquence  $f_0$  à une fréquence  $2f_0$  au droit de ses passages par zéro. Tous les échantillons de cette séquence sont nuls et il n'est dès lors pas possible de reconstituer la sinusoïde.

### Reconstruction : Formule d'interpolation de WHITTAKER

Comme nous venons de le voir, pour reconstruire la fonction  $g(t)$  à partir de la fonction échantillonnée  $g_s(t)$ , il suffit d'appliquer un filtre passe-bas dont la fréquence de coupure est comprise entre  $W$  et  $f_s - W$ , et dont le gain est égal à  $T_s$ . La formule de reconstruction (ou d'interpolation) de WHITTAKER est basée sur le choix d'un filtre passe-bas idéal de fréquence de coupure égale à la moitié de la fréquence d'échantillonnage  $f_s/2$ . La fonction de transfert du filtre de reconstruction est donc donnée par

$$H_r(f) = T_s \text{rect} \left( \frac{f}{f_s} \right)$$

dont la réponse impulsionnelle est

$$h_r(t) = \text{sinc}(f_s t) = \text{sinc} \left( \frac{t}{T_s} \right)$$

Le signal reconstruit, que nous noterons  $g_r(t)$ , est donc le résultat de la convolution de  $g_s(t)$  et de  $h_r(t)$  :

$$\begin{aligned} g_r(t) &= g_s(t) \otimes h_r(t) \\ &= \left( \sum_{k=-\infty}^{+\infty} g(kT_s) \delta(t - kT_s) \right) \otimes \text{sinc}\left(\frac{t}{T_s}\right) \\ &= \sum_{k=-\infty}^{+\infty} g(kT_s) \text{sinc}\left(\frac{t - kT_s}{T_s}\right) \end{aligned}$$

Il en résulte :

[Formule d'interpolation de WHITTAKER]. Soit  $g(t)$  un signal analogique à bande limitée  $[-W, +W]$ . Soit  $\{g(kT_s)\}$  la séquence de ses échantillons de pas  $T_s = 1/f_s$ . La fonction  $g(t)$  s'écrit comme la série de fonctions

$$\sum_{k=-\infty}^{+\infty} g(kT_s) \text{sinc}\left(\frac{t - kT_s}{T_s}\right) \quad (1.23)$$

L'interprétation de cette formule est assez simple. La fonction  $\text{sinc}\left(\frac{t - kT_s}{T_s}\right)$  est égale à 1 pour  $t = kT_s$  et égale à 0 pour tous les  $t = iT_s$ , tels que  $i \neq k$ . La somme de la série (1.23) se réduit donc à  $g(kT_s)$  pour  $t = kT_s$ . La formule signifie qu'entre ces valeurs cette série interpole *exactement*  $g(t)$ .

### Filtrage préalable à l'échantillonnage

Le théorème de SHANNON nous a appris que pour échantillonner correctement un signal, il faut choisir une fréquence d'échantillonnage supérieure au double de la plus haute fréquence du signal. Or, faut-il encore que le signal à échantillonner ait une bande passante limitée ! En effet, l'information analogique n'est naturellement pas à bande passante limitée. L'opération d'échantillonnage introduit alors un recouvrement entre les différentes répliques du spectre du signal de départ ; ce phénomène est appelé *repli de spectre* ou *aliasing*. L'opération de filtrage passe-bas (lors de la reconstruction) s'avère incapable de supprimer ce recouvrement et on risque de voir apparaître une série de fréquences initialement absentes du signal de départ. Deux approches permettent de supprimer ces effets indésirables :

1. Un filtrage passe-bas préalable à l'échantillonnage rend le signal à échantillonner à bande limitée.
2. Le signal est échantillonné à une fréquence légèrement supérieure au double de la plus haute fréquence du signal.

En raison de la difficulté qu'il y a à réaliser un filtre ayant un flanc raide au droit de la fréquence de coupure, il est d'usage de définir une bande de garde dans laquelle la transition est plus douce. La bande de garde, typiquement de l'ordre de 10 à 20%, entraîne donc une augmentation de la fréquence d'échantillonnage. On choisit donc couramment

$$f_s = 2,2 W$$

**Exemple.** Soit à produire un signal numérique à partir d'un signal musical s'étendant jusqu'à 20 [kHz]. En toute rigueur, une fréquence d'échantillonnage supérieure à 40 [kHz] suffit. Pour des questions de réalisation de filtre, on utilise plutôt 44,1 [kHz] dans le standard CD-Audio.

Afin d'illustrer le processus d'échantillonnage, nous fournissons ci-dessous 4 exemples de simulations réalisées sous Matlab. Pour ces exemples, aucun filtrage préalable à l'échantillonnage n'a été réalisée.

### Exemple 1

Considérons le signal suivant

$$g(t) = e^{-\frac{(t-0,4)^2}{0,02}} - e^{-\frac{(t-0,6)^2}{0,003}}$$

Il s'agit de la différence de deux gaussiennes dont le spectre est connu pour être à faible bande passante. On choisit une fréquence d'échantillonnage  $f_s = 64 [Hz]$ . La figure 1.11 illustre le processus d'échantillonnage : (a) Signal de départ  $g(t)$ , (b) Train d'impulsions de Dirac non pondérées, (c) Signal échantillonné  $g_s(t)$  et (d) Signal reconstruit  $g_r(t)$  en appliquant un filtre passe-bas idéal à la fonction échantillonnée  $g_s(t)$ .

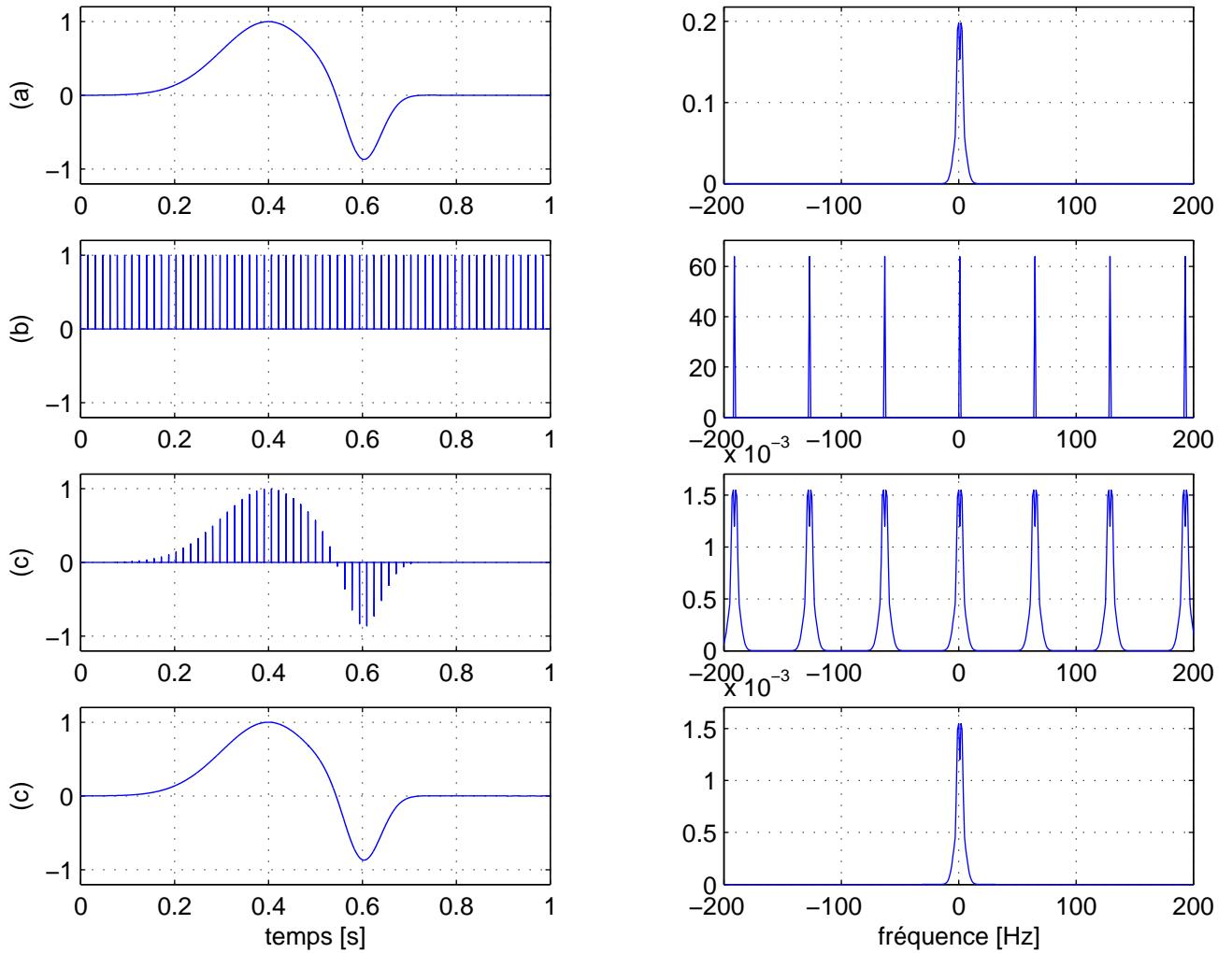


FIGURE 1.11 – Signal gaussien échantillonné à la fréquence  $f_s = 64 [Hz]$ .

On observe que le signal de départ  $g(t)$  est bien à bande limitée et que les répliques de son spectre sont bien distinctes dans le spectre de la fonction échantillonnée. Il n'y a donc pas d'aliasing ici. L'application du filtre passe-bas idéal de reconstruction, de bande passante égale à  $32 [Hz]$ , permet de reconstruire parfaitement le signal de départ.

### Exemple 2

Dans cet exemple, on considère le même signal  $g(t)$  que dans l'exemple 1. Par contre, on choisit à présent une fréquence d'échantillonnage  $f_s = 16 [Hz]$ . Le résultat de la simulation est fourni à la figure 1.12.

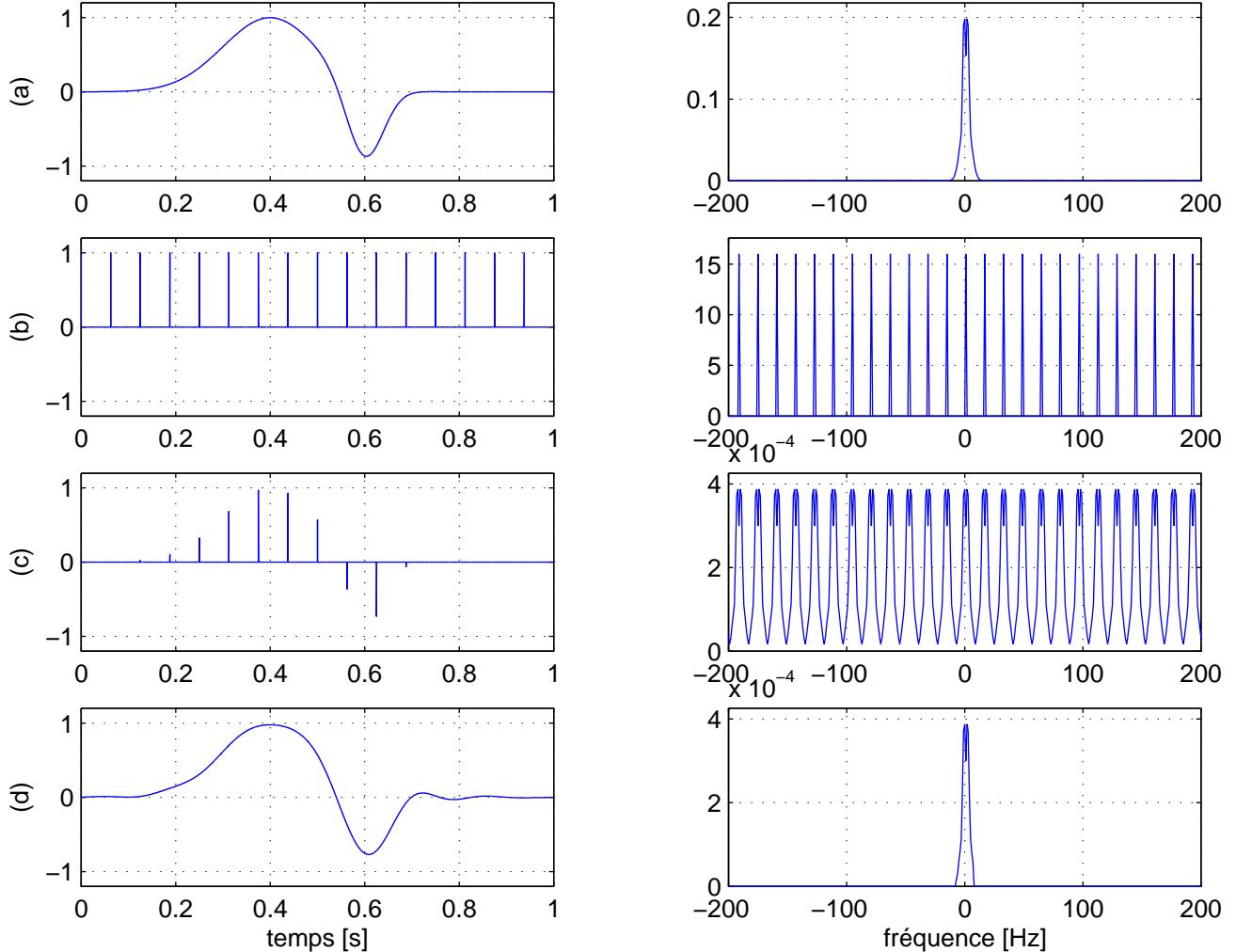


FIGURE 1.12 – Signal gaussien échantillonné à la fréquence  $f_s = 16 [Hz]$ .

La bande passante du signal de départ n'a donc pas changé. Par contre, la fréquence d'échantillonnage est nettement plus faible. On voit à présent apparaître, au niveau de la fonction échantillonnée, des chevauchements entre les répliques du spectre de  $g(t)$ ; l'aliasing est bien présent. L'application du filtre de reconstruction (de bande passante égale à 8 [Hz]) fournit un signal reconstruit  $g_r(t)$  dégradé par rapport au signal de départ  $g(t)$ . Cela n'a pas l'air si grave que cela mais ça aurait pu être pire... Voyons cela.

### Exemple 3

Considérons le signal

$$g(t) = \sin(2\pi 10t)$$

Il s'agit donc d'une sinusoïde d'amplitude 1 et de fréquence  $f_0 = 10 [Hz]$ . Nous choisissons ici une fréquence d'échantillonnage  $f_s = 64 [Hz]$ . Les conditions du théorème de SHANNON sont donc amplement satisfaites :  $64 > 2 \times 10$ . Les résultats de la simulation sont fournis à la figure

1.13. L'application du filtre passe-bas de bande passante 32 [Hz] fournit un signal reconstruit identique au signal de départ.

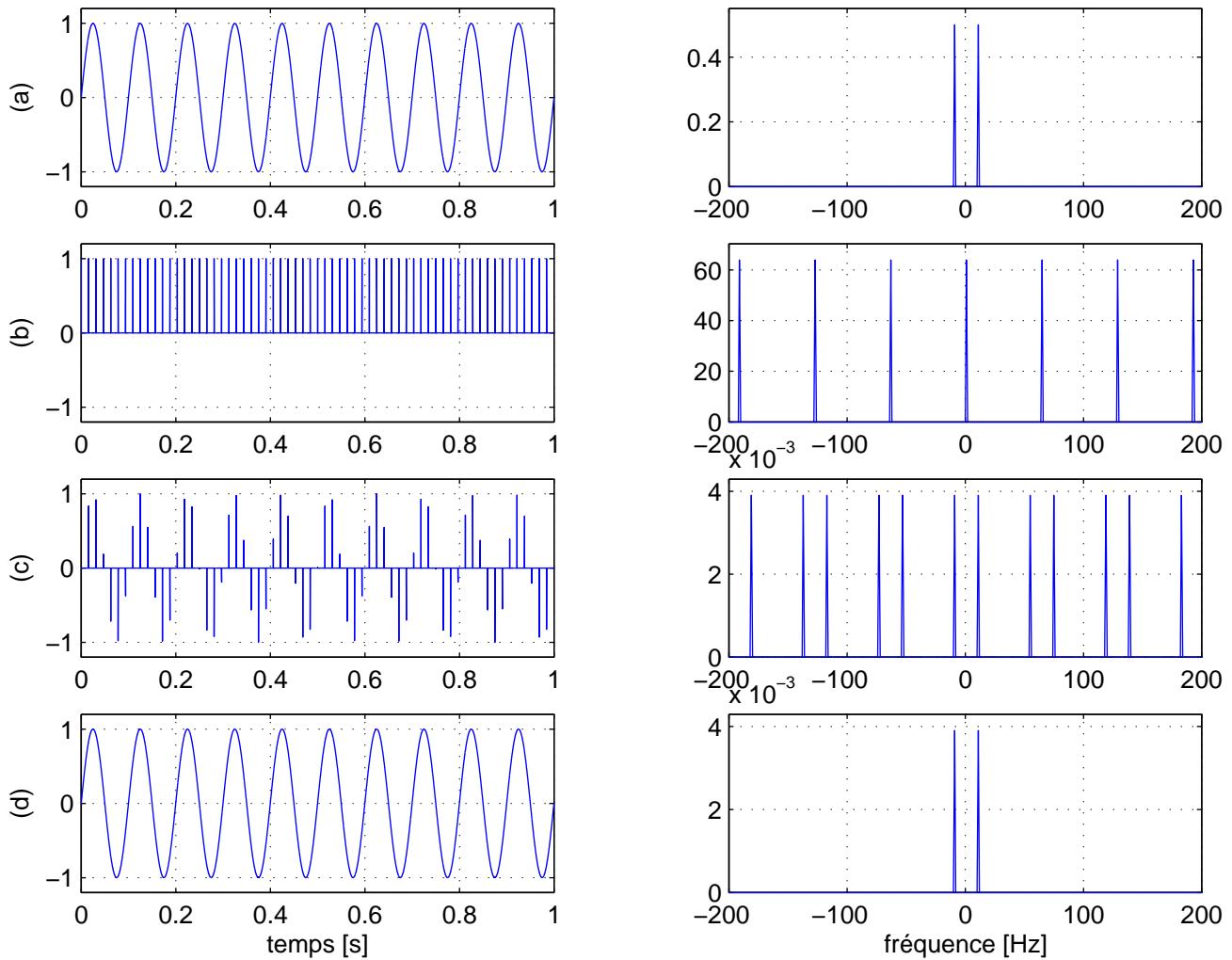


FIGURE 1.13 – Signal sinusoïdal de fréquence 10 [Hz] échantillonné à la fréquence  $f_s = 64$  [Hz].

#### Exemple 4

Dans cet exemple, nous considérons le même signal sinusoïdal qu'à l'exemple 3. Nous choisissons à présent une fréquence d'échantillonnage  $f_s = 16$  [Hz]. Les conditions du théorème de SHANNON ne sont donc plus satisfaites, en effet :

$$f_s = 16 \text{ [Hz]} < 2f_0 = 20 \text{ [Hz]}$$

Les résultats de la simulation sont fournis à la figure 1.14.

Le résultat de la reconstruction est assez surprenant ! Le signal reconstruit  $g_r(t)$ , bien que toujours sinusoïdal, n'a plus du tout la même fréquence que celle du signal de départ. Le processus d'échantillonnage a donc supprimé des fréquences et en a fait apparaître des nouvelles... La nouvelle fréquence de 6 [Hz] apparue est le résultat du *repli* de la fréquence  $f_0 = 10$  [Hz] autour de la fréquence de coupure du filtre passe-bas  $f_s/2 = 8$  [Hz]. C'est pour cela que l'aliasing porte le nom français de *repli de spectre*.

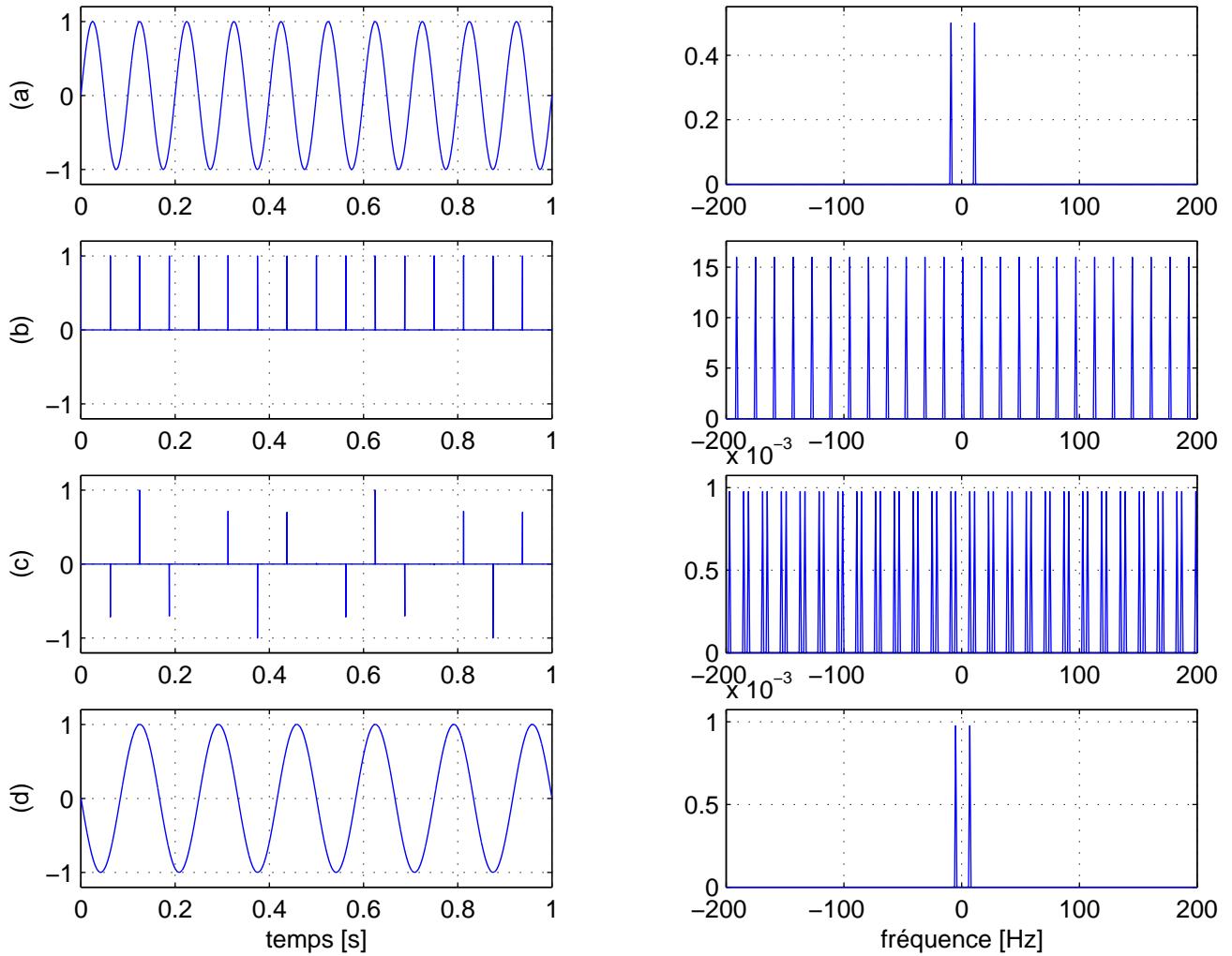


FIGURE 1.14 – Signal sinusoïdal de fréquence 10 [Hz] échantillonné à la fréquence 16 [Hz].

### 1.6.2 Quantification

Jusqu'à présent, nous avons étudié la première étape du processus de numérisation, à savoir la transformation d'un signal analogique  $g(t)$  en une séquence numérique  $\{g(kT_s)\}$ , en utilisant le processus d'échantillonnage :

$$g(t) \longrightarrow \{g(kT_s)\}$$

Pour un programme informatique, le traitement du signal  $g(t)$  peut commencer directement sur la séquence  $\{g(kT_s)\}$  en supposant que les échantillons sont représentés sous la forme de **double** ou **float**. Pour le stockage (sur CD ou DVD par exemple) ou la transmission sur la couche physique, il nous faut des bits. Comme, nous l'avons déjà mentionné, il existe une infinité de valeurs possibles pour les échantillons  $g(kT_s)$  et il nous faudrait donc une infinité de bits pour représenter un échantillon ; ce qui n'est pas pensable. Dès lors, chaque valeur d'échantillon va être remplacée par une valeur quantifiée faisant partie d'un ensemble fini de valeurs que nous appellerons niveaux discrets de quantification et noterons  $v_i$ ,  $i = 0, 1, 2, \dots, L - 1$ . Il y aura donc  $L$  niveaux de quantifications. Cela signifie que lors du processus de quantification, chaque valeur d'échantillon  $g(kT_s)$  va être remplacé par une des  $L$  valeurs  $v_i$  possibles.

**Définition [Quantification].** Le processus consistant à transformer un échantillon d'amplitude  $g(kT_s)$  en une amplitude  $v$  choisie dans un ensemble fini de valeurs possibles  $\{v_0, v_1, \dots, v_{L-1}\}$  est appelé quantification.

Évidemment, les valeurs de quantification  $v_i$  ne sont pas choisies de manière aléatoire, elles approchent au mieux les valeurs exactes des échantillons. Plus le nombre de niveaux de quantification sera élevé, meilleure sera la qualité de l'approximation. Cette *approximation* est responsable d'une *perte d'informations* par rapport au signal de départ ; cette perte est appelée *erreur de quantification*. L'erreur de quantification sera dès lors d'autant plus faible que le nombre de niveaux de quantification  $L$  sera élevé.

Le processus de quantification peut se représenter de la manière suivante :

$$\{g(kT_s)\} \longrightarrow \{q_k\}$$

où  $q_k$  appartient à l'ensemble fini  $\{v_0, v_1, \dots, v_{L-1}\}$  et représente la valeur quantifiée de  $g(kT_s)$ . La quantification transforme donc un signal numérique à valeurs réelles, continues dans l'intervalle théorique  $]-\infty, +\infty[$ , en un autre signal numérique à valeurs réelles, discrètes et prises dans l'ensemble  $\{v_0, v_1, \dots, v_{L-1}\}$ .

Mais à présent, il reste deux inconnues : combien de niveaux  $L$  de quantification doit-on choisir ? comment choisir les différents niveaux de quantification  $v_i$  ? Voyons cela... Le nombre de niveaux de quantification est directement lié à la qualité de la numérisation que l'on désire obtenir. Si l'on souhaite une faible erreur de quantification, il faudra choisir  $L$  élevé. Nous verrons très bientôt que ce choix dépend également d'autres paramètres comme par exemple, la bande passante du système de transmission utilisé, l'espace de stockage disponible ou encore la nécessité de faire du temps réel ou pas. Nous y reviendrons donc.

Le choix des valeurs  $v_i$  dépend essentiellement du signal  $g(t)$  que l'on désire numériser. En effet, il est inutile de quantifier des valeurs qui ne sont pas prises par le signal  $g(t)$ . Par exemple, si  $g(t)$  est un signal qui varie entre les valeurs -5 et +5, inutile de quantifier des valeurs supérieures à 5 ou inférieures à -5. Pour choisir les valeurs  $v_i$ , nous allons donc nous limiter à la *dynamique du signal*. La dynamique d'un signal  $g(t)$  correspond à l'intervalle de valeurs comprises entre la plus petite valeur prise par le signal,  $g_{min}$ , et la plus grande valeur prise par le signal,  $g_{max}$ , c'est-à-dire  $[g_{min}, g_{max}]$ . Quelque soit l'instant  $t$  où l'on échantillonne le signal  $g(t)$ , la valeur  $g(t)$  sera toujours comprise dans cet intervalle. La figure 1.15 illustre la notion de dynamique du signal.

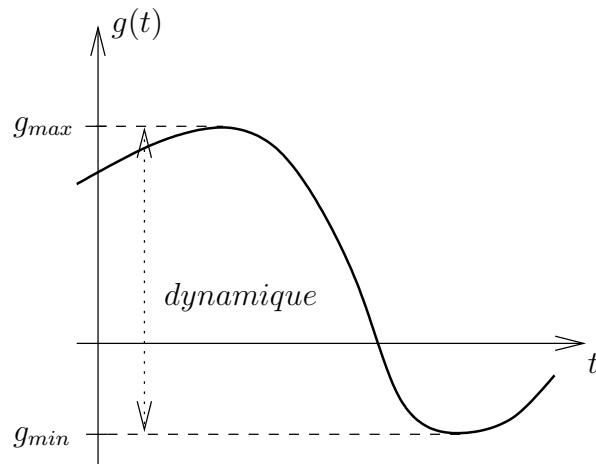


FIGURE 1.15 – Dynamique d'un signal.

La dynamique du signal est ensuite “découpée” en  $L$  intervalles  $I_i$  de telle sorte que l’union des  $L$  intervalles reforment la dynamique complète du signal  $g(t)$ . À chaque intervalle  $I_i$ , on associe une valeur  $v_i$  de telle sorte que chaque échantillon  $g(kT_s)$  dont la valeur tombe dans l’intervalle  $I_i$  soit remplacé par  $v_i$ . C’est donc à ce niveau-ci que se produit l’erreur de quantification. Afin de minimiser cette erreur, il est donc préférable de choisir la valeur  $v_i$  au milieu de l’intervalle  $I_i$ . La fonction caractéristique de la quantification prend alors la forme d’une fonction en escalier, appelée *courbe de quantification*. Ceci est illustré à la figure 1.16. La figure illustre également la quantification d’un échantillon  $g(kT_s)$ , tombant par exemple dans l’intervalle  $I_4$ , en la valeur quantifiée  $v_4$ .

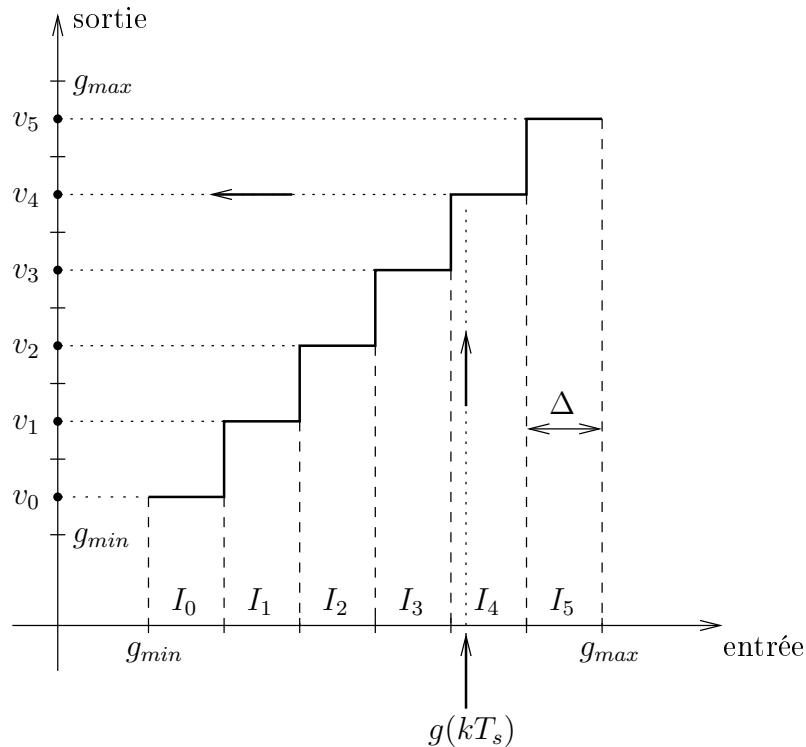


FIGURE 1.16 – Courbe de quantification linéaire à  $L = 6$  niveaux.

La largeur de chaque intervalle  $I_i$  est appelé *pas de quantification* et est noté  $\Delta_i$ . Dans le cas de la figure 1.16, tous les intervalles ont même longueur  $\Delta$ ; la quantification est alors dite *linéaire*. Le pas de quantification peut alors s’exprimer par

$$\boxed{\Delta = \frac{g_{\max} - g_{\min}}{L}}$$

Pour une dynamique du signal donnée, on remarque que l’approximation est d’autant meilleure que le pas de quantification  $\Delta$  est petit, et donc que le nombre  $L$  de niveaux de quantification est élevé. L’erreur de quantification est alors égale à

$$\boxed{e_q = \frac{\Delta}{2}}$$

### 1.6.3 Codage

À ce stade, nous avons déjà étudié les deux premières étapes du processus de numérisation d'un signal analogique  $g(t)$ . Ces deux étapes peuvent se représenter de la manière suivante :

$$g(t) \longrightarrow \{g(kT_s)\} \longrightarrow \{q_k\}$$

Nous disposons d'un signal numérique  $\{q_k\}$  dont les valeurs possibles se trouvent dans l'ensemble fini  $\{v_0, v_1, \dots, v_{L-1}\}$  où  $L$  est le nombre de niveaux de quantification. Reste maintenant la dernière étape, obtenir obtenir la séquence binaire si chère aux informaticiens...

La méthode consiste simplement à remplacer, dans la séquence  $\{q_k\}$ , chaque occurrence de  $v_i$  par une séquence de bits données. Mais combien de bits par valeur  $v_i$ ? Plusieurs possibilités existent. La plus simple est la suivante : il suffit de remplacer  $v_i$  par le nombre binaire correspondant à  $i$ . Pour  $L = 6$ , la table de codage est fournie à la table 1.1. Dès lors, chaque occurrence de la valeur  $v_0$  dans la séquence  $\{q_k\}$  sera remplacée par 000, chaque occurrence de la valeur  $v_1$  dans la séquence  $\{q_k\}$  sera remplacée par 001, et ainsi de suite...

$i$	Valeur quantifiée $v_i$	Code binaire $c_i$
0	$v_0$	000
1	$v_1$	001
2	$v_2$	010
3	$v_3$	011
4	$v_4$	100
5	$v_5$	101

TABLE 1.1 – Exemple de table de codage pour  $L = 6$  niveaux de quantification.

Quelques remarques s'imposent. Pour  $L = 6$ , les codes binaires 110 et 111 ne sont pas utilisés. Pour cette raison, on préfère généralement prendre une puissance de 2 pour le nombre de niveaux de quantification :

$$L = 2^R$$

où  $R$  est donc le nombre de bits utilisés pour coder un échantillon du signal analogique. Dans ce choix de codage, le nombre de bits utilisés par échantillon est identique pour tous les échantillons. Ceci n'est certainement pas optimal surtout si certaines valeurs sont plus fréquentes que d'autres. On pourrait dans ce cas utiliser un codage adapté aux données, comme par exemple le codage de HUFFMAN, qui fournit des codes de longueurs variables pour chaque échantillon (les valeurs les plus fréquentes sont codées sur moins de bits que les valeurs les moins fréquentes).

Mais revenons au cas simple du codage à longueur fixe. Rien oblige de choisir le codage fourni à la table 1.1. On pourrait très bien intervertir les codes utilisés pour les valeurs  $v_i$ . Il suffirait alors simplement de connaître la table de codage, lors du décodage, pour pouvoir retrouver les échantillons.

### 1.6.4 Synthétisons...

Nous avons à présent étudié les 3 étapes nécessaires à la numérisation d'un signal analogique  $g(t)$ . Ceci peut se représenter ainsi :

$$\boxed{g(t) \longrightarrow \{g(kT_s)\} \longrightarrow \{q_k\} \longrightarrow \{b_n\}}$$

où  $\{b_n\}$  est un signal numérique binaire constitué de la concaténation des codes correspondant à chaque échantillon  $q_k$ . Notons que l'indice n'est plus  $k$  (indice de chaque échantillon du signal analogique) mais maintenant  $n$  (indice de chaque bit dans la séquence binaire). En effet, plusieurs bits sont utilisés pour représenter un échantillon du signal de départ.

Rappelons à présent les différents paramètres importants du processus de numérisation :

- $f_s$  [echantillon/s] : fréquence d'échantillonnage qui doit être choisie, en respect du théorème de SHANNON, strictement supérieure au double de la plus haute fréquence de  $g(t)$  ;
- $R$  [bit/echantillon] : nombre de bits utilisés pour représenter un échantillon. La valeur de  $R$  est directement liée au nombre de niveaux  $L$  de quantification par la relation

$$L = 2^R$$

tandis que l'erreur de quantification est donnée par

$$e_q = \frac{\Delta}{2} = \frac{g_{max} - g_{min}}{2L}$$

Dès lors, la qualité de la numérisation est directement liée au nombre de bits pour représenter un échantillon, et donc à la quantité de bits nécessaires pour représenter un signal. Soit  $Q_b$  la quantité de bits nécessaires pour coder *une seconde* de signal.  $Q_b$  est alors simplement donné par

$$Q_b \text{ [bit/seconde]} = f_s \times R$$

Dès lors, si l'on souhaite réaliser la transmission en temps réel d'un signal numérisé (comme dans le cas du téléphone), il est nécessaire que le débit binaire  $R_b$  [bit/seconde] du canal de transmission soit supérieur au débit  $Q_b$  de l'information à transmettre. Si le temps réel est une nécessité et que le débit  $R_b$  du canal est fixé, il faudra adapter les paramètres de la numérisation pour que  $Q_b$  soit inférieur à  $R_b$ .

#### Exemple de la téléphonie numérique

En téléphonie numérique, l'important n'est pas la qualité irréprochable du signal mais bien l'aspect temps réel ; des coupures pendant une conversion, en attendant que des buffers se remplissent, ne seraient pas acceptables.

Bien qu'un signal audio s'étende de 0 à 20 [kHz], si l'on se limite à une bande fréquence de 0 à 4 [kHz], le signal filtré reste tout à fait compréhensible et tout à fait suffisant pour une communication téléphonique. Ainsi, le signal analogique téléphonique est pré-filtré par un filtre passe-bas de bande passante égale à 4 [kHz]. On peut donc choisir une fréquence d'échantillonnage égale à  $f_s = 8$  [kHz] = 8.000 [echantillon/s], tout en respectant le théorème de SHANNON. Un choix de  $R = 8$  [bit/echantillon] a été fait, ce qui conduit donc à

$$Q_b = 8.000 \text{ [echantillon/s]} \times 8 \text{ [bit/echantillon]} = 64.000 \text{ [bit/s]}$$

On comprend donc beaucoup mieux comment les opérateurs téléphoniques actuels peuvent proposer la téléphonie numérique en même temps que l'ADSL. Celui-ci propose en effet un débit actuel de l'ordre de 4 [Mbit/s] (en Belgique...), de quoi transmettre un nombre important de signaux téléphoniques en même temps si on voulait...

### Exemple du CD-Audio

Dans le cas de l'antique CD-Audio, les contraintes sont différentes. Le but n'est pas la transmission de l'information mais bien la qualité du son obtenu. En particulier ici, le stéréo est considéré. Donc, deux signaux sont numérisés, un pour la canal gauche et un pour le canal droit.

Ici, on garde toute la bande de fréquence audio, c'est-à-dire de 0 à 20 [kHz]. Une fréquence d'échantillonnage de 44,1 [kHz] a été choisie, en respect du théorème de SHANNON. Pour assurer une bonne qualité de la numérisation,  $R = 16$  [bit/echantillon] a été choisi. Par canal, il nous faut donc

$$Q_b(1 \text{ canal}) = 44.100 \text{ [echantillon/s]} \times 16 \text{ [bit/echantillon]} = 705.600 \text{ [bit/s]}$$

et donc

$$Q_b = 1.411.200 \text{ [bit/s]} \approx 0,168 \text{ [Mbyte/s]}$$

pour le signal stéréo. Etant donné qu'un CD-Audio présente une capacité de 700 [Mbyte], on peut espérer stocker une durée d'enregistrement égale à

$$\frac{700}{0,168} = 4161 \text{ [s]} \approx 70 \text{ [minutes]}$$

## 1.7 Transformée de FOURIER discrète

Jusqu'à présent, nous avons beaucoup parlé de transformée de FOURIER, mais toujours sous la forme de la définition (1.8). Cette définition s'appliquait uniquement à des signaux analogiques. Nous allons à présent voir qu'il existe une version de la transformée de FOURIER directement applicable à des signaux numériques, et donc directement utilisable dans des programmes informatiques.

### 1.7.1 Discréétisation de la transformée de FOURIER

Considérons un signal analogique  $g(t)$  à énergie finie et sa transformée de FOURIER

$$G(f) = \int_{-\infty}^{+\infty} g(t) e^{-j2\pi t f} dt \quad (1.24)$$

Afin de faire calculer cette intégrale par un programme, il est nécessaire de la discréétiser et d'utiliser un algorithme d'intégration fourni par les théories de l'analyse numérique. Tout d'abord, nous devons nous limiter à un intervalle de temps fini  $[t_0, t_0 + T[$  (donc de longueur  $T$  et commençant à l'instant  $t_0$ ), et discréétiser cet intervalle :

$$t = t_0 + n \Delta t \quad n = 0, 1, \dots, N - 1 \quad (1.25)$$

où

$$\Delta t = \frac{T}{N} \quad (1.26)$$

est le pas de discréétisation ou d'intégration numérique. D'un point de vue signal, tout se passe comme si on échantillonnait le signal  $g(t)$  avec une période d'échantillonnage égale à  $\Delta t$ , et donc une fréquence d'échantillonnage  $f_s = 1/\Delta t$ . Nous formons donc le signal numérique suivant

$$\{x_n\} = \{g(t_0 + n\Delta t)\} = \{g(t_0), g(t_0 + \Delta t), \dots, g(t_0 + (N - 1)\Delta t)\}$$

pour lequel nous nous limitons à un nombre fini  $N$  d'échantillons.

Nous pouvons à présent appliquer la bien connue règle du Trapèze pour calculer l'intégrale (1.24) :

$$\begin{aligned} G(f) &= \sum_{n=0}^{N-1} g(t_0 + n \Delta t) e^{-j2\pi(t_0+n\Delta t)f} \Delta t \\ &= \Delta t e^{-j2\pi t_0 f} \sum_{n=0}^{N-1} g(t_0 + n \Delta t) e^{-j2\pi n \Delta t f} \end{aligned}$$

Si les échantillons du signal  $g(t_0 + n\Delta t)$  sont stockés dans un vecteur de `float`, il est maintenant aisément d'écrire un programme permettant de calculer  $G(f)$  pour n'importe quelle valeur de la fréquence  $f$ . Oui, mais quelles valeurs de  $f$  allons-nous choisir ? et combien de valeurs de  $f$  différentes allons-nous considérer ? Tout d'abord, il est d'usage de calculer  $G(f)$  pour autant de valeurs de  $f$  qu'il y a d'échantillons du signal de départ. Donc, nous calculerons  $N$  "échantillons" de la transformée de FOURIER. D'un autre côté, nous savons que le module de  $G(f)$  est symétrique par rapport à l'origine tandis que sa phase est anti-symétrique par rapport à l'origine. Toute l'information obtenue pour des fréquences positives peut donc être utilisées pour retrouver l'information relatives aux fréquences négatives. Nous nous limiterons donc aux fréquences positives. En conclusions, nous allons calculer  $G(f)$  pour les valeurs de  $f$  suivantes :

$$f = m \Delta f \quad m = 0, \dots, N - 1$$

où  $\Delta f$  est le *pas fréquentiel* qu'il nous faudra fixer. Nous pouvons donc écrire

$$G(m \Delta f) = \Delta t e^{-j2\pi t_0 m \Delta f} \sum_{n=0}^{N-1} g(t_0 + n \Delta t) e^{-j2\pi n m \Delta t \Delta f} \quad m = 0, \dots, N - 1 \quad (1.27)$$

Grâce à cette dernière formule, il est possible de calculer  $N$  échantillons de la transformée de FOURIER en programmant une double boucle (une sur  $m$  et la seconde sur  $n$ ). Néanmoins, je vous souhaite bien du courage... Nous allons voir qu'il est possible de nous simplifier la vie.

Tout d'abord, sans nous en rendre vraiment compte, nous venons de former un nouveau signal numérique à valeurs complexes :

$$\{G(m \Delta f)\}$$

dont le nombre d'échantillons est limité à  $N$ . Ensuite, nous n'avons pas encore fixé le pas fréquentiel  $\Delta f$ . Si nous le prenons égal à

$$\Delta f = \frac{1}{N \Delta t} \quad (1.28)$$

l'expression (1.27) peut s'écrire plus simplement :

$$G(m \Delta f) = \Delta t e^{-j2\pi t_0 m \Delta f} \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{nm}{N}} \quad m = 0, \dots, N - 1$$

Courage, on arrive au bout ! Cette dernière expression peut finalement s'écrire, en introduisant la séquence  $\{X_m\}$ ,

$$\begin{cases} G(m \Delta f) = \Delta t e^{-j2\pi t_0 m \Delta f} X_m \\ X_m = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{nm}{N}} \end{cases}$$

avec  $m = 0, \dots, N-1$ . Mais pourquoi avoir fait apparaître deux expressions horribles alors qu'au départ, on n'en avait qu'une seule ? Tout simplement, parce qu'on a fait apparaître l'expression importante

$$X_m = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{nm}{N}} \quad m = 0, \dots, N-1$$

qui définit la *transformée de FOURIER discrète* (DFT en anglais pour *Discrete Fourier Transform*) de la séquence  $\{x_n\}$ . Qu'a-t-elle de si important cette nouvelle transformée ? Et bien d'avoir un nombre important d'implémentations déjà réalisées dans différents langages de programmation ! Donc, inutile de la reprogrammer ! Des logiciels comme Matlab ou Labview vous en offrent plusieurs implémentations que vous aurez (ou avez déjà eu) l'occasion de découvrir.

### 1.7.2 DFT - IDFT

Considérons un signal numérique réel  $\{x_n\}$  où  $n = 0, \dots, N-1$ .

**Définition [Transformée de FOURIER discrète - DFT].** La transformée de FOURIER discrète (DFT) de la séquence  $\{x_n\}$  est la séquence  $\{X_m\}$  définie par

$$X_m = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{nm}{N}} \quad m = 0, \dots, N-1 \quad (1.29)$$

La séquence  $\{x_n\}$  peut être retrouvée de manière exacte par transformée de FOURIER discrète inverse (IDFT) :

$$x_n = \frac{1}{N} \sum_{m=0}^{N-1} X_m e^{j2\pi \frac{nm}{N}} \quad n = 0, \dots, N-1 \quad (1.30)$$

La DFT présente des propriétés de périodicité que nous allons étudier et qui sont importantes pour bien comprendre ce qui se passe au niveau de l'implémentation.

#### Propriété 1

Tout d'abord, nous pouvons observer que

$$\boxed{X_{N-i} = X_{-i}} \quad (1.31)$$

En effet,

$$\begin{aligned} X_{N-i} &= \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{n(N-i)}{N}} \\ &= \sum_{n=0}^{N-1} x_n e^{-j2\pi n} e^{-j2\pi \frac{n(-i)}{N}} \\ &= X_{-i} \end{aligned}$$

Si on revient un instant à notre calcul de  $G(f)$  et que nous considérons que  $t_0 = 0$  (ce qui est le cas le plus fréquent), cette propriété conduit à la relation suivante :

$$G((N-i)\Delta f) = G(-i\Delta f)$$

ce qui montre que certains échantillons de  $G(f)$  que nous avons calculés pour des fréquences positives correspondent aussi à des échantillons de  $G(f)$  pour des fréquences négatives ! En

particulier, pour  $i = N/2$ ,

$$G\left(\frac{N}{2}\Delta f\right) = G\left(-\frac{N}{2}\Delta f\right)$$

### Propriété 2

Ensuite, nous pouvons observer que

$$\boxed{X_{aN+i} = X_i} \quad (1.32)$$

où  $a$  est un entier. En effet,

$$\begin{aligned} X_{aN+i} &= \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{n(aN+i)}{N}} \\ &= \sum_{n=0}^{N-1} x_n e^{-j2\pi na} e^{-j2\pi \frac{ni}{N}} \\ &= X_i \end{aligned}$$

À nouveau, si on revient au calcul de  $G(f)$  et que l'on considère que  $t_0 = 0$ , cette propriété conduit à

$$G((aN + i)\Delta f) = G(i\Delta f)$$

ce qui montre que les échantillons de  $G(f)$  que nous avons calculés se répètent sur l'axe des fréquences tous les

$$N\Delta f = N \frac{1}{N\Delta t} = \frac{1}{\Delta t}$$

qui correspond à la fréquence d'échantillonnage  $f_s$  du signal  $g(t)$  ! Nous observons donc à nouveau les conséquences du théorème de SHANNON ! La figure 1.17 illustre la situation. Les échantillons  $G(m\Delta f)$  pour  $m = 0, \dots, N - 1$  que nous avons calculés correspondent aux fréquences situées entre 0 et  $1/\Delta t$ , c'est-à-dire à la moitié de la réplique de  $G(f)$  à l'origine et à la moitié de la réplique de  $G(f)$  en  $f_s = 1/\Delta t$ .

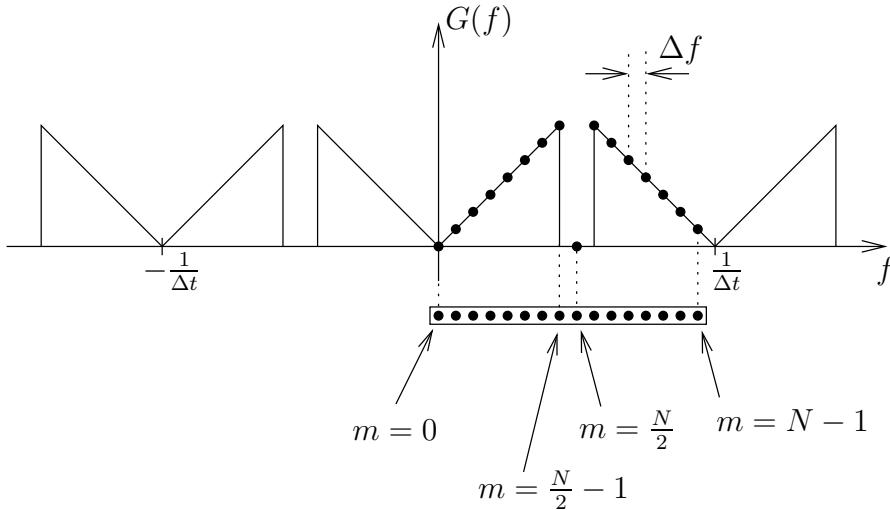


FIGURE 1.17 – Interprétation de la DFT en regard du théorème de SHANNON.

Afin d'afficher le spectre calculé de manière visuellement satisfaisante, il serait intéressant d'avoir la fréquence  $f = 0$  au milieu de la séquence  $\{X_m\}$ . En général, on permute donc les deux moitiés du vecteur contenant les échantillons de  $G(f)$ . Ceci est illustré à la figure 1.18.

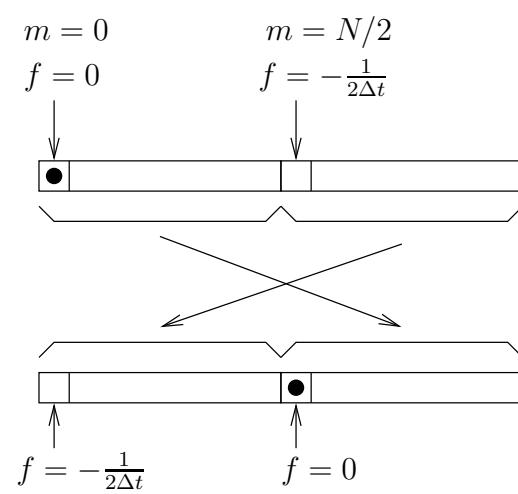


FIGURE 1.18 – Permutation des deux moitiés du vecteur contenant la DFT.



# Chapitre 2

## Transmission du signal

Dans ce chapitre, nous allons aborder différentes techniques permettant de transmettre un signal.

### 2.1 Introduction

Par transmission d'un signal, on entend son "transport" d'un point physique à un autre. Les exemples sont variés : communications réseaux entre ordinateurs, satellites, wifi, radio, ... À chaque fois, une liaison physique doit relier les deux extrémités de la communication : l'émetteur et le récepteur. Dans les communications entre ordinateurs, le support physique utilisé est principalement le câble (du type Ethernet), sur lequel on fait passer des signaux électriques, bien que le wifi prenne une place de plus en plus considérable. Pour ce dernier, ainsi que pour toutes les communications radio, il n'y a pas de support physique. L'information circule sous la forme d'une onde électromagnétique émise par une antenne alimentée par un signal électrique. Dans le cas des fibres optiques, il s'agit également d'utiliser une onde électromagnétique travaillant dans les longueurs d'onde associées à la lumière.

Dans tous les cas, la transmission de l'information se réalise en utilisant un signal analogique qui doit être adapté au support de transmission. L'espace au travers duquel transite le signal au cours de sa transmission est appelé **canal** de communication et l'opération qui consiste à adapter le signal à transmettre au canal s'appelle le **codage de canal**.

#### 2.1.1 La modulation

Considérons un câble (Ethernet ou paire torsadée, cela n'a aucune importance ici) ainsi qu'un signal  $g(t)$ , de type passe-bas comme un signal vocal par exemple, que l'on désire transmettre en utilisant ce câble. La première idée qui vient à l'esprit est d'appliquer le signal électrique  $g(t)$  aux bornes du câble. Et cela n'est pas une mauvaise idée... À la sortie, on retrouvera le signal  $g(t)$ , après un certain délai de transmission, et ayant éventuellement subi une atténuation d'amplitude. Mais hélas, ce n'est pas tout... La physique montre qu'un câble électrique présente les caractéristiques d'un filtre passe-bas et possède donc une bande passante limitée. La valeur de cette bande passante dépend essentiellement du type de câble.

Dès lors, afin d'assurer que le signal  $g(t)$  arrive à bon port sans trop de dégradation, il est nécessaire que sa propre bande passante  $W_g$  soit inférieure à celle du câble  $W_{canal}$ , comme l'illustre la figure 2.1. Dans cette situation, le signal est bien adapté au canal (sans rien faire), et on parle de **transmission en bande de base** car la bande de fréquence utilisée pour la transmission commence à la base de l'axe de fréquences. Tout a l'air bien mais n'y a-t-il pas un

gaspillage de bande passante au niveau du câble ?

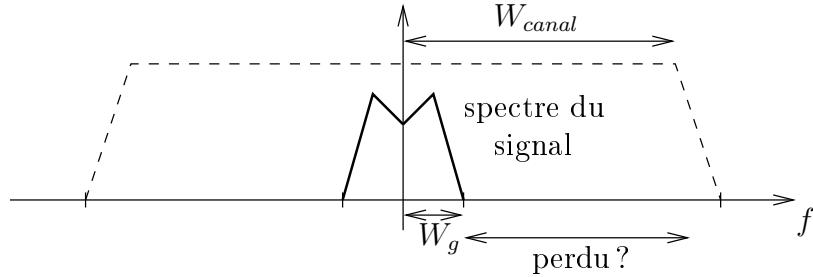


FIGURE 2.1 – Interprétation fréquentielle de la transmission d'un signal sur un câble.

Considérons à présent un second signal  $g_2(t)$ , de même caractéristiques fréquentielles que  $g(t)$ , que l'on désire transmettre sur ce même câble. Impossible si on est déjà en train de transmettre  $g(t)$ ... Si on tente de transmettre  $g(t)$  et  $g_2(t)$  tels quels en même temps sur le câble, leur spectres se chevaucheraient et tout ce que l'on obtiendrait à la sortie serait leur somme... Une première solution pourrait être de transmettre un puis l'autre. C'est ce que l'on appelle un **multiplexage temporel**. Il y a cependant un problème si l'on désire faire de la transmission stéréo en temps réel... Il faudrait dans ce cas trouver un moyen de transmettre ces deux signaux simultanément !

La solution serait alors d'utiliser la bande de fréquence non encore utilisée sur le câble afin d'y placer le spectre de  $g_2(t)$  sans chevauchement avec celui de  $g(t)$ . Mais pour cela, il faudrait être capable de "translater" le spectre de  $g_2(t)$  le long de l'axe de fréquence. Cette opération est possible et porte le nom de **modulation analogique**. Les transmissions de  $g(t)$  et de  $g_2(t)$  se font donc de manière simultanée mais dans des bandes de fréquences différentes. C'est ce que l'on appelle un **multiplexage fréquentiel**. Ceci est illustré à la figure 2.2. Après réception du signal, l'opération qui consiste à ramener le signal  $g_2$  dans sa bande de base est appelée **démodulation**.

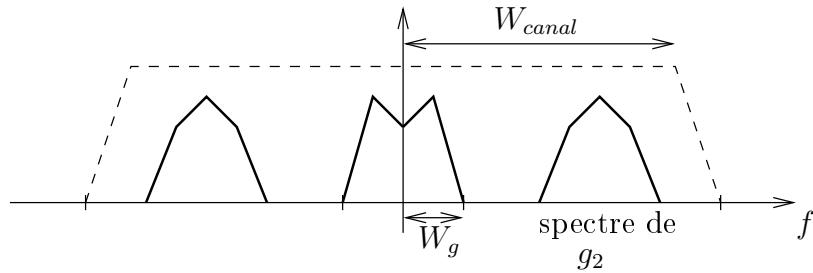


FIGURE 2.2 – Multiplexage fréquentiel et modulation.

Bien que nous ne sachions pas encore grand chose de la modulation, nous voyons déjà un de ses avantages : elle permet la transmission simultanée de plusieurs signaux sur le même canal. Néanmoins, dans certains cas, elle est obligatoire même si l'on n'a qu'un seul signal à transmettre. C'est le cas de toutes les transmissions radios. Considérons par exemple le même signal  $g(t)$  que l'on désire transmettre à présent à l'aide d'onde radio générée par une antenne. La première idée qui vient à l'esprit est à nouveau de connecter le signal électrique  $g(t)$  aux bornes de l'antenne d'émission (voir même en l'amplifiant). Le résultat est malheureusement

ici décevant. Au niveau de l'antenne de réception, rien (d'utilisable) n'a pu être capté... Cette situation est illustrée à la figure 2.3.

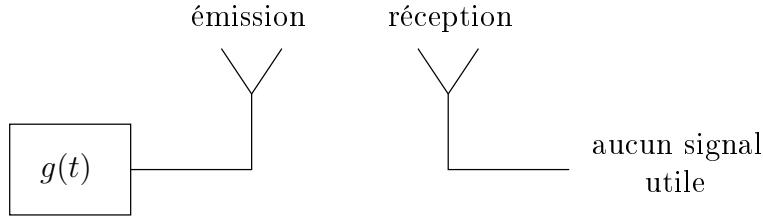


FIGURE 2.3 – Transmission radio d'un signal analogique en bande de base.

La physique (notamment les équations de MAXWELL de l'électromagnétisme) nous apprend que les ondes électromagnétiques radio travaillent dans le domaine des hautes fréquence allant jusqu'à 3.000 [GHz] et qu'un antenne, pour émettre de manière adéquate, doit avoir une taille  $l$  de l'ordre de grandeur de la longueur d'onde  $\lambda$  utilisée :

$$l \sim \lambda$$

Dès lors, qu'en est-il pour notre signal audio  $g(t)$ ? Si sa bande passante est de  $W_g = 20$  [kHz] (fréquence maximale audible), il nous faudrait, pour le transmettre tel quel, une antenne de taille minimale de l'ordre de

$$l \sim \lambda = \frac{c}{f} = \frac{300.000 \text{ [km/s]}}{20.000 \text{ [Hz]}} = 15 \text{ [km]}$$

ce qui n'est pas envisageable! La solution consiste alors à adapter le signal au canal en translatant son spectre sur l'axe des fréquences afin qu'une antenne de taille raisonnable puisse être utilisée. La modulation est donc ici une nécessité pour la transmission du signal ! La figure 2.4 illustre le schéma résultant.

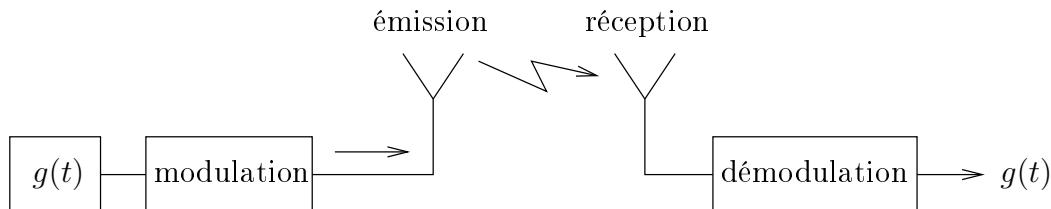


FIGURE 2.4 – Transmission radio d'un signal analogique.

En résumé, nous pouvons dire que la modulation est une opération extrêmement importante en transmission du signal. Elle permet de

- transmettre simultanément plusieurs signaux analogiques dans des bandes de fréquences différentes, en réalisant ainsi un multiplexage fréquentiel,
- adapter le signal à transmettre au canal de communication utilisé.

### 2.1.2 Schéma complet d'une transmission analogique

Le schéma global d'une communication analogique est présenté à la figure 2.5. Sur ce schéma, on voit bien apparaître les opérations de modulation et de démodulation dont nous avons déjà parlé. Mais ce n'est pas tout. En général, un pré-filtrage est réalisé sur le signal  $g(t)$  à transmettre afin d'être sûr de la bande de fréquence réellement utilisée par le signal. On obtient ainsi le signal  $m(t)$ , appelé signal utile ou plutôt ***signal modulant***. Avant d'être transmis, celui-ci doit être adapté au canal via l'opération de modulation. Le signal  $s(t)$  ainsi obtenu est appelé ***signal modulé***.

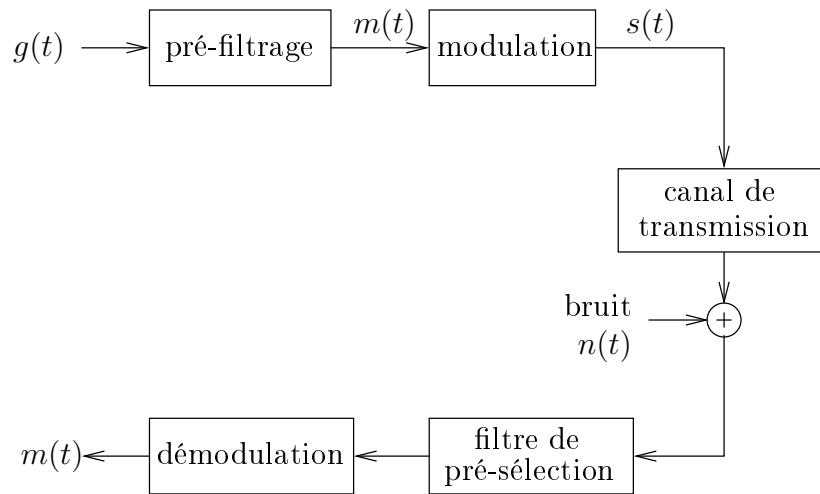


FIGURE 2.5 – Schéma global d'une transmission analogique.

Le signal modulé transite alors sur le canal de communication. Au cours de la transmission, un ou des signaux parasites (bruit thermique des circuits électroniques utilisés, signaux issus d'autres communications, ...), appelé ***bruit*** et noté  $n(t)$  peuvent s'ajouter au signal modulé. Ce bruit est en général à large bande et avant de réaliser l'opération de démodulation, il est nécessaire de réaliser un filtrage de pré-sélection afin de supprimer le bruit en dehors de la bande de fréquence du signal modulé. Ce filtre est également nécessaire dans le cas d'un multiplexage fréquentiel et cela afin de sélectionner le signal que l'on désire démoduler dans le signal multiplexé. Vient finalement l'opération de démodulation qui vise à récupérer le signal modulant  $m(t)$ .

La suite de ce chapitre est consacrée à l'étude de différentes techniques de modulation.

## 2.2 Généralités

Dans la suite de cet exposé, nous ferons deux hypothèses sur le signal modulant  $m(t)$  :

1. il est ***borné*** et ***normalisé*** de telle sorte que

$$-1 \leq m(t) \leq +1$$

2. il est ***à spectre limité*** à la bande de fréquences  $[-W, +W]$ , c'est-à-dire

$$M(f) = 0 \quad \text{si } |f| > W$$

On appelle alors ***bande de base*** la bande de fréquences  $[0, W]$ .

### 2.2.1 La porteuse

Le principe de la modulation est de modifier une sinusoïde de référence en fonction du signal modulant. Cette sinusoïde de référence sera notée  $c(t)$  et appelée **porteuse** (non modulée) :

$$c(t) = A_c \cos(2\pi f_c t + \Phi_c) \quad (2.1)$$

où

- $A_c$  est l'amplitude de la porteuse,
- $f_c$  est la fréquence de la porteuse, et
- $\Phi_c$  est la phase de la porteuse.

### 2.2.2 Le signal modulé

Pour incorporer le signal  $m(t)$ , on peut agir sur un de ces paramètres. D'une manière générale, la modulation consiste à remplacer une de ces caractéristiques par une fonction linéaire de  $m(t)$ . Le signal résultant est appelé **signal modulé** et est noté  $s(t)$ . Il prend la forme générale suivante :

$$s(t) = A(t) \cos \Phi_i(t)$$

où

- $A(t)$  est l'amplitude instantanée du signal modulé, et
- $\Phi_i(t)$  est l'angle (on parle aussi de phase) instantané du signal modulé.

On peut alors distinguer deux catégories de modulation analogique :

1. Les **modulations d'amplitude** : dans le cas où le signal modulant  $m(t)$  est incorporé à l'amplitude et non à l'angle :

$$s(t) = A(t) \cos(2\pi f_c t + \Phi_c) \quad (2.2)$$

2. Les **modulations angulaires** : dans le cas où le signal modulant  $m(t)$  est incorporé à l'angle et non à l'amplitude :

$$s(t) = A_c \cos \Phi_i(t) \quad (2.3)$$

Un paramètre important est la **fréquence instantanée**  $f_i(t)$  du signal modulé :

$$f_i(t) = \frac{1}{2\pi} \frac{d\Phi_i}{dt}(t) \quad (2.4)$$

Elle représente la fréquence du signal modulé à un instant  $t$  bien précis. Dans le cas de la porteuse non modulée, la fréquence instantanée est donnée par

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt}(2\pi f_c t + \Phi_c) = f_c$$

Elle ne varie donc pas au cours du temps. Il n'en sera plus de même dans le cas d'une modulation angulaire.

## 2.3 Modulation d'amplitude

Nous allons aborder ici quelques modulations d'amplitude classiques. Pour chacune d'elles, nous donnerons l'expression du signal modulé, son spectre, ainsi que la technique permettant de démoduler le signal.

### 2.3.1 Modulation AM

Il s'agit de la modulation historiquement la plus ancienne dans le domaine des télécommunications. En raison de sa faible efficacité, elle fut progressivement remplacée par d'autres techniques. Le terme AM vient simplement de l'anglais "Amplitude Modulation".

Pour cette modulation, le signal modulé  $s(t)$  est défini par

$$s(t) = A_c (1 + k_a m(t)) \cos(2\pi f_c t) \quad (2.5)$$

où

- $k_a$  est un paramètre de la modulation appelé *coefficient (ou taux) de modulation*. Il est toujours compris dans l'intervalle  $[0, 1]$
- $A(t) = A_c (1 + k_a m(t))$  est l'amplitude instantanée du signal modulé. On remarque que  $A(t)$  est bien une fonction linéaire du signal modulant  $m(t)$ .

La figure 2.6 illustre la construction du signal modulé. Dans la situation normale d'une modulation d'amplitude, la fréquence porteuse est beaucoup plus grande que la fréquence maximale de la bande de base  $W$  : la modulation correspond donc à une variation lente de l'amplitude instantanée.

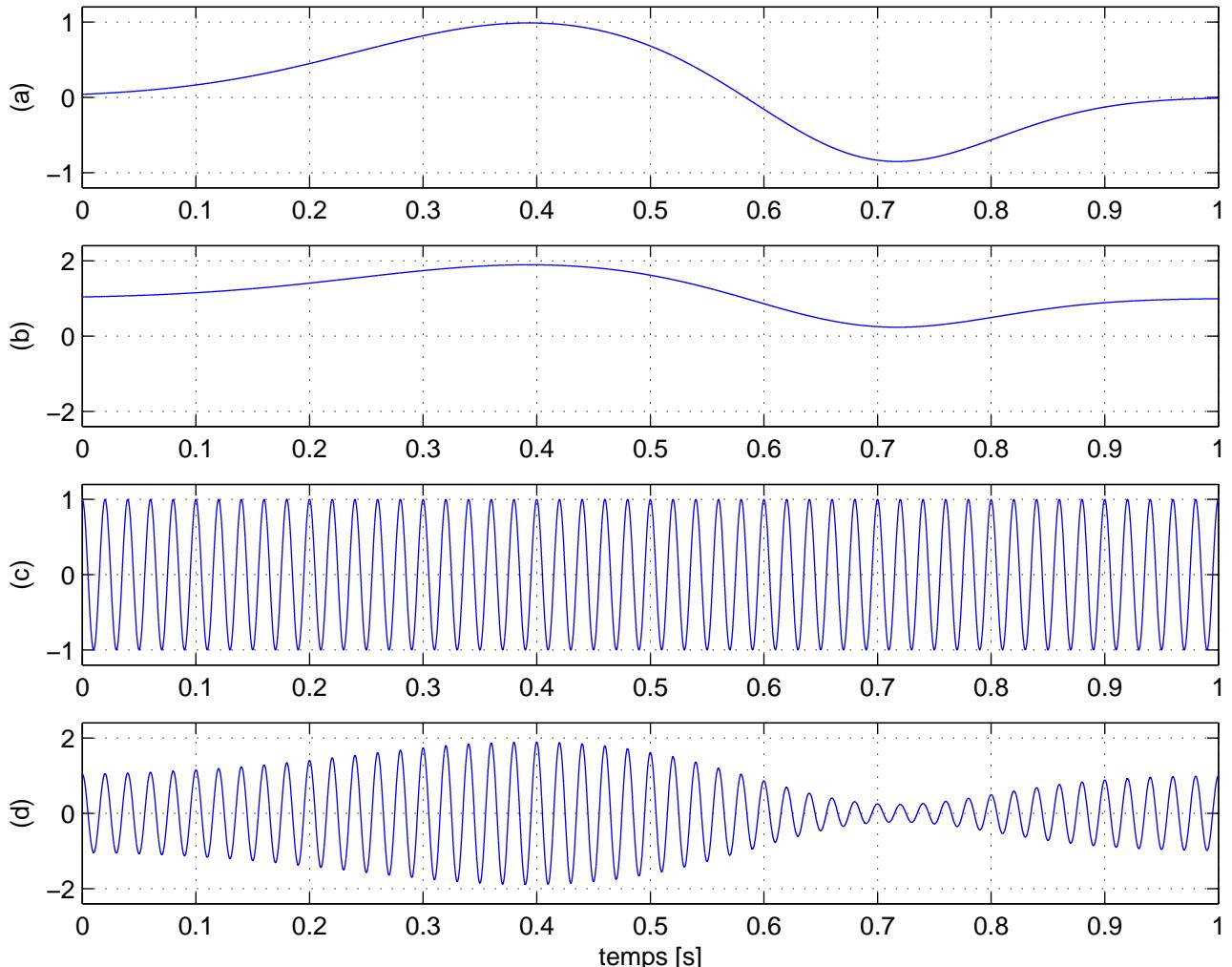


FIGURE 2.6 – Exemple de modulation AM pour  $A_c = 1$ ,  $k_a = 0,9$ ,  $f_c = 50 [Hz]$  : (a) Signal modulant  $m(t)$ . (b) Amplitude instantanée  $A(t)$ . (c) Porteuse  $c(t)$ . (d) Signal modulé  $s(t)$ .

### Spectre et bande passante

Afin d'analyser le comportement fréquentiel du signal modulé, nous allons calculer sa transformée de FOURIER. Tout d'abord, réécrivons-le sous la forme

$$s(t) = A_c \cos(2\pi f_c t) + k_a A_c m(t) \cos(2\pi f_c t)$$

Le premier terme est une cosinusoïde dont nous connaissons déjà la transformée de FOURIER :

$$\cos(2\pi f_c t) \rightleftharpoons \frac{\delta(f - f_c) + \delta(f + f_c)}{2}$$

Son spectre est donc constitué de deux raies situées en  $f = \pm f_c$ . Le second terme est plus compliqué. Il s'agit du produit du signal modulant  $m(t)$  et de la cosinusoïde  $\cos(2\pi f_c t)$ . Sa transformée de FOURIER correspond donc au produit de convolution des transformées de FOURIER respectives de  $m(t)$  et de  $\cos(2\pi f_c t)$  :

$$\begin{aligned} S(f) &= \int_{-\infty}^{+\infty} M(f - \lambda) \frac{\delta(\lambda - f_c) + \delta(\lambda + f_c)}{2} d\lambda \\ &= \frac{M(f - f_c) + M(f + f_c)}{2} \end{aligned}$$

Nous en arrivons donc à la propriété importante suivante :

*Lorsque l'on multiplie (dans le domaine temporel) un signal  $m(t)$  par une cosinusoïde à la fréquence  $f_c$ , cela revient en fréquentiel à recopier le spectre  $M(f)$  de  $m(t)$  à la fréquence  $+f_c$  et à la fréquence  $-f_c$ , l'amplitude étant divisée par le facteur 2.*

Finalement, la transformée de FOURIER du signal modulé est donné par

$$S(f) = \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] + \frac{k_a A_c}{2} [M(f - f_c) + M(f + f_c)] \quad (2.6)$$

et est donc constituée de

- deux raies de DIRAC situées en  $f = \pm f_c$ , il s'agit en fait de la porteuse, et
- la copie du spectre  $M(f)$  du signal modulant en  $f = \pm f_c$ , il s'agit de l'information utile du signal modulé. Dans les fréquences positives, la partie du spectre située au-dessus de  $f_c$  est appelée **bande latérale supérieure** (BLS, ou USB pour “Upper Side Band”) tandis que la partie du spectre située en-dessous de  $f_c$  est appelée **bande latérale inférieure** (BLI, ou LSB pour “Lower Side Band” ).

La figure 2.7 fournit un exemple de modulation AM, avec les spectres de tous les signaux intervenants.

La bande passante  $W_{AM}$  du signal modulé correspond donc à la somme des longueurs des bandes latérales inférieures et supérieures, donc

$$W_{AM} = 2W$$

qui correspond au double de la bande de base (ou au double de la bande passante du signal modulant).

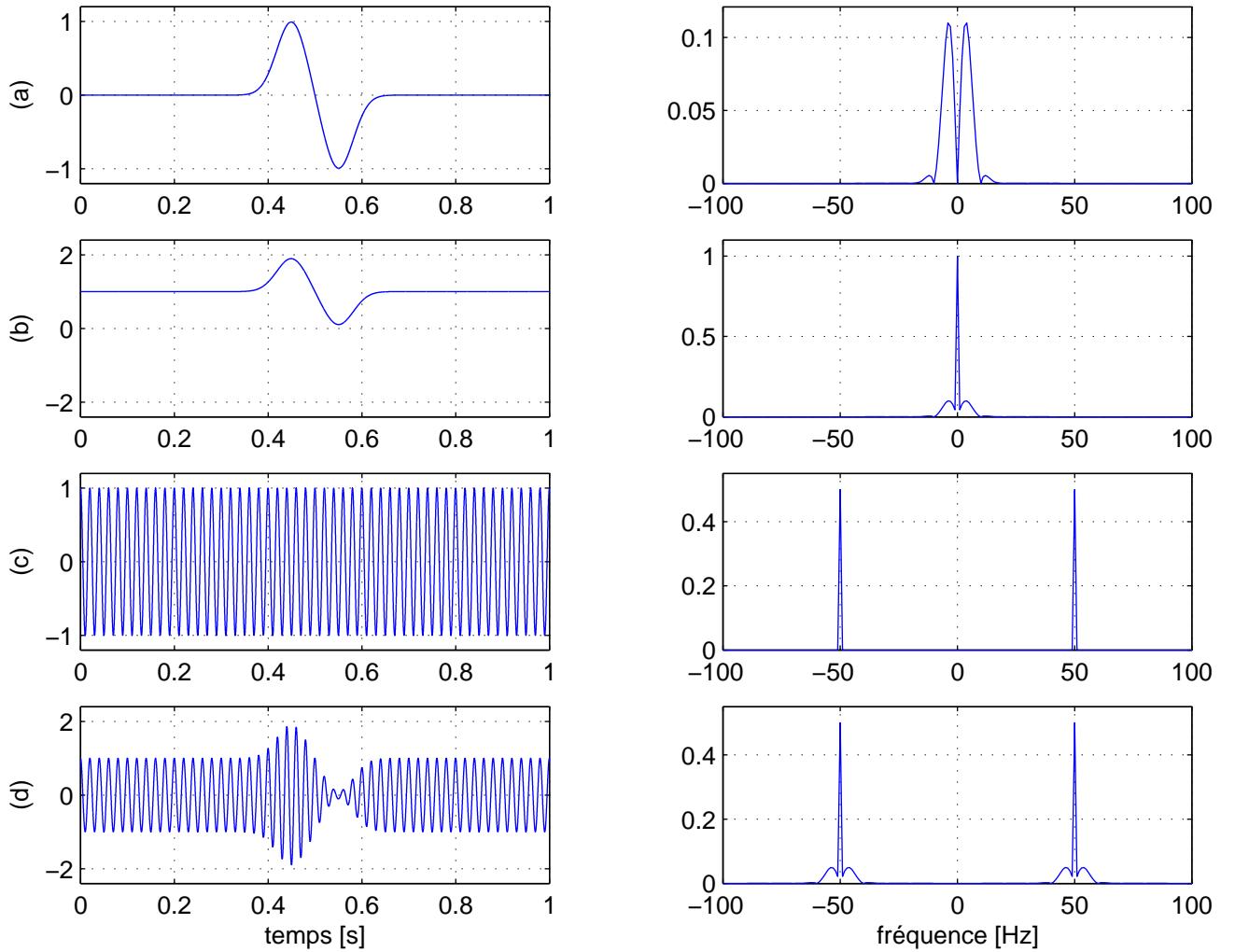


FIGURE 2.7 – Exemple de modulation AM pour  $A_c = 1$ ,  $k_a = 0.9$ ,  $f_c = 50 \text{ [Hz]}$  : (a) Signal modulant  $m(t)$ . (b) Amplitude instantanée  $A(t)$ . (c) Porteuse  $c(t)$ . (d) Signal modulé  $s(t)$ .

### Répartition de la puissance

Sur la figure 2.7d, on a l'impression que l'information utile du signal modulé (les deux bandes latérales) est très faible point de vue puissance par rapport aux deux raies de la porteuse. Il en est bien ainsi. On pourrait montrer que la plus grande partie de la puissance du signal modulé se situe dans la porteuse. La modulation AM se caractérise donc par un gaspillage de puissance et c'est la raison principale pour laquelle elle fut progressivement abandonnée au profit d'autres techniques de modulation.

Bon, si vous insistez, ne résistons pas à la tentation... Montrons cela pour un signal modulant sinusoïdal. Donc, soit le signal modulant suivant

$$m(t) = \cos(2\pi f_m t)$$

Le signal modulé a dans ce cas pour expression

$$\begin{aligned} s(t) &= A_c \cos(2\pi f_c t) + k_a A_c \cos(2\pi f_m t) \cos(2\pi f_c t) \\ &= A_c \cos(2\pi f_c t) + \frac{k_a A_c}{2} \cos(2\pi(f_c + f_m)t) + \frac{k_a A_c}{2} \cos(2\pi(f_c - f_m)t) \end{aligned}$$

Nous observons donc que

- le premier terme de cette dernière expression correspond à la porteuse  $c(t)$ , sa puissance est donnée par

$$P_c = \frac{A_c^2}{2}$$

- le second terme correspond à la bande latérale supérieure (BLS), sa puissance est donnée par

$$P_{BLS} = \frac{\left(\frac{k_a A_c}{2}\right)^2}{2} = \frac{k_a^2 A_c^2}{8} = \frac{k_a^2}{4} P_c$$

- le dernier terme correspond à la bande latérale inférieure (BLI), sa puissance est donnée par

$$P_{BLI} = \frac{\left(\frac{k_a A_c}{2}\right)^2}{2} = \frac{k_a^2 A_c^2}{8} = \frac{k_a^2}{4} P_c$$

Finalement, la puissance  $P_s$  du signal modulé est donc égale à

$$P_s = P_c + P_{BLI} + P_{BLS} = P_c \left(1 + \frac{k_a^2}{2}\right)$$

Le taux de modulation  $k_a$  a donc une très grande importance sur la répartition de la puissance émise. Si on considère un taux de modulation de 100% ( $k_a = 1$ ), la puissance totale du signal modulé est égale à

$$P_s = 1,5 P_c$$

et donc la puissance de la porteuse correspond à

$$P_c = \frac{2}{3} P_s$$

Cela signifie que la porteuse (qui ne contient aucune information utile) consomme 2/3 de la puissance totale d'émission !

### Démodulation par détecteur d'enveloppe

Le but est à présent de retrouver le signal modulant  $m(t)$  au départ du signal modulé  $s(t)$ .

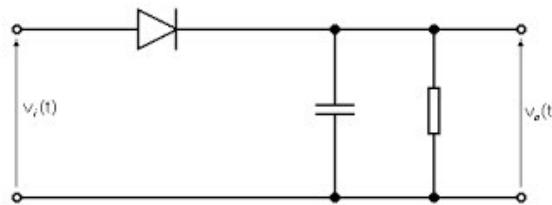


FIGURE 2.8 – Schéma du détecteur d'enveloppe.

Le démodulateur AM le plus courant et le plus simple est basé sur un circuit électronique appelé **détecteur d'enveloppe**. Sans entrer dans les détails (son circuit est donné à la figure 2.8 et son fonctionnement est représenté à la figure 2.9), celui-ci permet (à moindre frais car ce circuit est très simple) de retrouver l'amplitude instantanée  $A(t)$  du signal modulé, mais en **valeur absolue**, c'est-à-dire

$$|A(t)| = |1 + k_a m(t)|$$

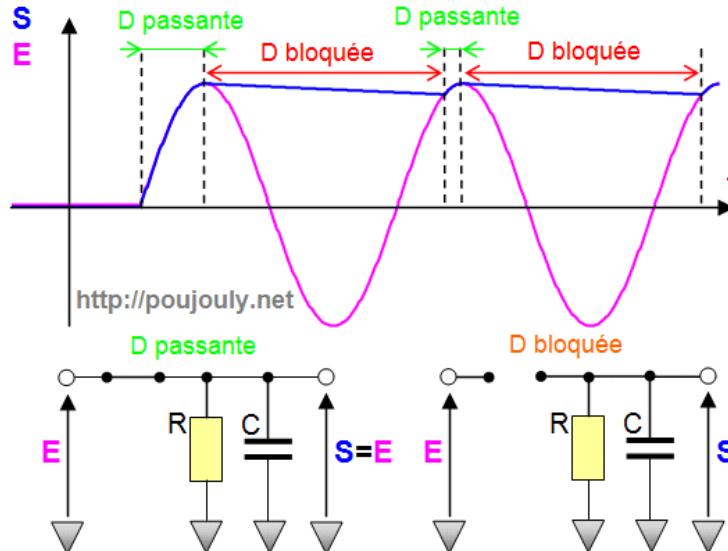


FIGURE 2.9 – Signaux et fonctionnement du détecteur d'enveloppe.

Lorsque le taux de modulation  $k_a$  est inférieure à 1 (et que  $m(t)$  vérifie bien l'hypothèse de départ  $-1 \leq m(t) \leq +1$ ) ,  $A(t)$  est toujours positif et la sortie du détecteur d'enveloppe est alors égale à

$$A(t) = 1 + k_a m(t)$$

Le signal modulant s'obtient alors simplement par

$$m(t) = \frac{A(t) - 1}{k_a}$$

Par contre, lorsque  $k_a$  est supérieur à 1, il est impossible de retrouver le signal  $m(t)$ . On dit qu'il y a alors **surmodulation**. Dans ce cas, il faudra alors utiliser la démodulation synchrone.

### Démodulation synchrone

Le principe de la démodulation synchrone est de

1. multiplier le signal modulé  $s(t)$  par une cosinusoïde de même fréquence et de même phase (d'où le terme **synchrone**) que la porteuse utilisée lors de la modulation,
2. filtrer le signal ainsi obtenu par un filtre passe-bas dont la bande passante correspond à la bande de base.

Le schéma correspondant à ce démodulateur est représenté à la figure 2.10. Étudions à présent son fonctionnement.

Après multiplication du signal modulé par la cosinusoïde, nous obtenons le signal

$$\begin{aligned} p(t) &= s(t) \cos(2\pi f_c t) \\ &= A_c(1 + k_a m(t)) \cos^2(2\pi f_c t) \\ &= \frac{A_c}{2}(1 + k_a m(t))(1 + \cos(4\pi f_c t)) \\ &= \frac{A_c}{2}(1 + k_a m(t)) + \frac{A_c}{2}(1 + k_a m(t)) \cos(4\pi f_c t) \end{aligned}$$

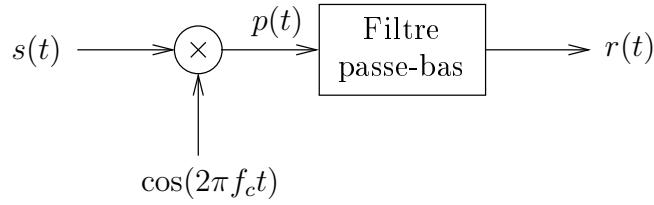


FIGURE 2.10 – Schéma de la démodulation synchrone.

Le premier terme de cette expression est un signal basse fréquence qui correspond à l'enveloppe du signal modulé  $s(t)$  à un facteur 1/2 près. Ce terme nous intéresse tout particulièrement car il est assez facile d'en sortir  $m(t)$ . Par contre, le second terme est un signal haute fréquence, qui correspond à la modulation AM de  $m(t)$  autour de la fréquence porteuse  $2f_c$ . Il constitue donc un terme parasite qu'il nous faut éliminer. Ceci est fait par l'application du filtre passe-bas qui nous fournit donc le signal suivant

$$r(t) = \frac{A_c}{2} (1 + k_a m(t))$$

L'interprétation fréquentielle de la démodulation synchrone est représentée schématiquement à la figure 2.11.

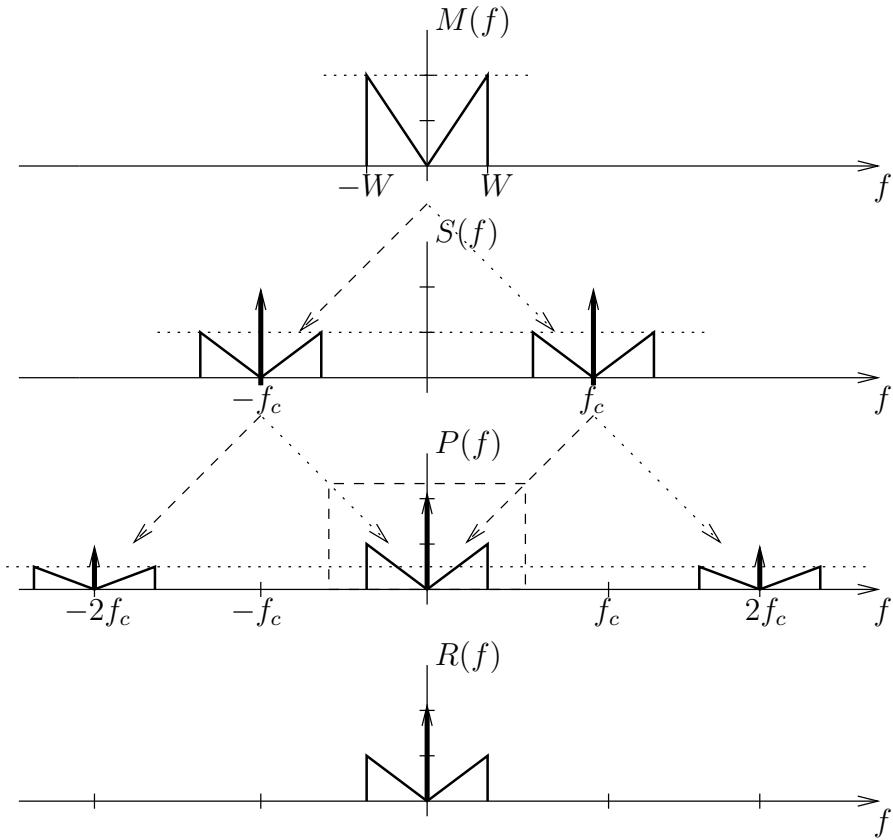


FIGURE 2.11 – Interprétation fréquentielle de la démodulation synchrone d'un signal AM.

Le signal modulant  $m(t)$  s'obtient alors facilement par

$$m(t) = \frac{\frac{2}{A_c} r(t) - 1}{k_a}$$

qui permet de supprimer la raie de DIRAC située en  $f = 0$ . En guise d'illustration, la figure 2.12 représente la démodulation du signal modulé de la figure 2.7.

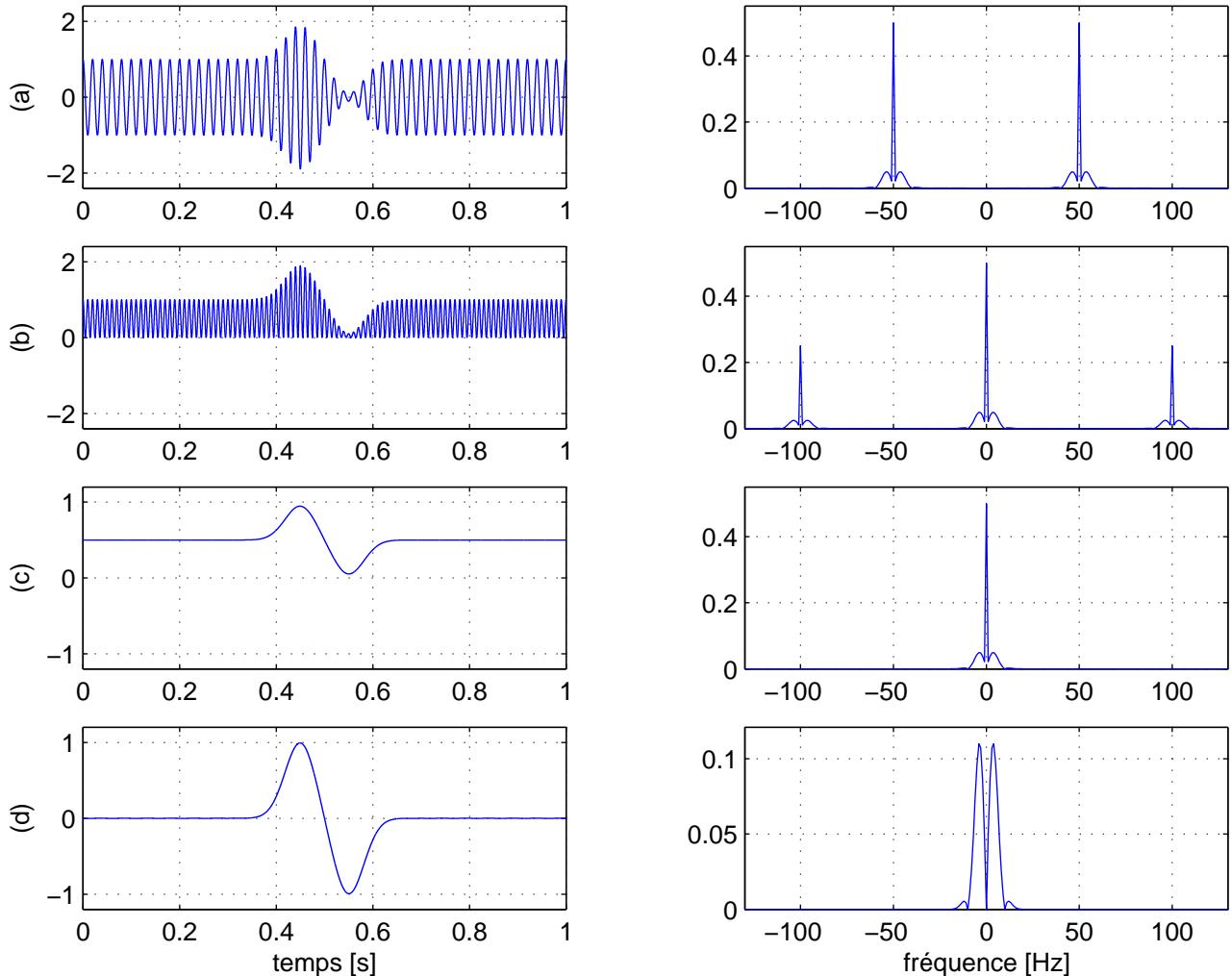


FIGURE 2.12 – Démodulation du signal AM ( $A_c = 1$ ,  $k_a = 0, 9$ ,  $f_c = 50$  [Hz] de la figure 2.7 : (a) Signal modulé  $s(t)$ . (b) Signal  $p(t)$ . (c) Signal  $r(t)$  obtenu par filtrage passe-bas du signal  $p(t)$ . (d) Signal  $m(t)$  retrouvé à partir de  $r(t)$  via la relation  $m(t) = (2r(t) - 1)/0, 9$ .

Avant de passer à la modulation suivante, une remarque concernant la cosinusoïde par laquelle on multiplie le signal modulé lors de la démodulation. On doit disposer d'une telle cosinusoïde au niveau du récepteur. Deux possibilités existent :

- la générer localement avec un circuit électronique adéquat. La difficulté est alors de la créer avec une fréquence et une phase strictement identiques à celles de la porteuse du signal modulé, sans quoi on observerait une dégradation plus ou moins importante dans le signal démodulé,
- filtrer le signal AM autour de la fréquence porteuse  $f_c$  avec un filtre passe-bande à bande étroite afin d'extraire exclusivement la porteuse. Ceci constitue un système de **récupération de porteuse**. La porteuse ainsi récupérée peut alors servir à démoduler le signal AM. Ceci constitue un avantage de la modulation AM par rapport à d'autres techniques de modulation qui ne transmettent pas la porteuse. Justement, puisqu'on parle...

### 2.3.2 Modulation DSB-SC

Nous venons de le voir, la modulation AM transmet, en plus de l'information utile  $m(t)$ , la porteuse. Ceci constitue un avantage à la démodulation mais un inconvénient majeur au niveau de la consommation de la puissance. Dès lors, pour remédier à cet inconvénient, pourquoi ne pas retirer cette porteuse du signal modulé ? Ceci conduit alors à la modulation d'amplitude à porteuse supprimée, ou encore modulation DSB-SC pour “Double Side-Band Suppressed Carrier”. Cette modulation constitue donc une modulation dérivée de la modulation AM.

Pour cette modulation, le signal modulé est défini par

$$s(t) = A_c m(t) \cos(2\pi f_c t)$$

Elle consiste donc simplement à multiplier le signal modulant  $m(t)$  par la porteuse  $c(t) = A_c \cos(2\pi f_c t)$ . La figure 2.13 (colonne de gauche) illustre la construction d'un signal modulé DSB-SC pour un signal modulant donné.

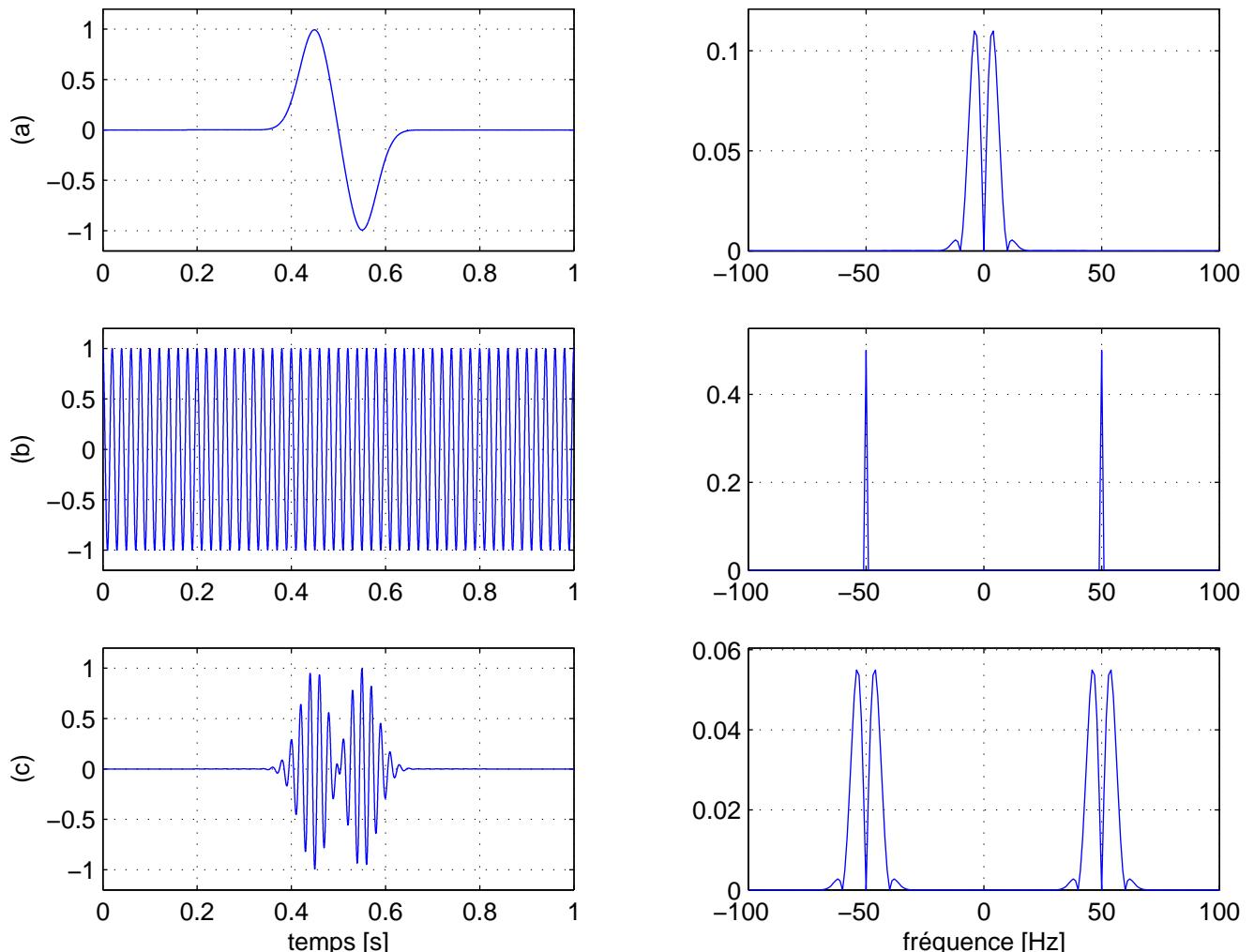


FIGURE 2.13 – Modulation DSB-SC pour laquelle  $A_c = 1$  et  $f_c = 50$  [Hz] : (a) Signal modulant  $m(t)$ . (b) Porteuse  $c(t)$ . (c) Signal modulé  $s(t) = m(t)c(t)$ .

## Spectre et bande passante

Le spectre du signal DSB-SC est obtenu en calculant la transformée de FOURIER de  $s(t)$  :

$$S(f) = \frac{A_c}{2} [M(f - f_c) + M(f + f_c)] \quad (2.7)$$

Il est donc constitué de la réplique du spectre du signal modulant en  $f = \pm f_c$  et est donc fort similaire à celui de la modulation AM. La différence essentielle est ici l'absence de la porteuse et donc des raies de Dirac en  $f = \pm f_c$ . La figure 2.13 montre les différentes signaux intervenant ainsi que leur spectres respectifs.

Comme pour la modulation AM, la bande passante  $W_{DSB-SC}$  du signal modulé correspond donc à la somme des longueurs des bandes latérales inférieures et supérieures, donc

$$W_{DSB-SC} = 2W$$

qui correspond au double de la bande de base (ou au double de la bande passante du signal modulant). La modulation DSB-SC n'apporte donc aucun gain en bande passante par rapport à la modulation AM. Le gain se situe au niveau de la puissance émise.

## Démodulation

Dans le cas de la modulation DSB-SC, un détecteur d'enveloppe n'est d'aucune utilité. En effet, directement appliqué au signal modulé, il nous fournirait l'enveloppe instantanée du signal qui est

$$|m(t)|$$

Nous perdons donc toute l'information du signe de  $m(t)$ .

Il est donc nécessaire ici d'utiliser le démodulateur synchrone de la figure 2.10. Pour un signal DSB-SC, nous obtenons donc, après multiplication par la cosinusoïde,

$$\begin{aligned} p(t) &= s(t) \cos(2\pi f_c t) \\ &= A_c m(t) \cos^2(2\pi f_c t) \\ &= \frac{A_c}{2} m(t) (1 + \cos(4\pi f_c t)) \\ &= \frac{A_c}{2} m(t) + \frac{A_c}{2} m(t) \cos(4\pi f_c t) \end{aligned}$$

Le premier terme de cette expression est un signal basse fréquence qui correspond au signal utile  $m(t)$  au facteur  $A_c/2$  près. Ce terme nous intéresse tout particulièrement car il est assez facile d'en sortir  $m(t)$ . Par contre, le second terme est un signal haute fréquence, qui correspond à la modulation DSB-SC de  $m(t)$  autour de la fréquence porteuse  $2f_c$ . Il constitue donc un terme parasite qu'il nous faut éliminer. Ceci est fait par l'application du filtre passe-bas qui nous fournit donc le signal suivant

$$r(t) = \frac{A_c}{2} m(t)$$

hors duquel on extrait facilement

$$m(t) = \frac{2}{A_c} r(t)$$

L'interprétation fréquentielle de la démodulation synchrone dans le cas d'un signal DSB-SC est représentée schématiquement à la figure 2.14.

En guise d'illustration, la figure 2.15 représente la démodulation du signal modulé de la figure 2.7.

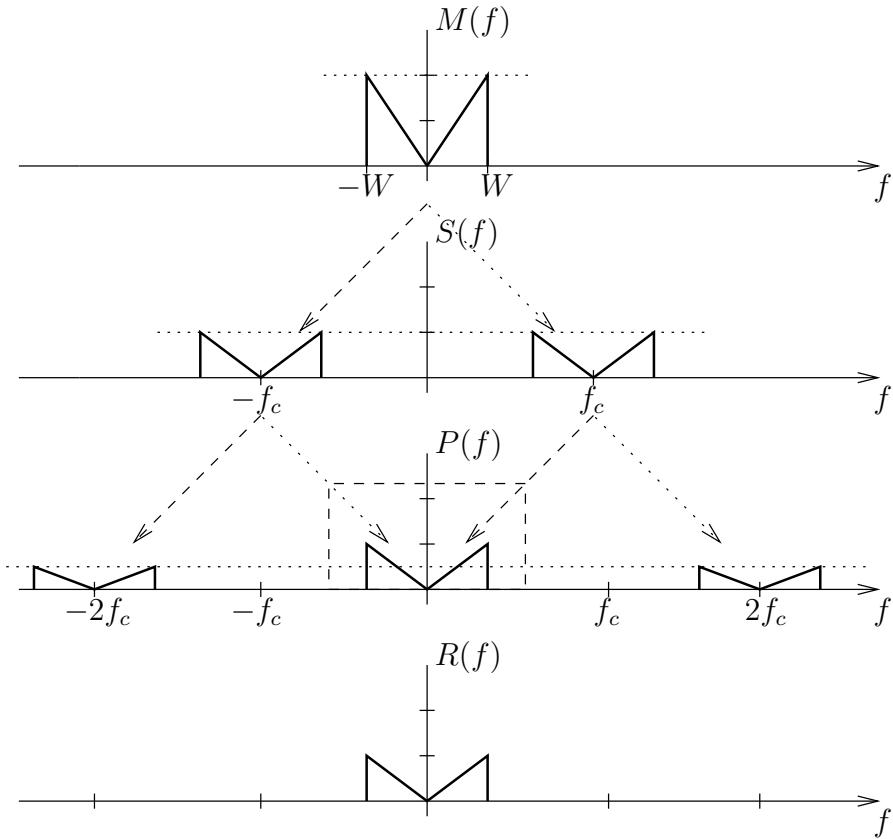


FIGURE 2.14 – Interprétation fréquentielle de la démodulation synchrone d'un signal DSB-SC.

### 2.3.3 Modulation BLU

Le passage de la modulation AM à la modulation DSB-SC a permis de gagner en puissance d'émission mais pas en bande passante. En effet, pour un même signal modulant  $m(t)$ , les bandes passantes des signaux modulés AM et DSB-SC sont identiques. Ne serait-il pas possible d'obtenir une modulation pour laquelle on bénéficierait de l'avantage de la modulation DSB-SC mais présentant une plus faible bande passante ? Bien si... Voyons cela.

Si on observe le spectre du signal DSB-SC, on remarque qu'il est symétrique autour de la fréquence porteuse  $f_c$ . Les bandes latérales supérieure BLS et inférieure BLI contiennent en effet la même information. Cela est dû au fait que le signal modulant  $m(t)$  est réel (voir propriétés de la transformées de FOURIER d'un signal réel). L'idée est donc de supprimer une des deux bandes latérales. Cela conduit à la modulation à bande latérale unique (BLU). Deux possibilités :

- on ne garde que la bande latérale supérieure, cela conduit à la modulation à bande latérale supérieure (BLS, ou USB pour “Upper Side Band”),
- on ne garde que la bande latérale inférieure, cela conduit à la modulation à bande latérale inférieure (BLI, ou LSB pour “Lower Side Band”)

Donc, pour obtenir un signal modulé BLU, deux étapes sont nécessaires :

1. Créer le signal modulé DSB-SC à la fréquence porteuse souhaitée, et
2. Réaliser un filtrage passe-bande afin de garder la bande latérale supérieure (modulation BLS) ou la bande latérale inférieure (modulation BLI).

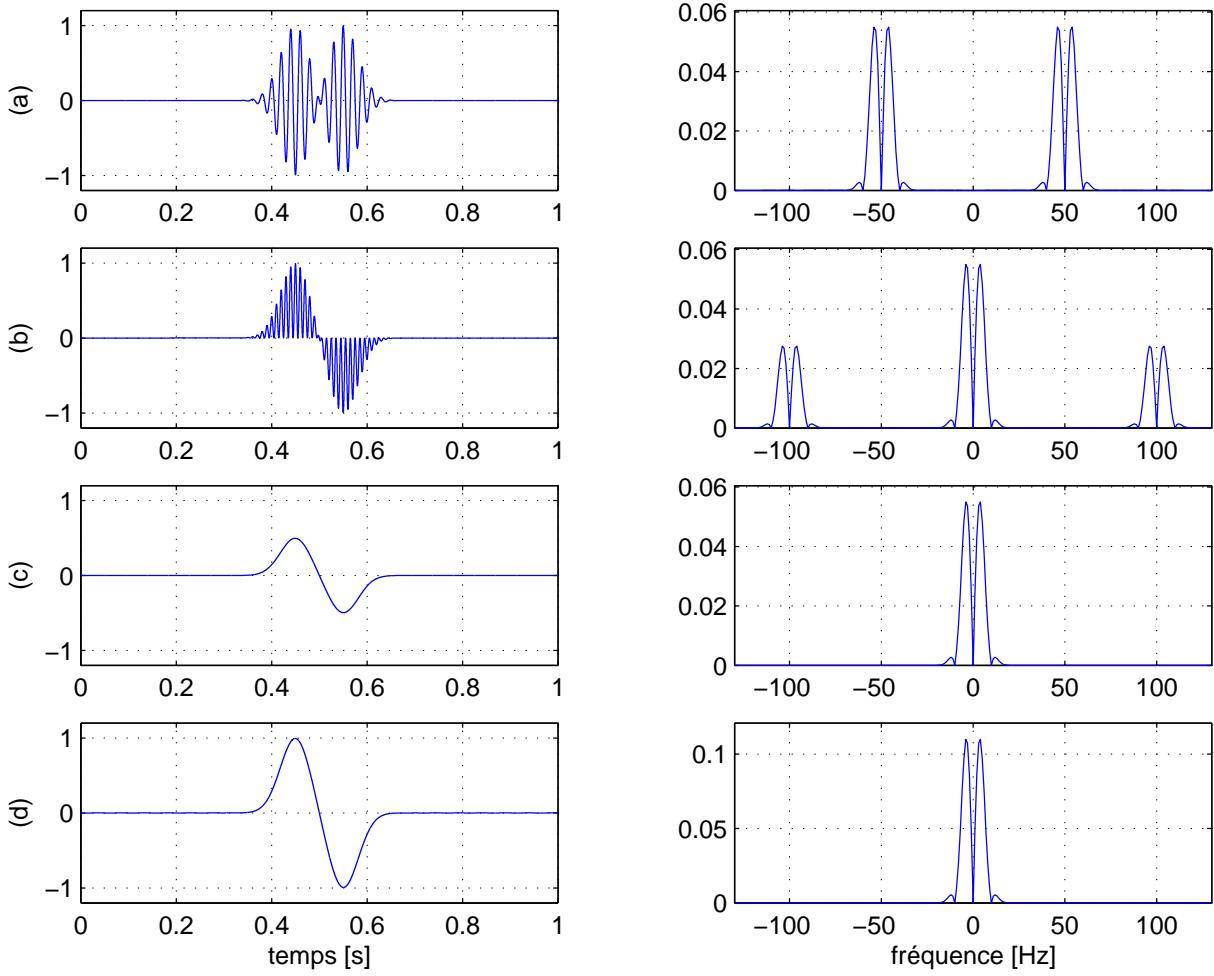


FIGURE 2.15 – Démodulation du signal DSB-SC ( $A_c = 1$ ,  $f_c = 50$  [Hz]) de la figure 2.13 : (a) Signal modulé  $s(t)$ . (b) Signal  $p(t)$ . (c) Signal  $r(t)$  obtenu par filtrage passe-bas du signal  $p(t)$ . (d) Signal  $m(t)$  retrouvé à partir de  $r(t)$  via la relation  $m(t) = 2r(t)$ .

Le schéma du modulateur est représenté à la figure 2.16. La figure 2.17 illustre la création d'un signal BLU du point de vue fréquentiel, tandis que la figure 2.18 présente un exemple de modulation BLS pour un signal modulant donné.

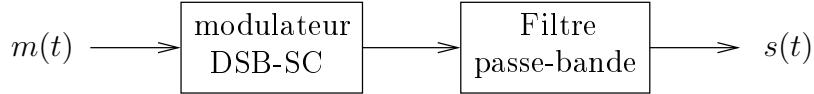


FIGURE 2.16 – Modulateur BLU.

Nous savons donc à présent ce qu'est la modulation BLU, mais nous n'avons encore aucune expression pour le signal modulé. On pourrait montrer par quelques calculs fastidieux qui décourageraient un bon nombre de nos lecteurs que celle-ci est donnée par

$$s_{BLS}(t) = A_c m(t) \cos(2\pi f_c t) - A_c \tilde{m}(t) \sin(2\pi f_c t) \quad (2.8)$$

$$s_{BLI}(t) = A_c m(t) \cos(2\pi f_c t) + A_c \tilde{m}(t) \sin(2\pi f_c t) \quad (2.9)$$

où  $\tilde{m}(t)$  est la transformée de HILBERT de  $m(t)$ . Bien qu'assez repoussantes, ces expressions sont intéressantes dans le sens qu'elles permettent de concevoir un modulateur dépourvu de

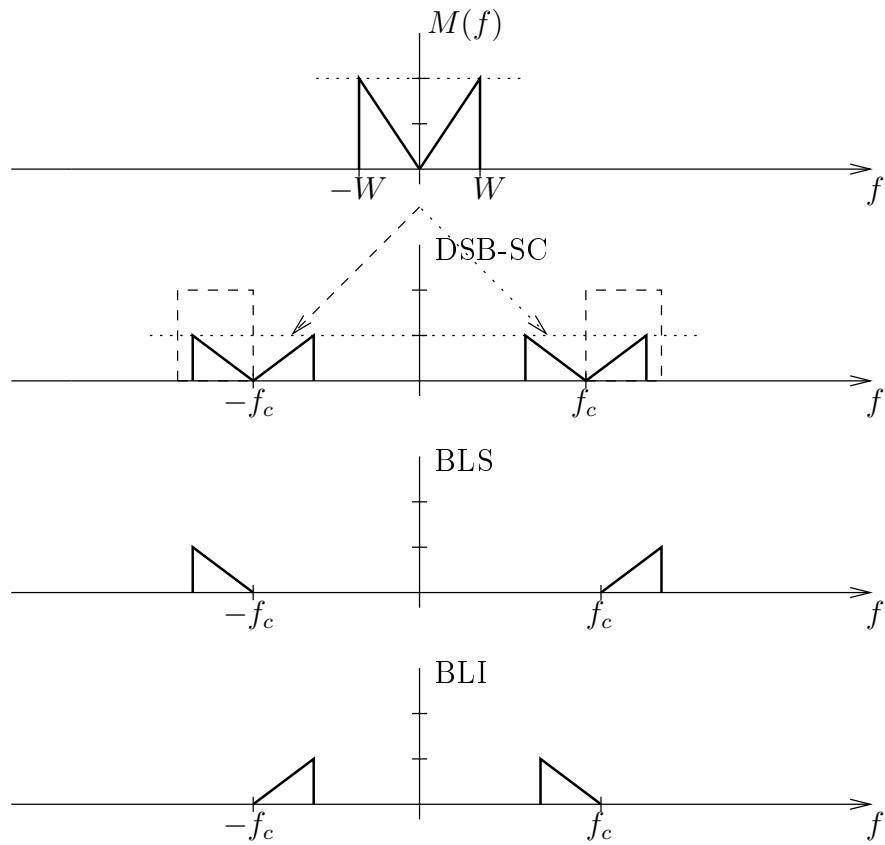


FIGURE 2.17 – Construction d'un signal BLU dans le domaine fréquentiel.

filtre passe-bande dont la réalisation pratique est difficile pour une bande passante située dans les hautes fréquences. Si nous disposons d'un circuit calculant la transformée de HILBERT d'un signal et des porteuses  $A_c \cos(2\pi f_c t)$  et  $A_c \sin(2\pi f_c t)$ , le modulateur de la figure 2.19 fait très bien l'affaire.

### Remarque

La transformée de HILBERT d'un signal  $m(t)$  est définie par

$$\tilde{m}(t) = m(t) \otimes \frac{1}{\pi t}$$

ou encore en fréquentiel par

$$\widetilde{M}(f) = -j \operatorname{sign}(f) M(f)$$

### Spectre et bande passante

L'expression mathématique de la transformée de FOURIER d'un signal modulé en BLU ne nous sera pas d'une grande utilité. Nous la passerons donc sous silence... Quoique... Non, n'insistons pas ! L'important est surtout de se rappeler qu'une des bandes latérales à été supprimée et que maintenant la bande passante du signal modulé est égale à

$$W_{BLU} = W$$

Elle est donc égale à la bande de base du signal modulant. Nous avons donc gagné un facteur 2 pour la bande passante par rapport à la modulation DSB-SC ou AM.

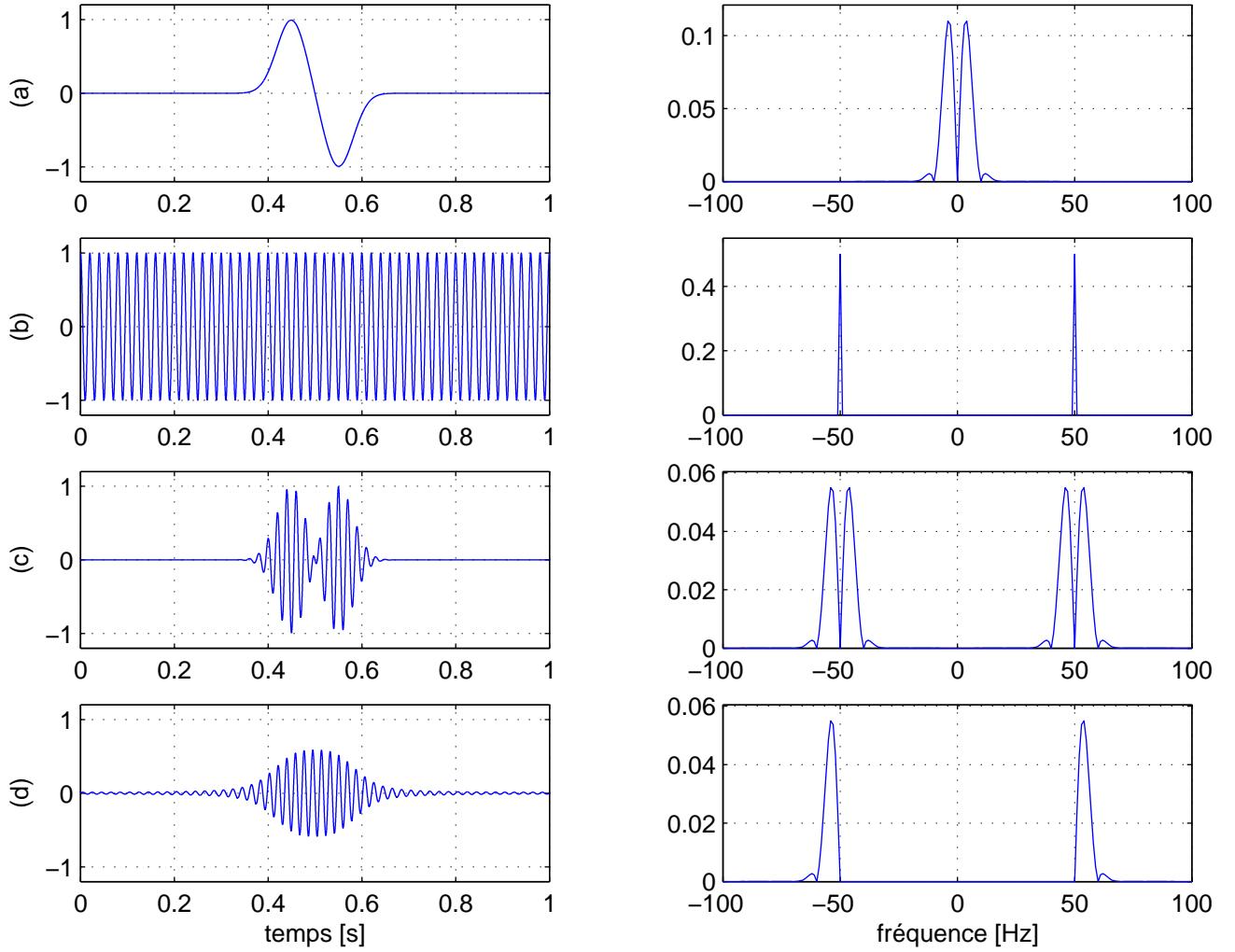


FIGURE 2.18 – Modulation BLS pour laquelle  $f_c = 50 \text{ [Hz]}$  : (a) Signal modulant  $m(t)$ . (b) Porteuse  $c(t)$ . (c) Signal modulé DSB-SC. (d) Signal modulé BLS obtenu par filtrage passe-bande du signal DSB-SC.

## Démodulation

Aussi surprenant que cela puisse l'être, la démodulation d'un signal BLU se réalise grâce au détecteur synchrone de la figure 2.10, comme pour les modulations AM et DSB-SC. Nous obtenons donc, après multiplication par la cosinusoïde,

$$\begin{aligned}
 p(t) &= s(t) \cos(2\pi f_c t) \\
 &= A_c m(t) \cos^2(2\pi f_c t) \pm A_c \tilde{m}(t) \sin(2\pi f_c t) \cos(2\pi f_c t) \\
 &= \frac{A_c}{2} m(t) (1 + \cos(4\pi f_c t)) \pm \frac{A_c}{2} \tilde{m}(t) \sin(4\pi f_c t) \\
 &= \frac{A_c}{2} m(t) + \frac{A_c}{2} m(t) \cos(4\pi f_c t) \pm \frac{A_c}{2} \tilde{m}(t) \sin(4\pi f_c t)
 \end{aligned}$$

Le premier terme de cette expression est un signal basse fréquence qui correspond au signal utile  $m(t)$  au facteur  $A_c/2$  près. Ce terme nous intéresse tout particulièrement car il est assez facile d'en sortir  $m(t)$ . Par contre, le second et le troisième termes sont des signaux haute fréquence, qui correspondent respectivement à la modulation DSB-SC de  $m(t)$  et de  $\tilde{m}(t)$  autour de la fréquence porteuse  $2f_c$ . Ils constituent donc des termes parasites qu'il nous faut éliminer. Ceci

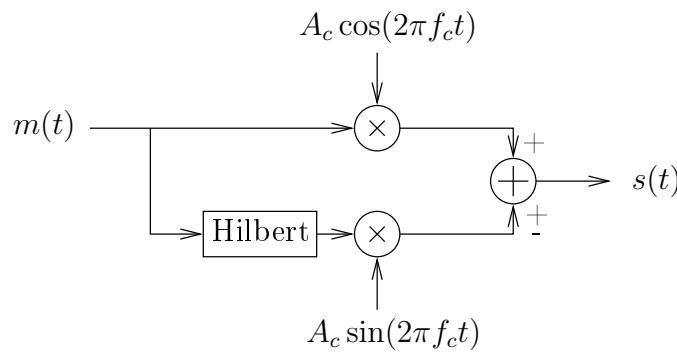


FIGURE 2.19 – Modulateur BLU utilisant la transformée de HILBERT.

est fait par l'application du filtre passe-bas qui nous fournit donc le signal suivant

$$r(t) = \frac{A_c}{2} m(t)$$

hors duquel on extrait facilement

$$m(t) = \frac{2}{A_c} r(t)$$

L'interprétation fréquentielle de la démodulation synchrone dans le cas d'un signal BLS est représentée schématiquement à la figure 2.20.

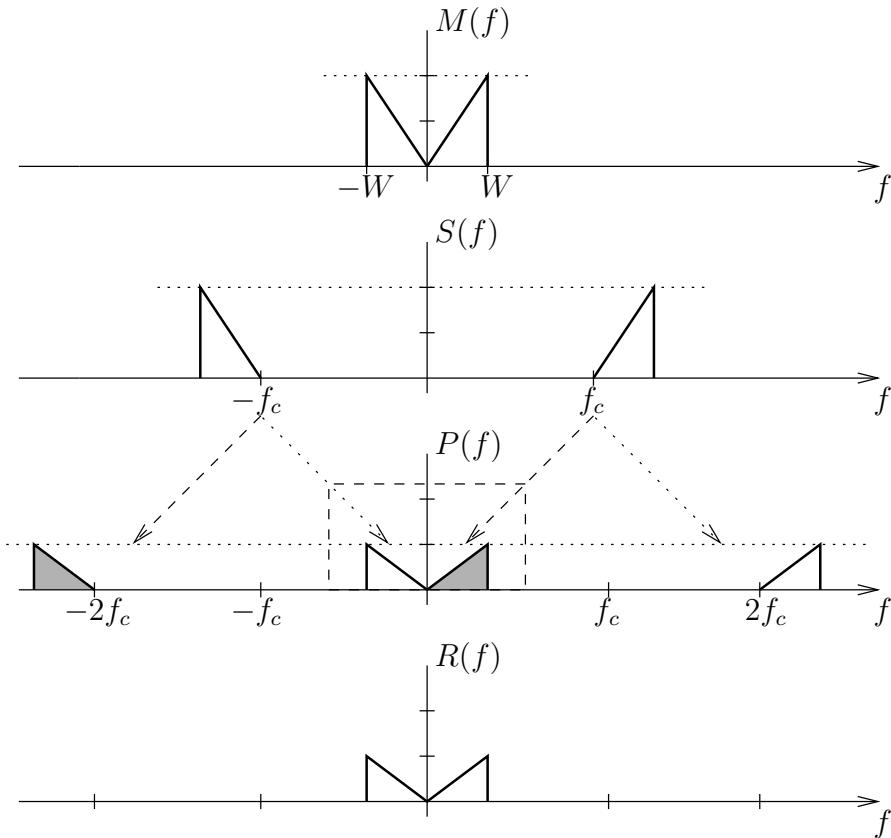


FIGURE 2.20 – Interprétation fréquentielle de la démodulation synchrone d'un signal BLS.

En guise d'illustration, la figure 2.21 représente la démodulation du signal modulé de la figure 2.18.

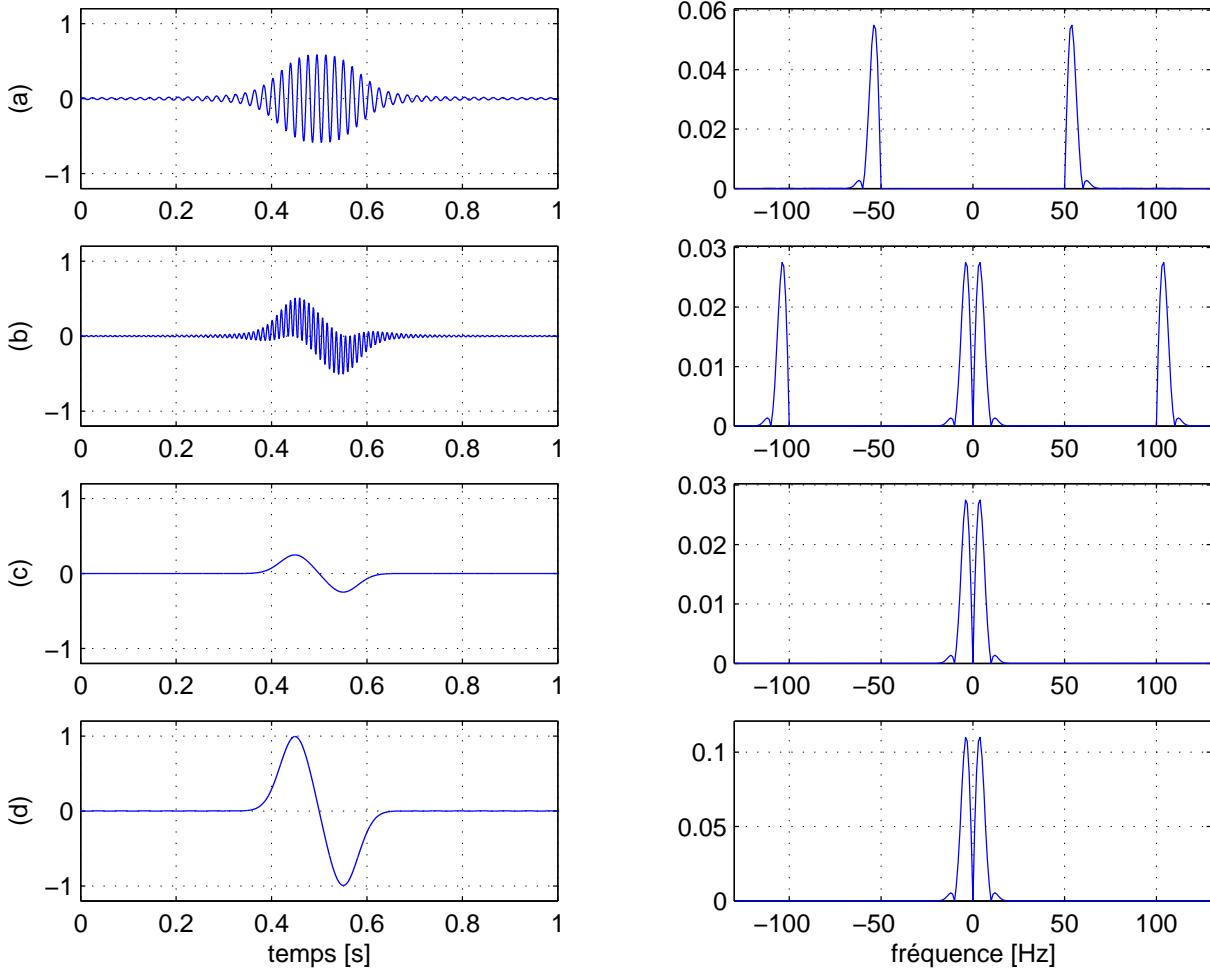


FIGURE 2.21 – Démodulation du signal BLS ( $f_c = 50$  [Hz]) de la figure 2.18 : (a) Signal modulé  $s(t)$ . (b) Signal  $p(t)$ . (c) Signal  $r(t)$  obtenu par filtrage passe-bas du signal  $p(t)$ . (d) Signal  $m(t)$  retrouvé à partir de  $r(t)$  via la relation  $m(t) = 4r(t)$ .

### 2.3.4 Modulation en quadrature

L'idée d'utiliser le cosinus et le sinus d'une porteuse peut également servir à l'émission de deux signaux modulants,  $m_1(t)$  et  $m_2(t)$ , ayant une même bande de base. On construit le signal suivant, appelé **signal modulé en quadrature**,

$$s(t) = A_c m_1(t) \cos(2\pi f_c t) + A_c m_2(t) \sin(2\pi f_c t) \quad (2.10)$$

Il s'agit de la somme de deux signaux DSB-SC modulé à la même fréquence porteuse  $f_c$ , l'un par le cosinus, l'autre par le sinus de la porteuse. Le premier terme est dit **en phase** avec la porteuse  $c(t) = A_c \cos(2\pi f_c t)$ , tandis que le second est dit **en quadrature** avec la porteuse  $c(t)$ . Le schéma du modulateur est donné à la figure 2.22. La figure 2.23 présente les différents signaux intervenant lors de la modulation en quadrature de deux signaux modulants donnés.

Il y a un réel intérêt à utiliser la modulation en quadrature. En effet, les deux signaux modulés (en DSB-SC), composant le signal modulé en quadrature, occupent la même bande de fréquence, ce qui signifie une réduction de la bande passante d'un facteur 2. Mais alors, comment retrouver les deux signaux  $m_1(t)$  et  $m_2(t)$  alors que leur spectres modulés se chevauchent autour de  $f_c$ ? Le démodulateur de la figure 2.24 fournit la solution.

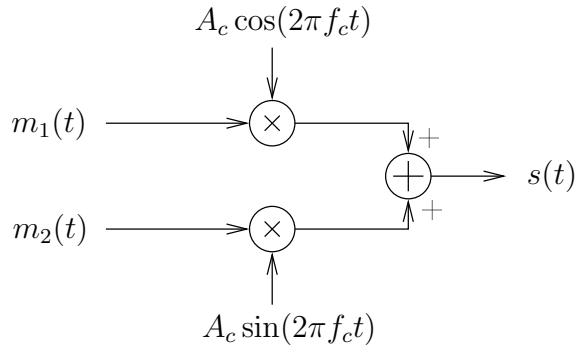


FIGURE 2.22 – Schéma du modulateur en quadrature.

Étudions tout d'abord la récupération du signal  $m_1(t)$ . Après multiplication du signal modulé  $s(t)$  par  $\cos(2\pi f_c t)$ , nous obtenons

$$\begin{aligned} p_1(t) &= s(t) \cos(2\pi f_c t) \\ &= A_c m_1(t) \cos^2(2\pi f_c t) + A_c m_2(t) \sin(2\pi f_c t) \cos(2\pi f_c t) \\ &= \frac{A_c}{2} m_1(t) (1 + \cos(4\pi f_c t)) + \frac{A_c}{2} m_2(t) \sin(4\pi f_c t) \\ &= \frac{A_c}{2} m_1(t) + \frac{A_c}{2} m_1(t) \cos(4\pi f_c t) + \frac{A_c}{2} m_2(t) \sin(4\pi f_c t) \end{aligned}$$

Le premier terme de cette expression est un signal basse fréquence qui correspond au signal utile  $m_1(t)$  au facteur  $A_c/2$  près. Ce terme nous intéresse tout particulièrement car il est assez facile d'en sortir  $m_1(t)$ . Par contre, le second et le troisième termes sont des signaux haute fréquence, qui correspondent respectivement à la modulation DSB-SC de  $m_1(t)$  et de  $m_2(t)$  autour de la fréquence porteuse  $2f_c$ . Ils constituent donc des termes parasites qu'il nous faut éliminer. Ceci est fait par l'application du filtre passe-bas qui nous fournit donc le signal suivant

$$r_1(t) = \frac{A_c}{2} m_1(t)$$

hors duquel on extrait facilement

$$m_1(t) = \frac{2}{A_c} r_1(t)$$

La récupération du signal  $m_2(t)$  est tout à fait analogue. Après multiplication du signal modulé  $s(t)$  par  $\sin(2\pi f_c t)$ , nous obtenons

$$\begin{aligned} p_2(t) &= s(t) \sin(2\pi f_c t) \\ &= A_c m_1(t) \cos(2\pi f_c t) \sin(2\pi f_c t) + A_c m_2(t) \sin^2(2\pi f_c t) \\ &= \frac{A_c}{2} m_1(t) \sin(4\pi f_c t) + \frac{A_c}{2} m_2(t) (1 - \cos(4\pi f_c t)) \\ &= \frac{A_c}{2} m_2(t) - \frac{A_c}{2} m_2(t) \cos(4\pi f_c t) + \frac{A_c}{2} m_1(t) \sin(4\pi f_c t) \end{aligned}$$

Le premier terme de cette expression est un signal basse fréquence qui correspond au signal utile  $m_2(t)$  au facteur  $A_c/2$  près. Ce terme nous intéresse tout particulièrement car il est assez facile d'en sortir  $m_2(t)$ . Par contre, le second et le troisième termes sont des signaux haute fréquence, qui correspondent respectivement à la modulation DSB-SC de  $m_2(t)$  et de  $m_1(t)$  autour de la

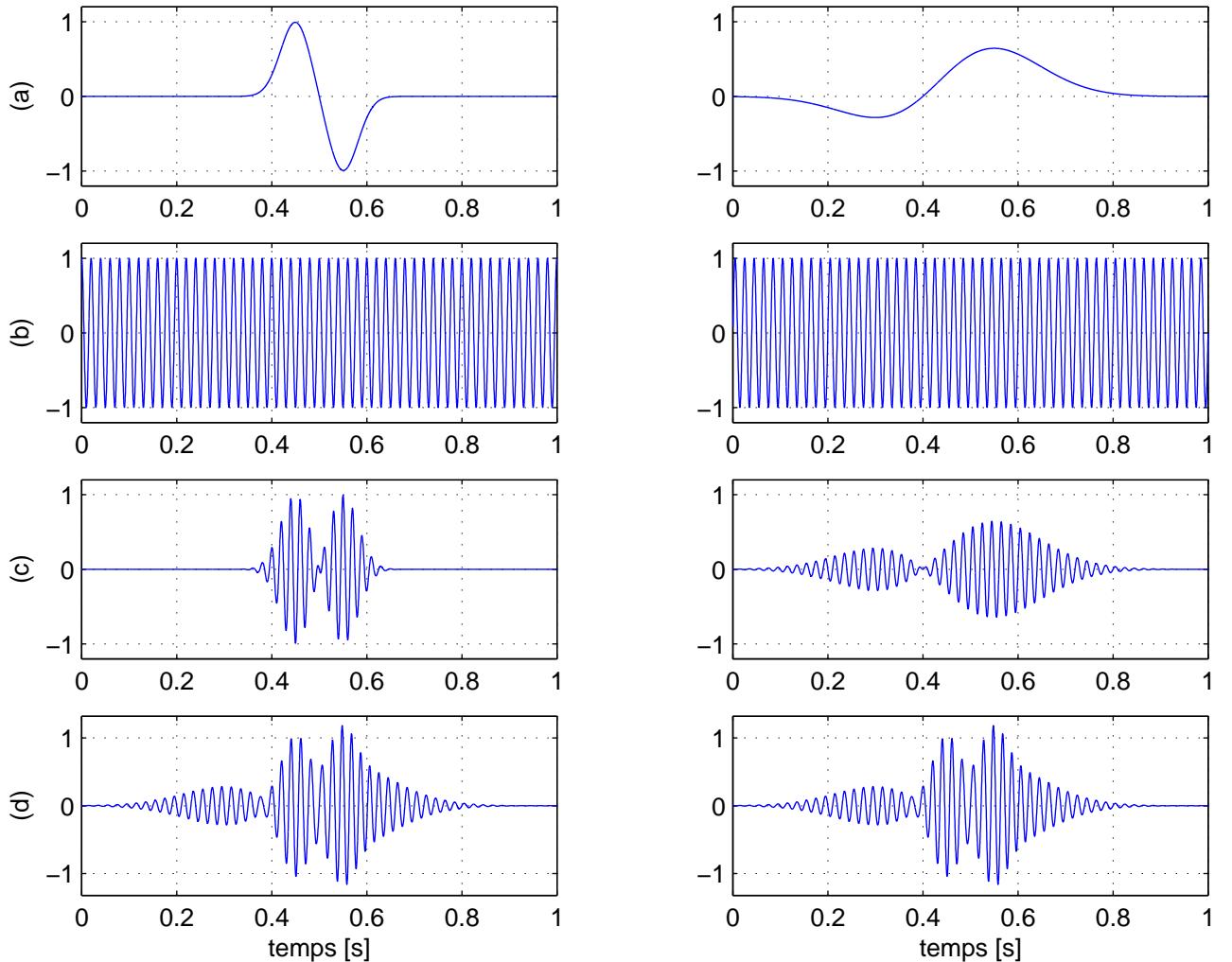


FIGURE 2.23 – Modulation en quadrature ( $f_c = 50 [Hz]$ ,  $A_c = 1$ ) : (a) Signaux modulants,  $m_1(t)$  à gauche et  $m_2(t)$  à droite. (b) Porteuses,  $A_c \cos(2\pi f_c t)$  à gauche et  $A_c \sin(2\pi f_c t)$  à droite. (c) Signaux en phase et en quadrature,  $A_c m_1(t) \cos(2\pi f_c t)$  à gauche et  $A_c m_2(t) \sin(2\pi f_c t)$  à droite. (d) Signal modulé  $s(t)$ , identique à gauche et à droite.

fréquence porteuse  $2f_c$ . Ils constituent donc des termes parasites qu'il nous faut éliminer. Ceci est fait par l'application du filtre passe-bas qui nous fournit donc le signal suivant

$$r_2(t) = \frac{A_c}{2} m_2(t)$$

hors duquel on extrait facilement

$$m_2(t) = \frac{2}{A_c} r_2(t)$$

En guise d'illustration, la figure 2.25 présente les signaux intervenant lors de la démodulation du signal de la figure 2.23.

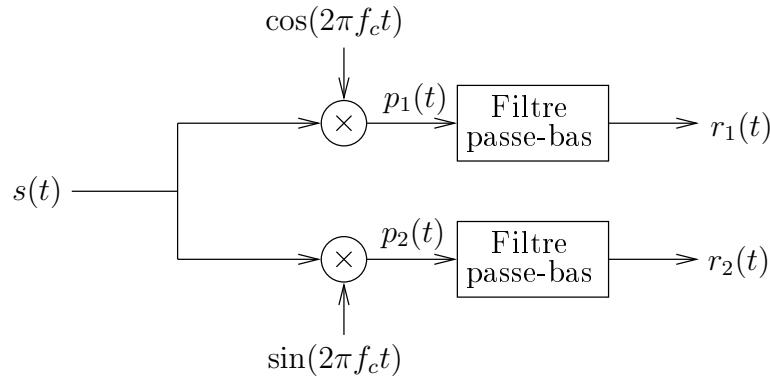


FIGURE 2.24 – Démodulateur en quadrature.

## 2.4 Modulation angulaire

Dans la modulation angulaire, le signal modulé prend la forme

$$s(t) = A_c \cos \Phi_i(t)$$

où, pour rappel,  $\Phi_i(t)$  est l'angle instantané du signal modulé. En l'absence de modulation, celui-ci est égal à

$$\Phi_i(t) = 2\pi f_c t + \phi_c$$

La modulation angulaire consiste à introduire le signal modulant  $m(t)$  dans  $\Phi_i(t)$  de telle sorte que celui-ci soit une fonction linéaire de  $m(t)$ . Deux possibilités existent :

- on remplace  $\phi_c$  par une fonction linéaire de  $m(t)$ . Dans ce cas, on obtient une **modulation de phase**.
- on remplace la fréquence instantanée de la porteuse (actuellement constante et égale à  $f_c$  dans le cas non modulé) par une fonction linéaire de  $m(t)$ . Dans ce cas, on obtient une **modulation de fréquence**.

Néanmoins, comme nous allons le voir, une modulation de phase s'accompagne nécessairement d'une modulation de fréquence et réciproquement. De plus, la fonction cosinus avec son angle ainsi modifié n'est plus fonction linéaire du temps, ce qui va nous poser quelques soucis pour déterminer le spectre du signal modulé. Enfin, remarquons que la modulation angulaire ne modifie pas l'amplitude  $A_c$  de la porteuse.

### Quelques paramètres importants

On appelle **déviation instantanée de phase** (ou d'angle) la grandeur

$$\Delta\Phi_i(t) = \Phi_i(t) - (2\pi f_c t + \phi_c) \quad (2.11)$$

c'est-à-dire l'écart entre l'angle du signal modulé et l'angle de la porteuse non modulée. L'amplitude de la déviation instantanée de phase

$$\beta = \max |\Delta\Phi_i(t)| \quad (2.12)$$

est appelée **indice de modulation**.

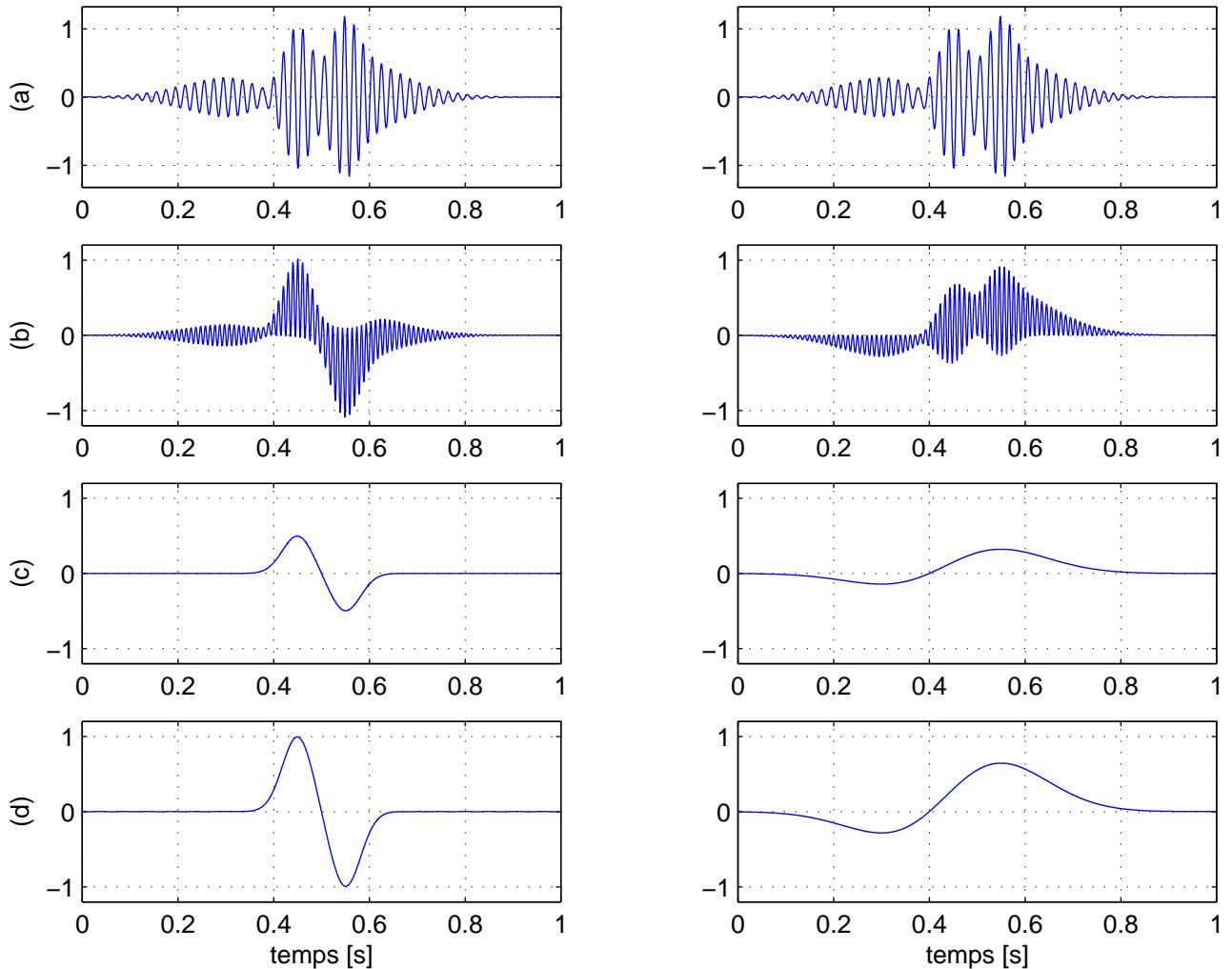


FIGURE 2.25 – Démodulation en quadrature du signal de la figure 2.23 : (a) Signal modulé  $s(t)$  à gauche et droite. (b) Multiplication par les porteuses,  $p_1(t)$  à gauche et  $p_2(t)$  à droite. (c) Filtrage passe-bas,  $r_1(t)$  à gauche et  $r_2(t)$  à droite. (d) Récupération de  $m_1(t)$  et  $m_2(t)$  via les relations  $m_1(t) = 2r_1(t)$  et  $m_2(t) = 2r_2(t)$ .

On définit également la *déviation instantanée de fréquence*  $\Delta f_i(t)$  comme l'écart entre la fréquence instantanée  $f_i(t)$  du signal modulé et la fréquence de la porteuse  $f_c$  :

$$\Delta f_i(t) = f_i(t) - f_c \quad (2.13)$$

Le maximum de la déviation instantanée de fréquence fournit l'*excursion de fréquence*  $\Delta f$  définie par

$$\boxed{\Delta f = \max |\Delta f_i(t)|} \quad (2.14)$$

Les modulation angulaire consiste à faire varier, selon une loi linéaire bien précise, une des quantités  $\Delta\Phi_i(t)$  ou  $\Delta f_i(t)$ . Étant donné les définitions qui précèdent, il apparaît qu'on ne peut faire varier l'une sans l'autre ; une modulation de phase entraîne une modulation de fréquence et réciproquement.

### 2.4.1 Modulation de phase (PM)

La modulation de phase (PM pour “Phase Modulation”) s’exprime sous la forme du signal modulé suivant

$$s(t) = A_c \cos(2\pi f_c t + \phi_i(t))$$

où  $\phi_i(t)$  est donnée par

$$\phi_i(t) = k_p m(t)$$

$k_p$  étant appelé *sensibilité du modulateur*. Dès lors, le signal modulé PM peut s’écrire

$$s(t) = A_c \cos(2\pi f_c t + k_p m(t)) \quad (2.15)$$

La fréquence instantanée est alors égale à

$$f_i(t) = \frac{1}{2\pi} \frac{d\Phi_i(t)}{dt} = \frac{1}{2\pi} \frac{d}{dt}(2\pi f_c t + k_p m(t)) = f_c + \frac{k_p}{2\pi} \frac{dm}{dt}(t) \quad (2.16)$$

il s’en suit que la modulation de phase modifie la fréquence de la porteuse. La déviation de fréquence instantanée vaut

$$\Delta f_i(t) = \frac{k_p}{2\pi} \frac{dm}{dt}(t)$$

Un exemple de modulation PM pour un signal modulant donné est présenté à la figure 2.26c.

### 2.4.2 Modulation de fréquence (FM)

Dans la modulation de fréquence (FM pour “Frequency Modulation”), la fréquence instantanée  $f_i(t)$  du signal modulé est une fonction linéaire du signal modulant  $m(t)$  :

$$f_i(t) = f_c + k_f m(t) \quad (2.17)$$

où  $k_f$  est appelé *sensibilité du modulateur*. L’expression du signal modulé s’obtient à calculant tout d’abord l’angle (ou la phase) instantanée en se rappelant de (2.4) :

$$\begin{aligned} \Phi_i(t) &= 2\pi \int_{-\infty}^t f_i(\tau) d\tau \\ &= 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \end{aligned}$$

Dès lors, le signal modulé vaut

$$s(t) = A_c \cos \left( 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) \quad (2.18)$$

Un exemple de modulation FM pour signal modulant donné est présenté à la figure 2.26d.

Les relations (2.16) et (2.17) mettent bien en évidence qu’une modulation de phase entraîne une modulation de fréquence, et vice versa. On peut même affirmer que la modulation de phase se réduit à une modulation de fréquence par le signal modulant préalablement dérivé. Inversement, une modulation de fréquence est une modulation de phase de la primitive du signal modulant. Ceci est illustré à la figure 2.27. Dans la suite de cette étude, nous nous limiterons donc à la modulation FM.

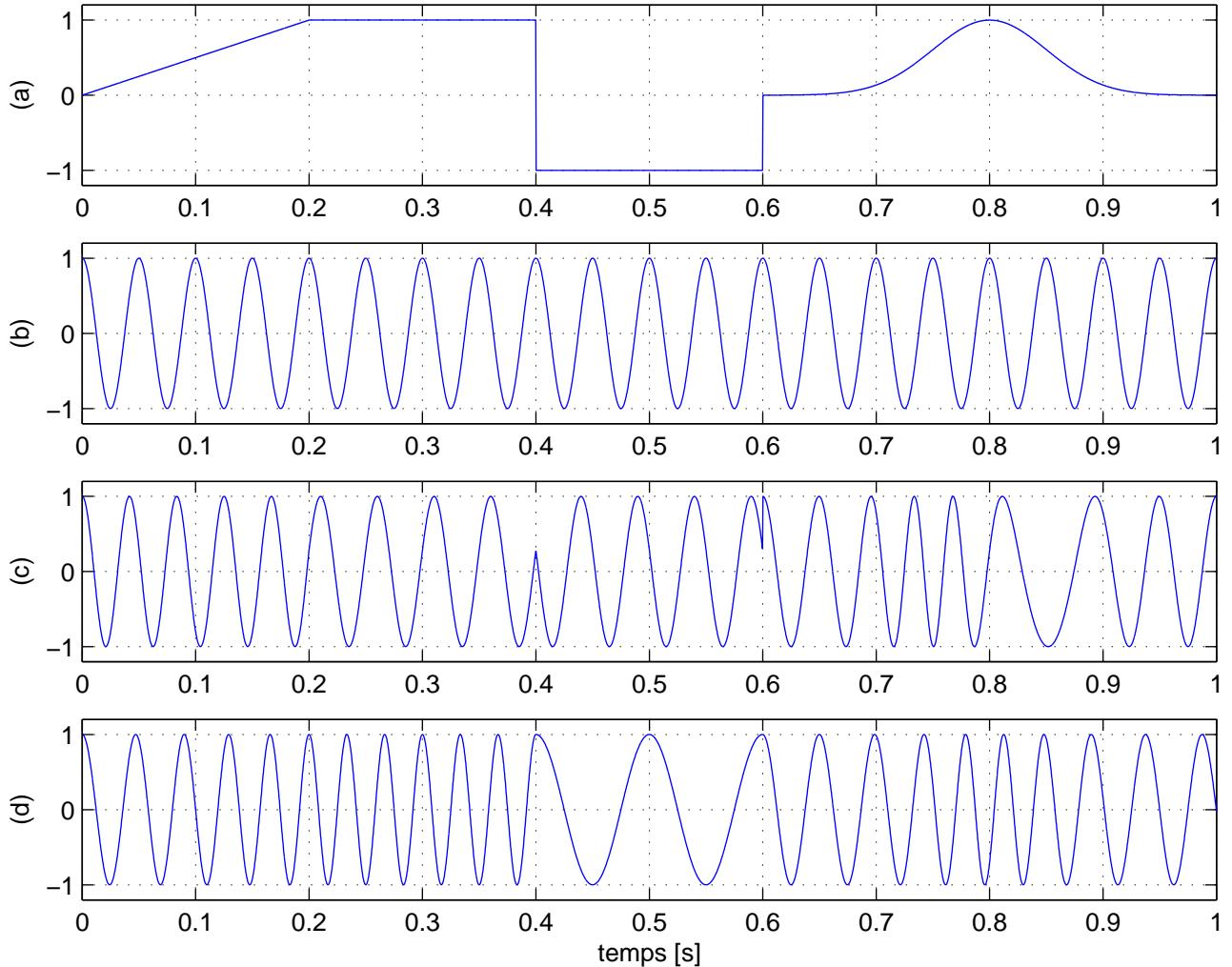


FIGURE 2.26 – Exemples de modulation PM et FM : (a) Signal modulant  $m(t)$ . (b) Porteuse  $c(t) = A_c \cos(2\pi f_c t)$  pour  $A_c = 1$  et  $f_c = 20$  [Hz]. (c) Signal modulé PM pour  $k_p = 5$ . (d) Signal modulé FM pour  $k_f = 10$ .

### Analyse spectrale de la modulation FM

L'étude fréquentielle de la modulation FM n'est pas aisée, mais est néanmoins indispensable si on veut déterminer la bande passante d'un signal FM. Cela est dû à la non-linéarité liant le signal modulé  $s(t)$  au signal modulant  $m(t)$ . Comme nous allons le voir, nous allons devoir ruser... Nous devrons user d'approximations ou nous limiter aux cas de signaux simples, en essayant d'extrapoler au cas général.

Pour rappel, un signal modulé FM peut s'écrire sous la forme

$$s(t) = A_c \cos(2\pi f_c t + \theta(t))$$

où la phase instantanée, que nous avons notée  $\theta(t)$  ici, est donnée par

$$\theta(t) = 2\pi k_f \int_{-\infty}^t m(\tau) d\tau$$

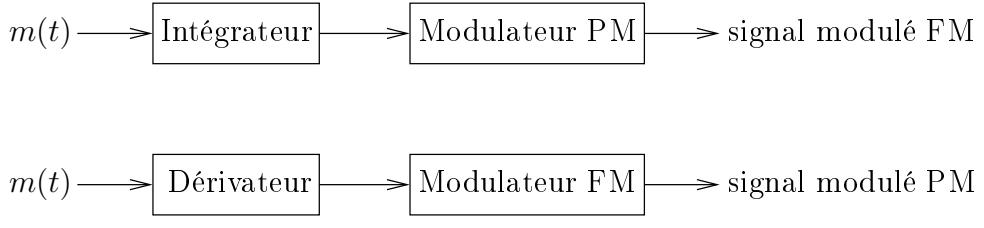


FIGURE 2.27 – Lien entre modulation de phase et de fréquence.

À partir de maintenant, nous allons considérer deux cas :

- la **modulation FM à bande étroite** (on parle de NBFM pour “Narrow Band Frequency Modulation”) qui correspond à

$$|\theta(t)| \ll 1 \text{ Radian}$$

Dans ce cas, nous pourrons user des approximations suivantes

$$\cos \theta(t) \approx 1 \quad \text{et} \quad \sin \theta(t) \approx \theta(t)$$

- la **modulation FM à large bande** pour laquelle l’approximation précédente n’est plus valable. Dans ce cas, nous devrons nous limiter à l’étude du spectre du signal modulé FM pour un signal modulant sinusoïdal.

### 2.4.3 Modulation FM à bande étroite (NBFM)

En utilisant une relation bien connue de la trigonométrie, nous pouvons réécrire le signal modulé sous la forme

$$s(t) = A_c \cos(2\pi f_c t) \cos \theta(t) - A_c \sin(2\pi f_c t) \sin \theta(t)$$

L’utilisation des approximations citées plus haut conduit alors à

$$s(t) = A_c \cos(2\pi f_c t) - A_c \theta(t) \sin(2\pi f_c t)$$

#### Spectre et bande passante

Nous sommes maintenant en mesure de calculer la transformée de FOURIER du signal modulé :

$$S(f) = \frac{A_c}{2} (\delta(f - f_c) + \delta(f + f_c)) - \frac{A_c}{2j} (\Theta(f - f_c) - \Theta(f + f_c))$$

où  $\Theta(f)$  est la transformée de FOURIER de  $\theta(t)$  :

$$\Theta(f) = 2\pi k_f \frac{M(f)}{j2\pi f} = -jk_f \frac{M(f)}{f}$$

La transformée de FOURIER du signal modulé peut alors finalement s’écrire

$$S(f) = \frac{A_c}{2} (\delta(f - f_c) + \delta(f + f_c)) + \frac{k_f A_c}{2} \left( \frac{M(f - f_c)}{f - f_c} - \frac{M(f + f_c)}{f + f_c} \right) \quad (2.19)$$

Le spectre du signal modulé est donc composé de :

- deux raies de DIRAC situées en  $f = \pm f_c$ , il s'agit en fait de la porteuse ;
- les répliques du spectre de  $m(t)$ , pondéré par le facteur  $1/f$ , autour des fréquences  $f = \pm f_c$ .

Le spectre d'un signal FM à bande étroite est donc très similaire à celui d'un signal AM, mis à part le facteur  $1/f$  modifiant la forme du spectre de  $m(t)$ . La bande passante d'un signal NBFM est donc égale à

$$W_{NBFM} = 2W \quad (2.20)$$

où  $W$  est la bande de base du signal modulant.

La figure 2.28 illustre la modulation FM à bande étroite pour un signal modulant donné. On y observe que le signal  $\theta(t)$  reste bien inférieur à 1 radian, ce qui confirme la validité de l'approximation réalisée. Mais que se passe-t-il lorsque l'approximation n'est plus valide, c'est-à-dire lorsque la sensibilité  $k_f$  du modulateur est trop importante ? La figure 2.29 fournit la réponse. Dans ce cas, le spectre du signal modulé occupe une bande passante nettement supérieure ; nous sommes dans le cas d'une modulation de fréquence à large bande. Pour son étude, nous allons devoir nous limiter à un signal modulant sinusoïdal.

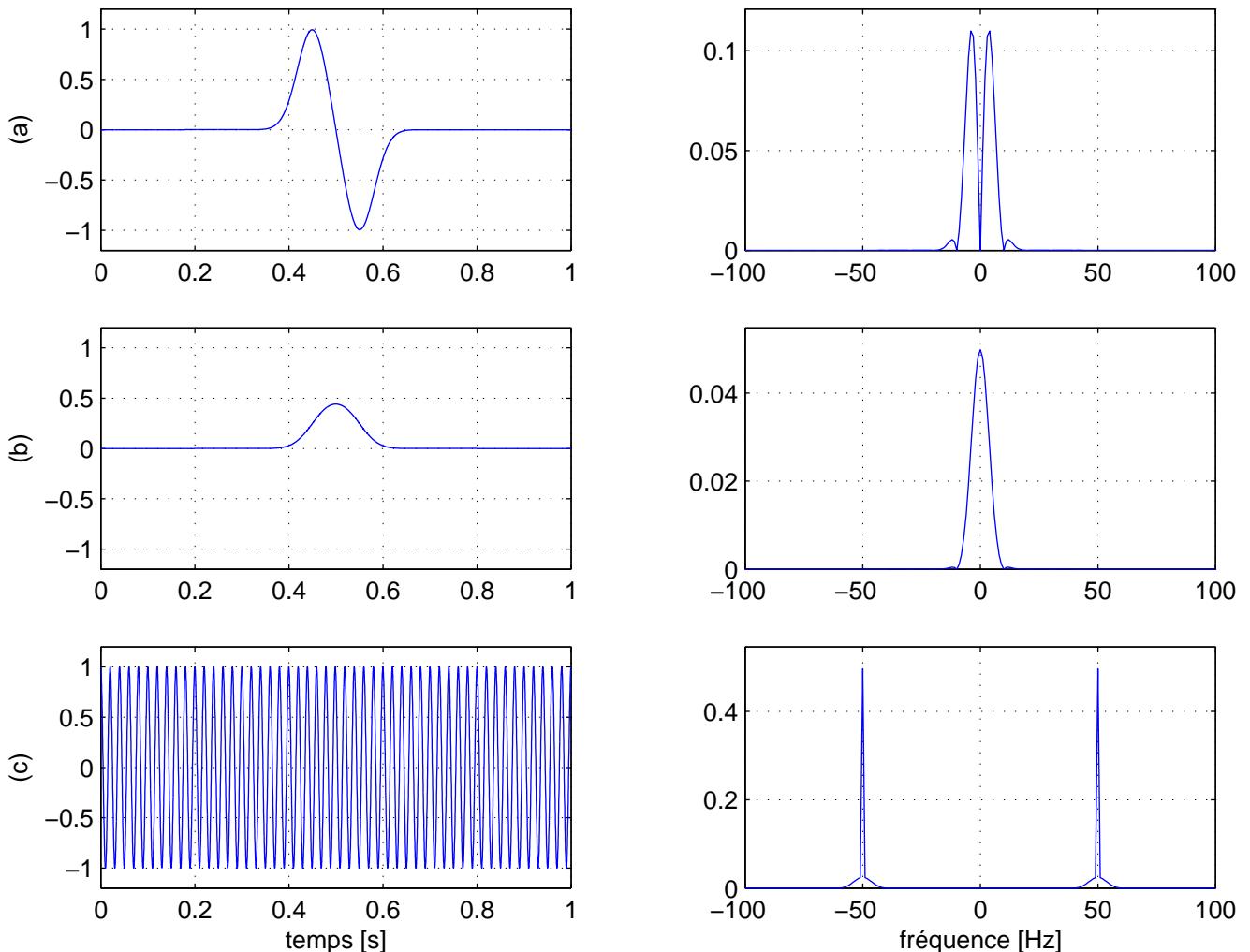


FIGURE 2.28 – Modulation FM à bande étroite ( $A_c = 1$ ,  $f_c = 50$  [Hz],  $k_f = 1$ ) : (a) Signal modulant  $m(t)$ . (b) Déphasage  $\theta(t)$  du signal modulé. (c) Signal modulé  $s(t)$ .

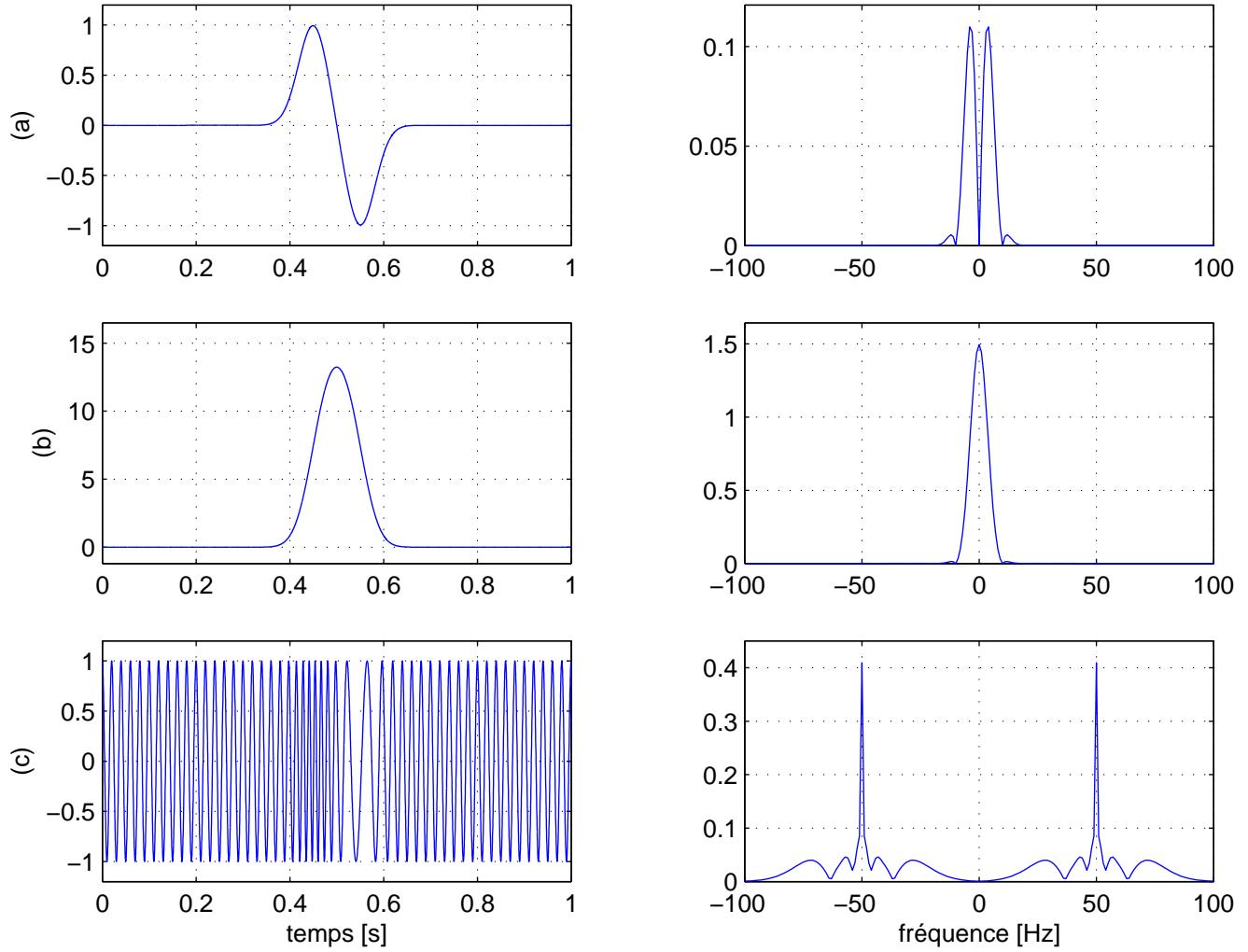


FIGURE 2.29 – Modulation FM à large bande ( $A_c = 1$ ,  $f_c = 50$  [Hz],  $k_f = 30$ ) : (a) Signal modulant  $m(t)$ . (b) Déphasage  $\theta(t)$  du signal modulé. (c) Signal modulé  $s(t)$ .

## Démodulation

Comme nous l'avons vu, le signal modulé NBFM

$$s(t) = A_c \cos(2\pi f_c t) - A_c \theta(t) \sin(2\pi f_c t)$$

où

$$\theta(t) = 2\pi k_f \int_{-\infty}^t m(\tau) d\tau$$

est assez similaire à celui de la modulation AM. En effet, on reconnaît la porteuse et la modulation DSB-SC du signal  $\theta(t)$ . Le démodulateur synchrone peut-il alors être utilisé dans le cas de la NBFM ? Presque... Le démodulateur à utiliser est celui de la figure 2.30. On y reconnaît le démodulateur synchrone dont on a remplacer  $\cos(2\pi f_c t)$  par  $\sin(2\pi f_c t)$ . De plus, la présence du déivateur est justifiée par le fait que l'on désire retrouver  $m(t)$  et non  $\theta(t)$ . Étudions à présent son fonctionnement.

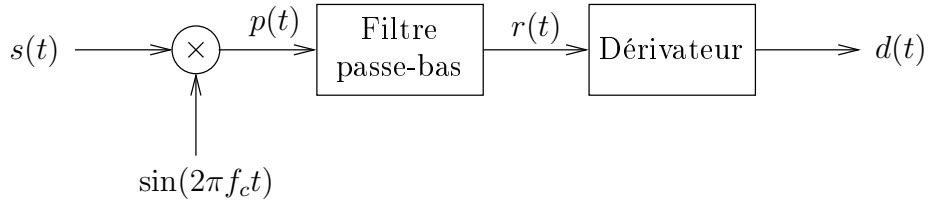


FIGURE 2.30 – Démodulateur NBFM.

Le signal  $p(t)$  obtenu par la multiplication du signal modulé par  $\sin(2\pi f_c t)$  est égal à

$$\begin{aligned}
 p(t) &= s(t) \sin(2\pi f_c t) \\
 &= A_c \cos(2\pi f_c t) \sin(2\pi f_c t) - A_c \theta(t) \sin^2(2\pi f_c t) \\
 &= \frac{A_c}{2} \sin(4\pi f_c t) - \frac{A_c}{2} \theta(t) (1 - \cos(4\pi f_c t)) \\
 &= \frac{A_c}{2} \sin(4\pi f_c t) - \frac{A_c}{2} \theta(t) + \frac{A_c}{2} \theta(t) \cos(4\pi f_c t)
 \end{aligned}$$

Le premier terme de cette expression est une sinusoïde pure à la fréquence  $2f_c$ . Le second terme correspond au signal utile  $\theta(t)$  au facteur  $A_c/2$  près. Ce terme nous intéresse tout particulièrement car c'est hors de celui-ci que nous allons extraire  $m(t)$ . Par contre, le troisième terme est un signal haute fréquence, qui correspond à la modulation DSB-SC de  $\theta(t)$  autour de la fréquence porteuse  $2f_c$ . Il constitue donc, avec la sinusoïde pure, un terme parasite qu'il nous faut éliminer. Ceci est fait par l'application du filtre passe-bas qui nous fournit donc le signal suivant

$$r(t) = -\frac{A_c}{2} \theta(t) = -\frac{A_c}{2} 2\pi k_f \int_{-\infty}^t m(\tau) d\tau$$

Reste maintenant le déivateur qui fournit le signal

$$d(t) = \frac{dr(t)}{dt} = -\pi k_f A_c m(t)$$

Le signal modulant se récupère finalement par la relation

$$m(t) = -\frac{d(t)}{\pi k_f A_c}$$

En guise d'illustration, la démodulation du signal modulé NBFM de la figure 2.28 est présentée à la figure 2.31.

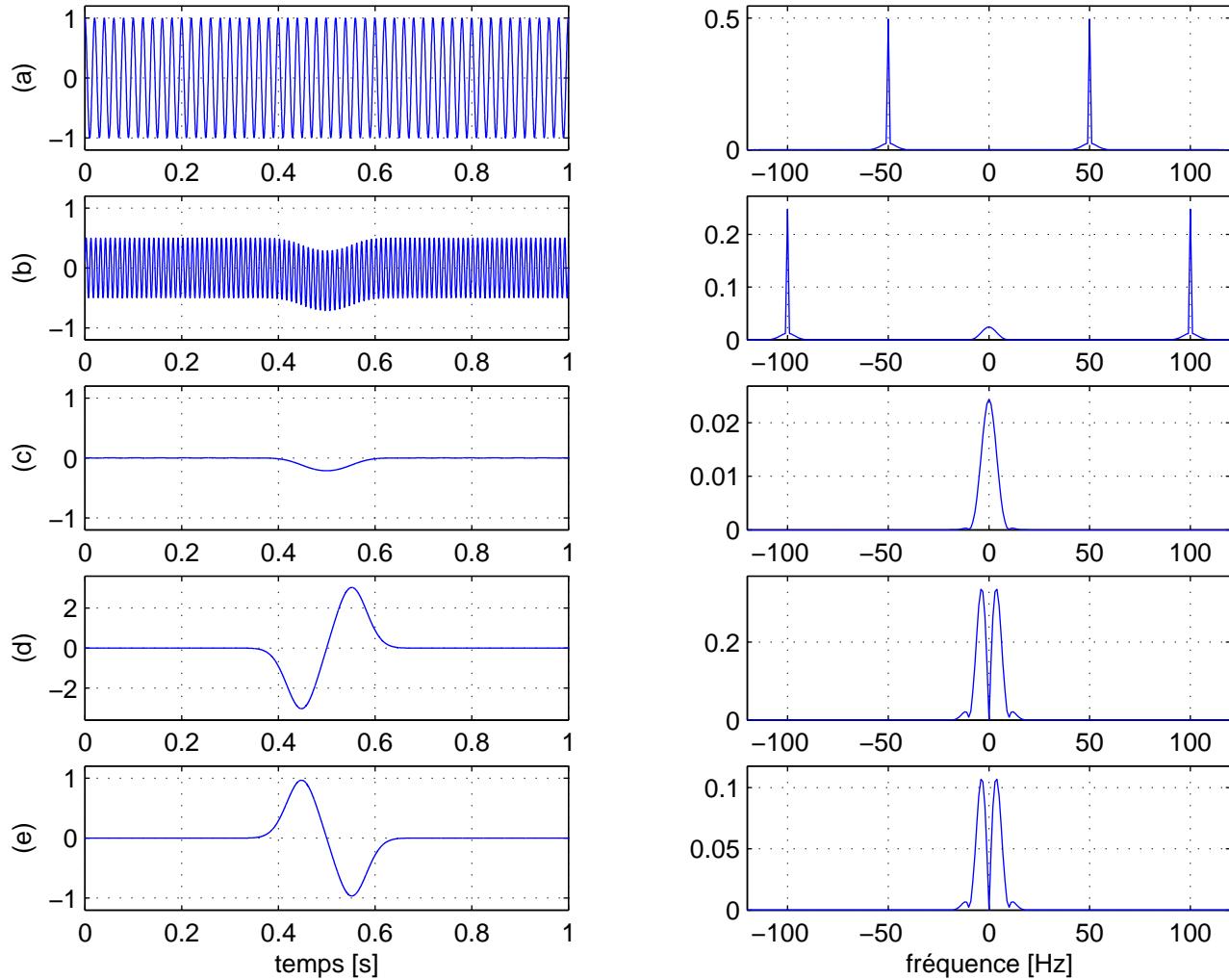


FIGURE 2.31 – Démodulation du signal NBFM ( $A_c = 1$ ,  $f_c = 50$  [Hz],  $k_f = 1$ ) de la figure 2.28 : (a) Signal modulé  $s(t)$ . (b) Signal  $p(t)$ . (c) Signal  $r(t)$ . (d) Signal  $d(t)$ . (e) Signal  $m(t)$  récupéré via la relation  $m(t) = -d(t)/\pi$ .

### Cas d'un signal modulant cosinusoidal

Considérons à présent un signal modulant cosinusoidal

$$m(t) = A_m \cos(2\pi f_m t)$$

où  $A_m$  et  $f_m$  sont respectivement l'amplitude et la fréquence du signal modulant. La fréquence instantanée du signal modulé vaut alors

$$\begin{aligned} f_i(t) &= f_c + k_f A_m \cos(2\pi f_m t) \\ &= f_c + \Delta f \cos(2\pi f_m t) \end{aligned}$$

où  $\Delta f = k_f A_m$  est l'excursion de fréquence, notion déjà définie plus haut. On voit que celle-ci est proportionnelle à l'amplitude du signal modulant mais qu'elle ne dépend pas de sa fréquence.

L'angle instantané du signal modulé peut à présent se calculer facilement

$$\begin{aligned}\Phi_i(t) &= 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \\ &= 2\pi f_c t + \frac{2\pi k_f A_m}{2\pi f_m} \sin(2\pi f_m t) \\ &= 2\pi f_c t + \frac{\Delta f}{f_m} \sin(2\pi f_m t) \\ &= 2\pi f_c t + \beta \sin(2\pi f_m t)\end{aligned}$$

où  $\beta$ , l'indice de modulation, est égal à

$$\boxed{\beta = \frac{\Delta f}{f_m}}$$

Finalement, le signal modulé peut s'écrire

$$\boxed{s(t) = A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t))}$$

sans la moindre approximation. L'approximation réalisée plus haut se traduit ici par

$$\beta \ll 1$$

Dans ce cas, le signal modulé vaut, en utilisant l'approximation,

$$\begin{aligned}s(t) &= A_c \cos(2\pi f_c t) + \beta A_c \sin(2\pi f_m t) \sin(2\pi f_c t) \\ &= A_c \cos(2\pi f_c t) + \frac{\beta A_c}{2} (\cos(2\pi(f_c - f_m)t) - \cos(2\pi(f_c + f_m)t))\end{aligned}$$

Le spectre du signal NBFM d'un signal modulant cosinusoidal se compose donc de

- une raie de DIRAC située aux fréquences  $f = \pm f_c$ , il s'agit de la porteuse ;
- deux raies de DIRAC situées de part et d'autre de la fréquence  $f_c$ , en  $f_c + f_m$  et  $f_c - f_m$ .

La bande passante d'un tel signal est donc égale à  $2f_m$ , qui est bien égal à deux fois la bande de base du signal modulant.

#### 2.4.4 Modulation FM à large bande

Le signal modulé

$$s(t) = A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t))$$

peut encore s'écrire, sans approximation,

$$s(t) = A_c \cos(2\pi f_c t) \cos(\beta \sin(2\pi f_m t)) - A_c \sin(2\pi f_c t) \sin(\beta \sin(2\pi f_m t))$$

En utilisant les propriétés des fonctions de BESSEL, on pourrait montrer que

$$\boxed{s(t) = A_c \sum_{n=-\infty}^{+\infty} J_n(\beta) \cos(2\pi(f_c + n f_m)t)}$$

où  $J_n(\beta)$  sont les fonctions de BESSEL de première espèce, et ce toujours sans aucune approximation. Pour information, la figure 2.32 présente les fonctions de BESSEL de première espèce  $J_n(x)$  pour  $n = 1, 2, 3$ .

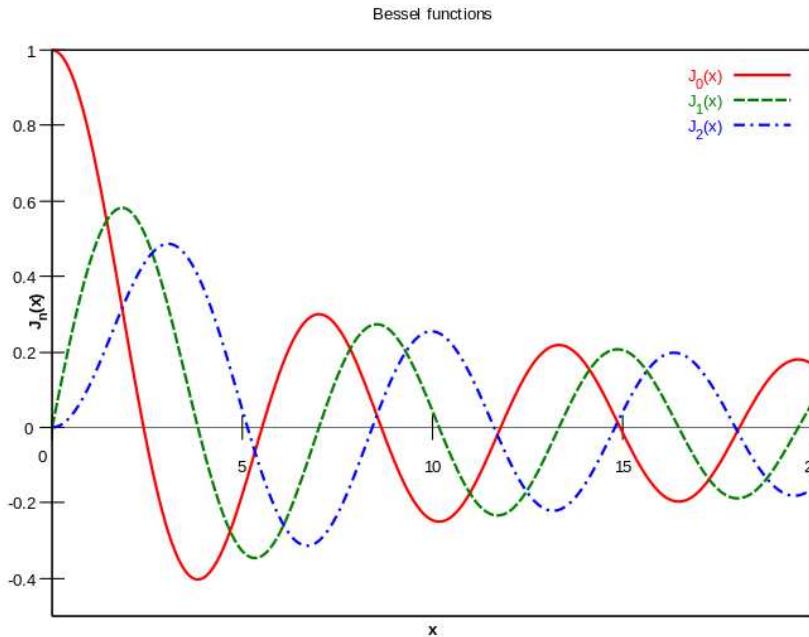


FIGURE 2.32 – Fonctions de Bessel de première espèce  $J_n(x)$ .

### Spectre et bande passante

Le signal modulé peut être vu comme une somme contenant une infinité de cosinusoïdes. Il est à présent possible de calculer la transformée de FOURIER :

$$S(f) = \frac{A_c}{2} \sum_{n=-\infty}^{+\infty} J_n(\beta) (\delta(f - (f_c + n f_m)) + \delta(f + (f_c + n f_m)))$$

Le spectre du signal modulé est donc composé d'une infinité de raies de DIRAC situées de part en d'autre de la fréquence porteuse  $f_c$ , séparées par des intervalles de largeur égale à  $f_m$ , et dont les amplitudes sont directement liées à l'indice de modulation  $\beta$ . Des exemples de spectres de signaux modulés FM sont présentés aux figures 2.33 et 2.34.

Étant donné l'expression de la transformée de FOURIER d'un signal modulé FM à large bande, la bande passante d'un tel signal modulé est *théoriquement infinie*. Il apparaît néanmoins que la puissance est principalement véhiculée par la porteuse et quelques harmoniques autour de cette fréquence. Dès lors, en pratique, la bande passante peut être considérée comme finie et sa valeur a été déterminée empiriquement. La bande passante d'un signal FM a donc été définie de telle sorte qu'elle contienne 98% de la puissance du signal. Ceci a conduit, après un certain nombre de calculs que nous passerons sous silence, à l'expression suivante

$$W_{FM} = 2(\Delta f + f_m) = 2 f_m (\beta + 1) \quad (2.21)$$

Pour une modulation FM à bande étroite ( $\beta \ll 1$ ), cette relation se simplifie en

$$W_{NBFM} = 2f_m$$

relation que nous avons déjà énoncée plus haut.

**Exemple.** Prenons comme exemple le signal FM dont le spectre est présenté à la figure 2.34d. Nous avons donc  $f_m = 20 [Hz]$  et  $\beta = 5$ . La bande passante est donc égale à

$$W_{FM} = 2 \cdot 20 (5 + 1) = 240 [Hz]$$

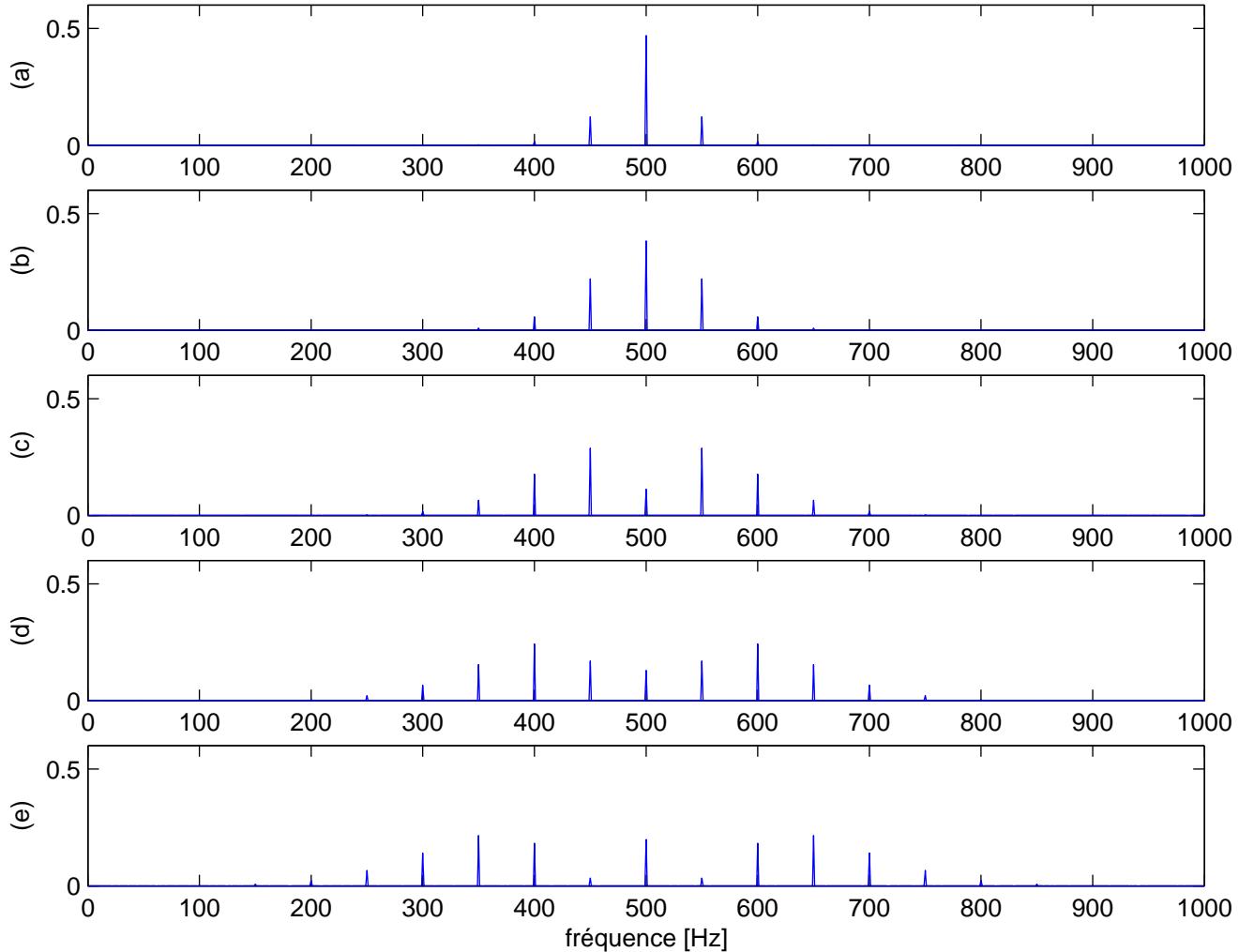


FIGURE 2.33 – Spectre d'un signal FM ( $A_c = 1$ ,  $f_c = 500$  [Hz]) pour lequel  $f_m = 50$  [Hz] est maintenu constant : (a)  $\Delta f = 25$  [Hz] ( $\beta = 0,5$ ). (b)  $\Delta f = 50$  [Hz] ( $\beta = 1$ ). (c)  $\Delta f = 100$  [Hz] ( $\beta = 2$ ). (d)  $\Delta f = 150$  [Hz] ( $\beta = 3$ ). (e)  $\Delta f = 200$  [Hz] ( $\beta = 4$ ).

La formule (2.21) est valable pour un signal modulant cosinusoidal de fréquence  $f_m$ . Qu'en est-il à présent pour un signal  $m(t)$  quelconque de bande de base  $W$ ? La formule précédente peut alors être généralisée en remplaçant  $f_m$  par  $W$  dans le calcul de l'indice de modulation  $\beta$

$$\boxed{\beta = \frac{\Delta f}{W}}$$

La formule (2.21) reste alors valable pour un signal modulant quelconque et porte le nom de règle de CARSON.

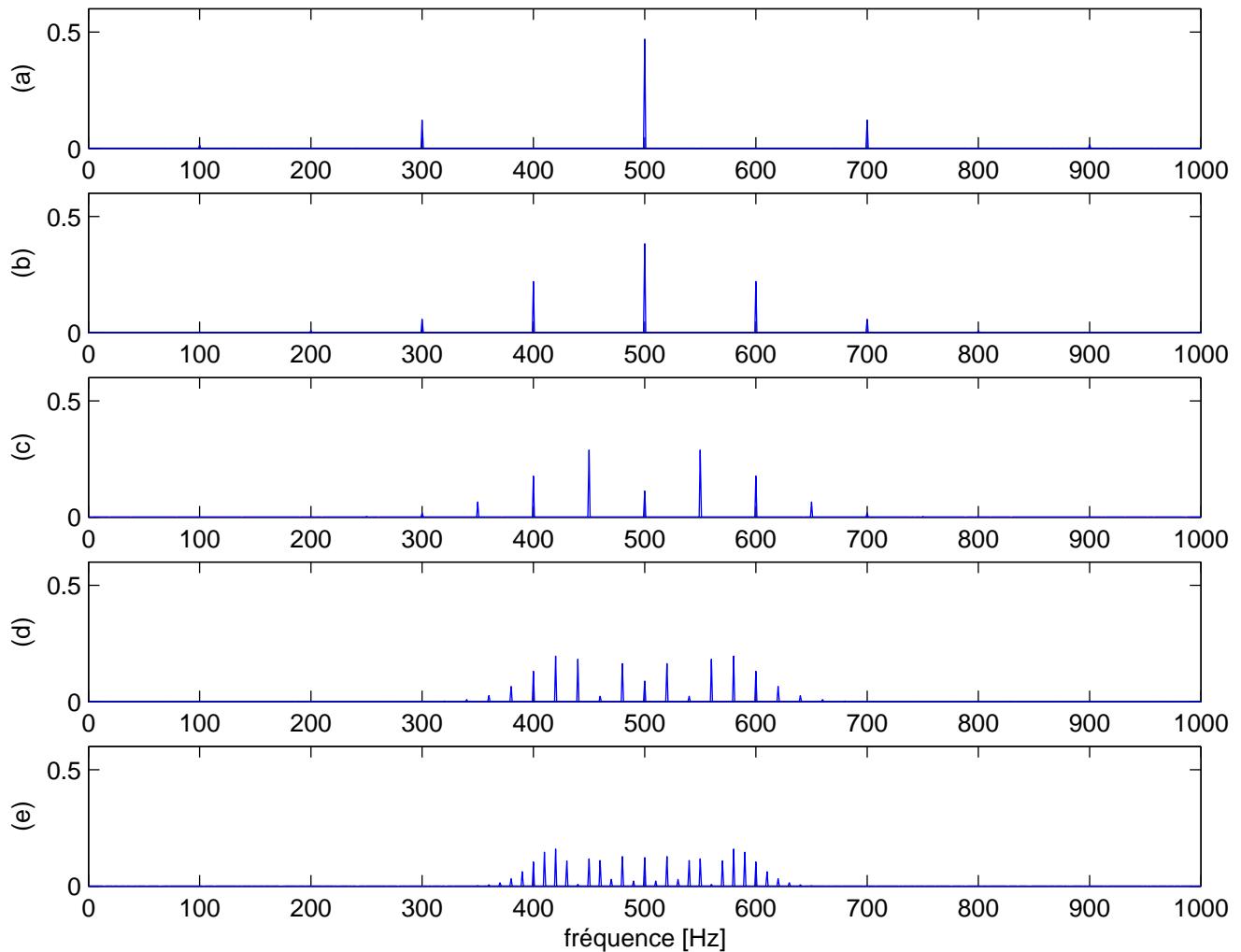


FIGURE 2.34 – Spectre d'un signal FM ( $A_c = 1$ ,  $f_c = 500$  [Hz]) pour lequel  $\Delta f = 100$  [Hz] est maintenu constant : (a)  $f_m = 200$  [Hz] ( $\beta = 0,5$ ). (b)  $f_m = 100$  [Hz] ( $\beta = 1$ ). (c)  $f_m = 50$  [Hz] ( $\beta = 2$ ). (d)  $f_m = 20$  [Hz] ( $\beta = 5$ ). (e)  $f_m = 10$  [Hz] ( $\beta = 10$ ).

**Exemple.** On désire connaître le nombre de stations radio FM que l'on peut placer dans la bande de fréquence allant de 88 [MHz] à 108 [MHz] (soit 20 [MHz] de largeur totale) sachant que la bande de base des signaux utiles est égale à  $W = 15$  [kHz] et que l'excursion de fréquence  $\Delta f$  est limité à 75 [kHz]. L'indice de modulation est égal à

$$\beta = \frac{\Delta f}{W} = \frac{75}{15} = 5$$

La bande passante d'un signal FM est donc égale à

$$W_{FM} = 2 \cdot 15 (1 + 5) = 180 \text{ [kHz]}$$

Dès lors, le nombre de stations maximum est égal à

$$\frac{20.000}{180} = 111,111 \approx 111 \text{ stations}$$

## Démodulation

Il existe de nombreux démodulateur de signaux FM à large bande. Nous en présenterons un ici. Celui-ci est présenté à la figure 2.35. Étudions à présent son fonctionnement.

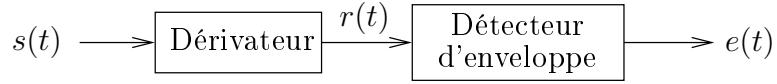


FIGURE 2.35 – Démodulateur FM à large bande.

Le signal modulé

$$s(t) = A_c \cos \left( 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) \quad (2.22)$$

appliquée à l'entrée du déivateur fournit, en sortie, le signal

$$\begin{aligned} r(t) &= -A_c \left( 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right)' \sin \left( 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) \\ &= -A_c (2\pi f_c + 2\pi k_f m(t)) \sin \left( 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) \end{aligned}$$

dont l'enveloppe est une fonction linéaire du signal modulant  $m(t)$ . Le détecteur d'enveloppe fournit dès lors le signal

$$e(t) = A_c (2\pi f_c + 2\pi k_f m(t))$$

à condition que  $k_f m(t) > f_c$ . La récupération de  $m(t)$  peut alors s'obtenir par

$$m(t) = \frac{\frac{e(t)}{A_c} - 2\pi f_c}{2\pi k_f}$$

La figure 2.36 illustre la démodulation d'un signal FM à large bande.

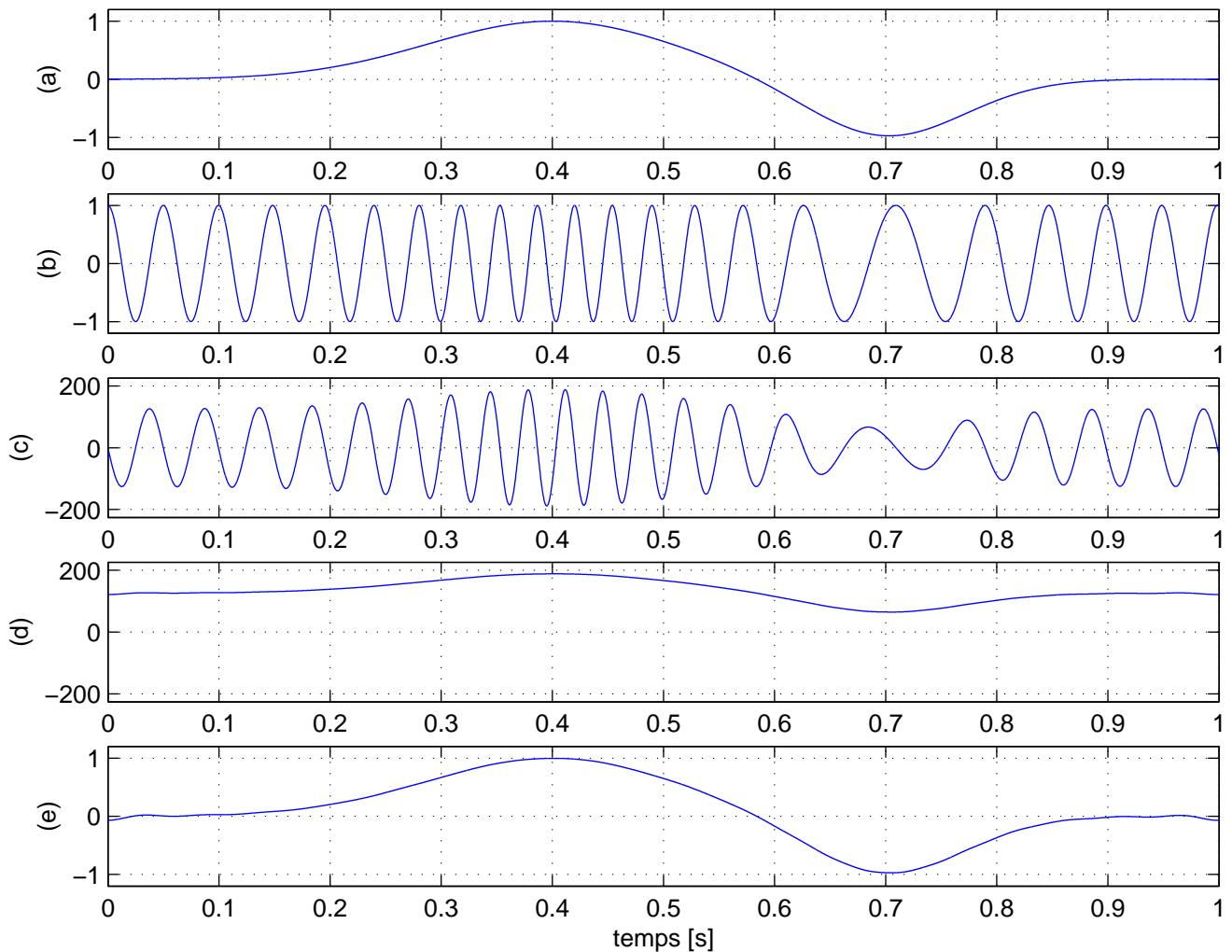


FIGURE 2.36 – Démodulation d'un signal FM à large bande : (a) Signal modulant  $m(t)$ . (b) Signal modulé FM ( $A_c = 1$ ,  $f_c = 20$  [Hz],  $k_f = 10$ ). (c) Signal dérivé  $r(t)$ . (d) Sortie du détecteur d'enveloppe  $e(t)$ . (e) Signal modulant  $m(t)$  récupéré via la relation  $m(t) = (e(t) - 40\pi)/(20\pi)$ .

## 2.5 Multiplexage en fréquence (FDM)

L'utilisation de certains supports de transmission exige un partage adéquat des ressources fréquentielles. La technique réalisant ce partage est appelée multiplexage en fréquence ou “Frequency Division Multiplexing” (FDM). La figure 2.37 illustre le principe. On dispose d'une série de signaux  $m_i(t)$  en bande de base à transmettre **simultanément**. Au moyen de modulateurs accordés à des fréquences porteuses spécifiques, le spectre de chaque signal est déplacé le long de l'axe des fréquences et ajouté au signal **multiplex** de manière à couvrir une certaine plage fréquentielle, tout en évitant un chevauchement en ménageant des **bandes de garde** entre les signaux.

Le signal multiplexé est transmis au récepteur qui doit extraire un à un tous les signaux au moyen de démodulateurs adéquats accordés aux mêmes fréquences qu'à l'émission. Voir figure 2.38.

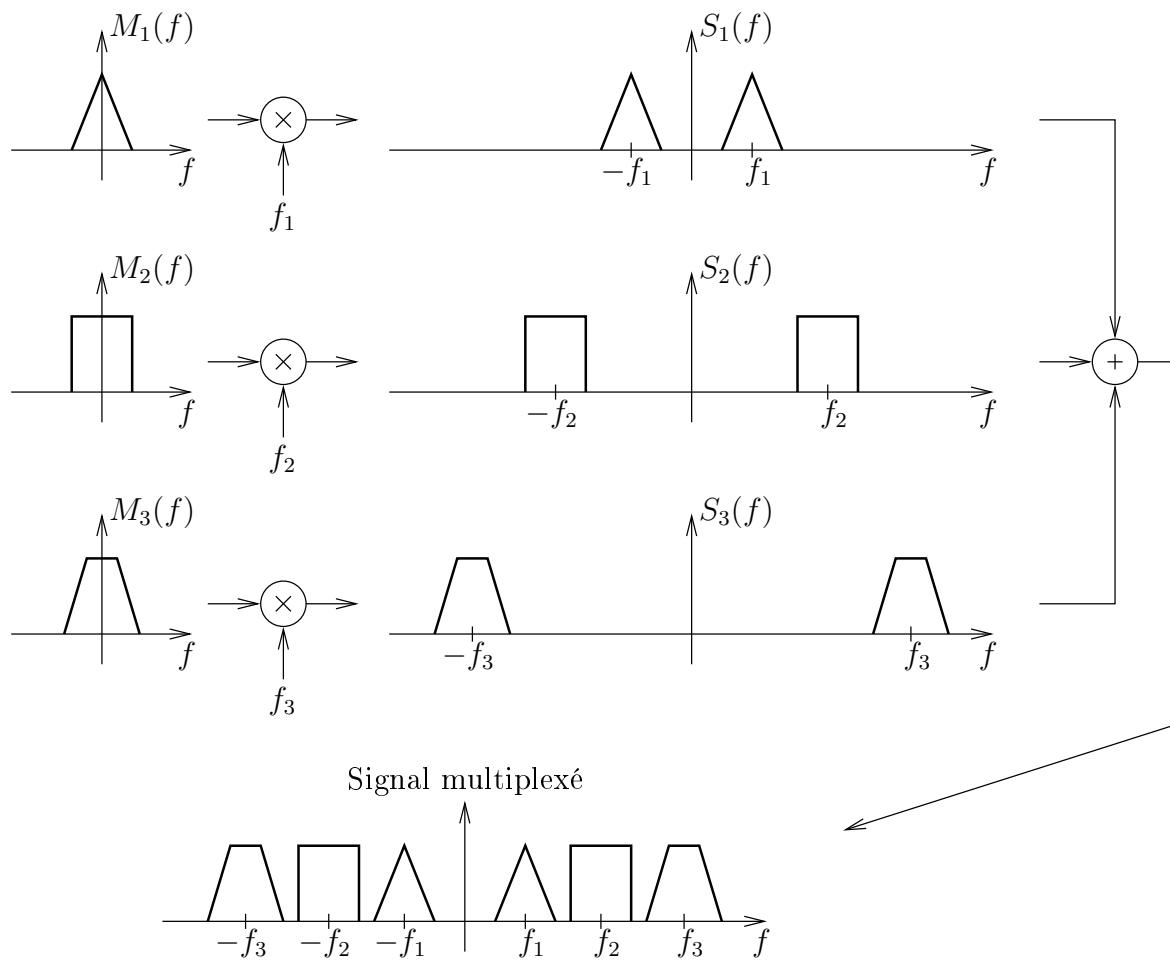


FIGURE 2.37 – Principe du multiplexage en fréquence (on s'est limité pour cet exemple à la modulation DSB-SC).

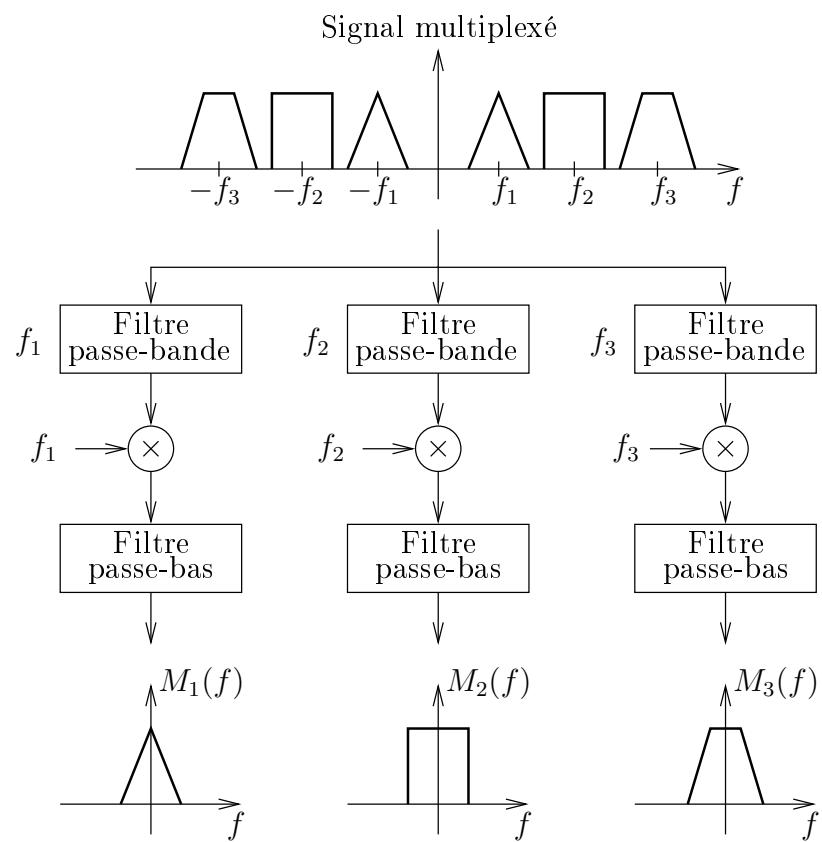


FIGURE 2.38 – Principe du démultiplexage fréquentiel.



## Deuxième partie

### Codage de source



# Chapitre 3

## Introduction à la théorie de l'information, du codage et de la compression

Ce chapitre a pour but de d'introduire les notions de la théorie de l'information et du codage. Il aborde également les techniques de compression de données les plus courantes, comme le codage de HUFFMAN ou encore l'algorithme de LEMPEL-ZIV qui est à la base de la compression Zip. Néanmoins, la théorie de l'information concerne également le domaine de la transmission de données.

### 3.1 Introduction

L'information que nous rencontrons dans la nature est par nature analogique. Nous pouvons citer la musique (le son), les images, les vidéos, ... Mais d'autres formes d'information existent. Par exemple, un texte (issus d'un livre par exemple) peut être vu comme une information discrète (on dit également numérique) car il est constitué d'éléments distincts que sont les lettres utilisées. Nous pouvons donc classer les sources d'informations en deux grandes catégories :

- les **sources analogiques** : son, images, vidéos, ... En gros, tout ce qui peut être représenté par une fonction mathématique de une (son) ou plusieurs (image et video) variables.
- les **sources discrètes** : textes mais également données binaires stockées sur ordinateur (CD, DVD, disques durs, ...). En gros, tout ce qui peut être représenté comme un flot de **symboles**. Dans le cas des textes, les symboles sont les lettres de l'alphabet utilisé tandis que dans le cas de données informatiques (les fichiers), les symboles utilisés sont les bits 1 et 0.

Toutefois, aujourd'hui, les données analogiques sont la plupart du temps stockées et transmises sous forme numérique (images JPEG, video MPEG, audio MP3, ...). D'où l'importance de l'étude des sources discrètes d'informations.

#### 3.1.1 Modèle mathématique d'une source

Une source transmet de l'information à un récepteur, celui-ci ne connaissant pas, en général, l'information qu'il va recevoir. Considérons un exemple simple. Deux personnes discutent. L'un d'eux prononce un mot commençant par "cha". Celui qui écoute peut avoir une idée de la fin du mot : "riot", "teau", ... mais certainement pas "gnon". Nous pouvons en déduire que l'information est en général **imprédictible** ou partiellement imprédictible. Nous pouvons même aller plus loin. Qu'est-ce qui est intéressant pour la personne qui écoute ? Sûrement pas une suite de mots

qu'il connaît à coup sûr. Pour lui, ce n'est plus une information car il connaît déjà la réponse ; il est d'autant plus intéressé qu'il ne peut pas prédire la suite du message.

Nous arrivons aux constatations suivantes :

1. Une source d'informations émet en général un message **aléatoire** (on dit également non déterministe).
2. Ce qui rend une information intéressante est son caractère **imprédictible**. Une information est ainsi d'autant plus riche qu'elle est **peu probable**.

Ces constatations nous mènent à modéliser une source de données sous la forme d'une variable aléatoire. Dans le contexte qui nous concerne, il s'agira d'une **variable aléatoire discrète**, car ne pouvant prendre qu'un nombre fini (discret) de valeurs, c'est-à-dire les symboles émis par cette source.

### 3.1.2 Source discrète sans mémoire

Donc, dans la suite, nous allons modéliser une source par une variable aléatoire discrète  $X$ . Cette source dispose d'un alphabet constitués d'éléments ou symboles  $\{x_1, x_2, x_3, \dots, x_K\}$  où  $K$  est la taille de l'alphabet, nous dirons qu'il s'agit d'un alphabet  $K$ -aire. Ces symboles sont associés (ou émis) pour constituer un message. Emettre un message consiste donc à émettre une succession de symboles appartenant à une source.

Chaque symbole  $x_k$  de l'alphabet a une certaine probabilité d'utilisation notée

$$p_k = P(X = x_k)$$

pour  $k = 1, 2, 3, \dots, K$ . Bien sûr, étant donné la théorie des probabilités, nous avons

$$\sum_{k=1}^K p_k = 1$$

Dans la suite de cet exposé, nous parlerons surtout d'une catégorie simplifiée de source discrète. Il s'agit des sources discrète sans mémoire. Si nous notons  $p(x_i, x_j)$  la probabilité qu'une source discrète émette successivement le symbole  $x_i$  puis le symbole  $x_j$ , nous avons la définition suivante.

**Définition (Source discrète sans mémoire).** Une **source discrète sans mémoire** est une source discrète pour laquelle la probabilité d'émission d'un symbole est indépendante de ce qui a été émis avant ou sera émis après, ce qui peut se traduire par

$$p(x_i, x_j) = P(X = x_i)P(X = x_j) = p_i p_j$$

pour  $i, j = 1, 2, \dots, K$ .

## 3.2 Mesure de l'information

Nous allons à présent donner une définition précise de la notion d'information. Celle-ci est due à C. SHANNON [1916-2001], ingénieur en génie électrique et mathématicien américain. Il est considéré comme l'un des pères de la théorie de l'information.

### 3.2.1 Quantité d'information

Avant de donner la définition attendue, nous émettons les remarques suivantes, déjà citées ou non :

- La quantité d'information d'un symbole est d'autant plus grande que celui-ci est peu probable.
- La quantité d'information de deux symboles successifs émis par une source discrète sans mémoire doit être égale à la somme de leur quantités d'information respectives.

Ainsi, la quantité d'information, que nous noterons  $I$ , est une fonction qui doit avoir les propriétés suivantes :

1.  $I$  est une fonction continue de la probabilité  $p_k$ .
2.  $I(p_k)$  croît si  $p_k$  décroît. Donc,  $I$  est une fonction décroissante.
3.  $I(p_i, p_j) = I(p_i) + I(p_j)$  où  $I(p_i, p_j)$  est la quantité d'information des deux symboles successifs  $x_i$  et  $x_j$ .
4. Un symbole certain, s'il existe, possède une quantité d'information nulle :  $I(p_k) = 0$  si  $p_k = P(X = x_k) = 1$ .

Une fonction mathématique bien connue remplit les conditions 1, 3 et 4. Il s'agit de  $\log(p_k)$ . Pour respecter la condition 2, il suffit de prendre  $-\log(p_k) = \log\left(\frac{1}{p_k}\right)$ .

**Définition (Quantité d'information).** La quantité d'information d'un symbole  $x_k$  de probabilité  $p_k$  est définie par

$$I(x_k) = -\log(p_k) = \log\left(\frac{1}{p_k}\right) \quad (3.1)$$

Il reste néanmoins encore une inconnue : la base de la fonction logarithme utilisée. Plusieurs valeurs de cette base ont été proposées mais c'est la base initialement proposée par SHANNON, c'est-à-dire 2, qui est la plus utilisée. Dans ce cas, l'unité de la quantité d'information est le bit. Cette unité a été rebaptisée le shannon en hommage à son inventeur mais cette appellation reste peu usitée. Par la suite, sauf précision du contraire, c'est la base 2 et donc cette unité (le bit) que nous utiliserons.

**Exemple.** Considérons une source discrète sans mémoire ayant deux symboles (on parle encore de source binaire) équiprobables  $x_1 = 1$  et  $x_2 = 0$  dans son alphabet. Les symboles étant équiprobables, nous avons

$$p_1 = p_2 = \frac{1}{2}$$

L'information de chaque symbole est donc égale à

$$I(x_1) = I(x_2) = -\log_2\left(\frac{1}{2}\right) = 1 \text{ bit}$$

Le symbole d'une source binaire, qui est un bit, possède donc une quantité d'information de 1 bit, d'où l'équivalence. Par la suite, sauf précision contraire, nous omettrons d'écrire la base 2 de la fonction logarithme.

### 3.2.2 Entropie d'une source

Nous disposons à présent de la définition de la quantité d'information pour chaque symbole de la source. Mais qu'en est-il de la source en général ? Il est courant de considérer la quantité d'information moyenne de chaque symbole de la source. Pour une source, notée  $S$ , modélisée par une variable aléatoire  $X$ , la grandeur  $I(X)$  est également une variable aléatoire dont l'espérance mathématique ("la valeur la plus probable") est son espérance mathématique. Cette valeur constitue la moyenne recherchée et porte le nom d'entropie de la source.

**Définition (Entropie d'une source).** L'entropie d'une source  $S$ , présentant  $K$  symboles  $x_k$  ( $k = 1, 2, \dots, K$ ) de probabilité  $p_k$ , est notée  $H(S)$  (ou encore  $H(X)$ ), et est définie par

$$H(S) = E[I(X)] = \sum_{k=1}^K p_k I(x_k) = \sum_{k=1}^K p_k \log\left(\frac{1}{p_k}\right) = -\sum_{k=1}^K p_k \log p_k \quad (3.2)$$

Son unité est le bits/symbole.

Cette définition n'est pas sans rappeler la définition de l'entropie d'un système thermodynamique ou celle d'un ensemble de molécules en théorie cinétique des gaz, dans laquelle l'entropie représente le désordre (incertitude ?) dans le gaz.

**Exemple.** Considérons une source binaire de deux symboles  $x_1 = 1$  et  $x_2 = 0$  de probabilités respectives  $p_1 = p$  et  $p_2 = 1 - p$ . L'entropie de cette source est donnée par

$$H(S) = p \log\left(\frac{1}{p}\right) + (1-p) \log\left(\frac{1}{1-p}\right) = -p \log p - (1-p) \log(1-p)$$

et est donc une fonction de la probabilité  $p$ . Cette fonction est illustrée à la figure (3.1). Cette entropie, c'est-à-dire l'information moyenne de la source, est maximale lorsque  $p = \frac{1}{2}$ , c'est-à-dire lorsque les deux symboles sont équiprobables, ou encore lorsque la source est la plus imprédictible possible. Par contre, lorsque  $p = 0$  ou  $p = 1$ , la source est totalement prédictible et son entropie est nulle.

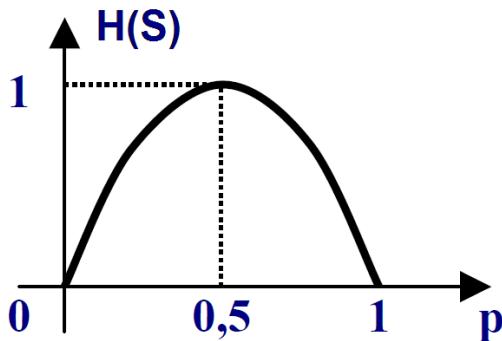


FIGURE 3.1 – Entropie d'une source binaire.

Cet exemple est un cas particulier d'une source de  $K$  symboles. On peut démontrer (voir annexe 3.6.1) que son entropie est maximale lorsque tous les symboles sont équiprobables, c'est-à-dire lorsque  $p_k = \frac{1}{K}$  pour tout  $k = 1, 2, \dots, K$ . L'entropie devient alors

$$H(S) = \sum_{k=1}^K \frac{1}{K} \log K = \log K$$

Ce qui permet d'énoncer le théorème suivant.

**Théorème (Maximum de l'entropie).** L'entropie d'une source discrète sans mémoire ayant un alphabet de  $K$  symboles respecte la relation suivante :

$$0 \leq H(S) \leq \log K \quad (3.3)$$

### 3.2.3 Entropie jointe entre deux sources

Cette notion permet de mesurer le degré de similitude entre deux sources. Comme application, nous mettrons en évidence (voir section 3.2.7) l'effet d'une probabilité d'erreur de transmission lors une communication binaire. En effet, lors d'une transmission numérique sur un canal de communication, il est possible de commettre une erreur lors de la réception à cause d'une bande passante trop limitée du canal ou de la présence de bruit.

Considérons deux sources discrètes  $X$  et  $Y$ , sans mémoire, et présentant les alphabets respectifs  $\{x_1, x_2, \dots, x_N\}$  et  $\{y_1, y_2, \dots, y_M\}$ . Nous notons  $p_{XY}(x_i, y_j) = P(X = x_i \text{ et } Y = y_j)$  la probabilité que les sources  $X$  et  $Y$  émettent respectivement les symboles  $x_i$  et  $y_j$  à un moment donné. Ceci permet de définir la quantité d'information jointe due aux deux symboles :

$$I(x_i, y_j) = \log \left( \frac{1}{p_{XY}(x_i, y_j)} \right)$$

**Définition (Entropie jointe).** L'entropie jointe des deux sources  $X$  et  $Y$  est la quantité d'information moyenne jointe entre deux symboles de la source :

$$H(X, Y) = \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{1}{p_{XY}(x_i, y_j)} \right) \quad (3.4)$$

### Cas particulier de deux sources indépendantes

Si les deux sources  $X$  et  $Y$  sont indépendantes, l'émission du symbole  $x_i$  par la source  $X$  n'influence en aucune façon le symbole  $y_j$  par la source  $Y$ . Mathématiquement, cela signifie que les variables aléatoires  $X$  et  $Y$  sont indépendantes et nous pouvons écrire

$$p_{XY}(x_i, y_j) = P(X = x_i) P(Y = y_j) = p_X(x_i) p_Y(y_j)$$

où nous avons posé  $p_X(x) = P(X = x)$  et  $p_Y(y) = P(Y = y)$  pour simplifier les notations. Dans ce cas, l'entropie jointe (3.4) des deux sources peut s'écrire

$$\begin{aligned} H(X, Y) &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{1}{p_{XY}(x_i, y_j)} \right) \\ &= \sum_{i=1}^N \sum_{j=1}^M p_X(x_i) p_Y(y_j) \log \frac{1}{p_X(x_i) p_Y(y_j)} \\ &= \sum_{i=1}^N \sum_{j=1}^M p_X(x_i) p_Y(y_j) \log \frac{1}{p_X(x_i)} + \sum_{i=1}^N \sum_{j=1}^M p_X(x_i) p_Y(y_j) \log \frac{1}{p_Y(y_j)} \\ &= \sum_{j=1}^M p_Y(y_j) \sum_{i=1}^N p_X(x_i) \log \frac{1}{p_X(x_i)} + \sum_{i=1}^N p_X(x_i) \sum_{j=1}^M p_Y(y_j) \log \frac{1}{p_Y(y_j)} \\ &= \sum_{i=1}^N p_X(x_i) \log \frac{1}{p_X(x_i)} + \sum_{j=1}^M p_Y(y_j) \log \frac{1}{p_Y(y_j)} \end{aligned}$$

Ce qui peut finalement s'écrire

$$\boxed{H(X, Y) = H(X) + H(Y)} \quad (3.5)$$

Dès lors, si deux sources sont indépendantes, l'information moyenne jointe des deux sources est la somme des informations moyennes respectives de chaque source.

### 3.2.4 Quantité d'information mutuelle

Dans le cas de sources dépendantes, la formule (3.5) n'est plus valable. Mathématiquement, les variables aléatoires  $X$  et  $Y$  ne sont plus indépendantes. Les deux sources émettent donc des messages ayant une certaine similitude, on dit qu'il y a de la **redondance** entre les deux sources. L'information jointe des deux sources doit donc être inférieure à la somme des informations moyennes respectives de chaque source. Voyons cela.

$$\begin{aligned}
 H(X, Y) &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{1}{p_{XY}(x_i, y_j)} \right) \\
 &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_X(x_i) p_Y(y_j)}{p_{XY}(x_i, y_j)} \right) + \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{1}{p_X(x_i) p_Y(y_j)} \right) \\
 &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_X(x_i) p_Y(y_j)}{p_{XY}(x_i, y_j)} \right) \\
 &\quad + \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \frac{1}{p_X(x_i)} + \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \frac{1}{p_Y(y_j)} \\
 &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_X(x_i) p_Y(y_j)}{p_{XY}(x_i, y_j)} \right) \\
 &\quad + \sum_{i=1}^N \log \frac{1}{p_X(x_i)} \sum_{j=1}^M p_{XY}(x_i, y_j) + \sum_{j=1}^M \log \frac{1}{p_Y(y_j)} \sum_{i=1}^N p_{XY}(x_i, y_j) \\
 &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_X(x_i) p_Y(y_j)}{p_{XY}(x_i, y_j)} \right) \\
 &\quad + \sum_{i=1}^N \log \frac{1}{p_X(x_i)} p_X(x_i) + \sum_{j=1}^M \log \frac{1}{p_Y(y_j)} p_Y(y_j) \\
 &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_X(x_i) p_Y(y_j)}{p_{XY}(x_i, y_j)} \right) + H(X) + H(Y)
 \end{aligned}$$

où nous avons utilisé le fait que

$$\sum_{j=1}^M p_{XY}(x_i, y_j) = p_X(x_i) \text{ et } \sum_{i=1}^N p_{XY}(x_i, y_j) = p_Y(y_j)$$

Dans ce résultat, le premier terme est un terme supplémentaire par rapport aux cas des sources indépendantes. C'est le terme d'information mutuelle. Sachant que  $p_{XY}(x_i, y_i) \geq p_X(x_i) p_Y(y_i)$ <sup>1</sup>, ce terme est négatif. Nous pouvons alors introduire la définition suivante.

**Définition (Quantité d'information mutuelle).** La quantité d'information mutuelle  $I(X, Y)$  entre deux sources discrètes  $X$  et  $Y$ , sans mémoire, est définie par

$$I(X, Y) = \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_{XY}(x_i, y_j)}{p_X(x_i) p_Y(y_j)} \right) \quad (3.6)$$

1. Ceci provient de la théorie des probabilités. Si  $A$  représente l'occurrence d'un certain événement et  $B$  celle d'un autre événement, alors  $P(A \text{ et } B) = P(A) + P(B) - P(A \text{ ou } B) \geq P(A) + P(B)$

De là, nous pouvons écrire

$$H(X, Y) = H(X) + H(Y) - I(X, Y) \quad (3.7)$$

Cette dernière expression illustre bien ce que nous pressentions : l'information jointe des deux sources doit donc être inférieure à la somme des informations moyennes respectives de chaque source. La différence entre les deux grandeurs est la quantité d'information mutuelle  $I(X, Y)$  qui représente donc la redondance entre les deux sources. La figure 3.2 donne une interprétation ensembliste de cette constatation, cette représentation porte le nom de diagramme de VENNE. Bien sûr, dans le cas de sources indépendantes, nous retrouvons la relation (3.5).

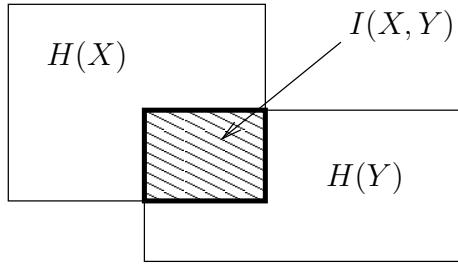


FIGURE 3.2 – Interprétation ensembliste de la quantité d'information mutuelle de deux sources (diagramme de VENNE).

### 3.2.5 Entropie conditionnelle

Afin de compléter nos définitions de la mesure de l'information, il est également possible de définir une entropie conditionnelle, c'est-à-dire la mesure de l'information que nous apporte une source  $X$  étant donné que l'on sait que la source  $Y$  a émis un certain symbole de son alphabet.

Si nous notons  $p_{X|Y}(x_i|y_j) = P(X = x_i|Y = y_j)$  la probabilité conditionnelle que la source  $X$  émette le symbole  $x_i$  alors que la source  $Y$  a émis le symbole  $y_j$ , la quantité d'information conditionnelle du symbole  $x_i$  conditionnellement à  $y_j$  est donnée par

$$I(x_i|y_j) = \log \frac{1}{p_{X|Y}(x_i|y_j)} \quad (3.8)$$

De là, nous déduisons l'entropie conditionnelle (de  $X$  conditionnellement à  $y_j$ ) :

$$H(X|y_j) = \sum_{i=1}^N p_{X|Y}(x_i|y_j) I(x_i|y_j) = \sum_{i=1}^N p_{X|Y}(x_i|y_j) \log \frac{1}{p_{X|Y}(x_i|y_j)} \quad (3.9)$$

Et finalement, en faisant la moyenne pour tous les symboles  $y_j$  que la source  $Y$  peut émettre, nous avons la définition suivante de l'entropie conditionnelle moyenne.

**Définition (Entropie conditionnelle moyenne).** L'entropie conditionnelle moyenne d'une source  $X$  conditionnellement à la source  $Y$  est définie par

$$H(X|Y) = \sum_{j=1}^M p_Y(y_j) H(X|y_j) = \sum_{j=1}^M p_Y(y_j) \sum_{i=1}^N p_{X|Y}(x_i|y_j) \log \frac{1}{p_{X|Y}(x_i|y_j)} \quad (3.10)$$

L'entropie conditionnelle  $H(X|Y)$  peut être vue comme l'information que peut encore nous apporter la source  $X$  alors que l'on dispose déjà de l'information de la source  $Y$ .

### 3.2.6 Lien entre entropie conditionnelle et information mutuelle

La relation que nous allons établir provient de la théorie des probabilités, en particulier de la relation de BAYES qui s'exprime par

$$p_{XY}(x, y) = p_{X|Y}(x|y) p_Y(y) = p_{Y|X}(y|x) p_X(x) \quad (3.11)$$

En repartant de la définition (3.6) de la quantité d'information mutuelle, nous pouvons écrire

$$\begin{aligned} I(X, Y) &= \sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) \log \left( \frac{p_{XY}(x_i, y_j)}{p_X(x_i) p_Y(y_j)} \right) \\ &= \sum_{i=1}^N \sum_{j=1}^M p_{X|Y}(x_i|y_j) p_Y(y_j) \log \left( \frac{p_{X|Y}(x_i|y_j) p_Y(y_j)}{p_X(x_i) p_Y(y_j)} \right) \\ &= \sum_{i=1}^N \sum_{j=1}^M p_{X|Y}(x_i|y_j) p_Y(y_j) \log p_{X|Y}(x_i|y_j) + \sum_{i=1}^N \sum_{j=1}^M p_{X|Y}(x_i|y_j) p_Y(y_j) \log \frac{1}{p_X(x_i)} \\ &= \sum_{j=1}^M p_Y(y_j) \sum_{i=1}^N p_{X|Y}(x_i|y_j) \log p_{X|Y}(x_i|y_j) + \sum_{i=1}^N \sum_{j=1}^M p_{Y|X}(y_j|x_i) p_X(x_i) \log \frac{1}{p_X(x_i)} \\ &= -H(X|Y) + \sum_{i=1}^N p_X(x_i) \log \frac{1}{p_X(x_i)} \sum_{j=1}^M p_{Y|X}(y_j|x_i) \\ &= -H(X|Y) + \sum_{i=1}^N p_X(x_i) \log \frac{1}{p_X(x_i)} \\ &= -H(X|Y) + H(X) \end{aligned}$$

Un résultat semblable peut être établi en permutant le rôle de  $x$  et  $y$ . D'où les deux expressions équivalentes de la quantité d'information mutuelle :

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (3.12)$$

Etant donné la relation (3.7), nous pouvons encore donner l'expression suivante de l'entropie jointe des deux sources  $X$  et  $Y$  :

$$H(X, Y) = H(X) + H(Y) - I(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (3.13)$$

Ceci peut être interprété de la manière suivante. L'information jointe que nous apporte deux sources  $X$  et  $Y$  est donc l'information que la première source  $X$  ( $Y$ ) nous apporte augmentée de l'information que nous apporte encore la source  $Y$  ( $X$ ) étant donné que nous disposons déjà de l'information de la source  $X$  ( $Y$ ). Une représentation graphique (diagramme de VENNE) de ces notions est fournie à la figure 3.3. Pour être complet, nous pouvons encore écrire

$$H(X) = H(X|Y) + I(X, Y) \text{ et } H(Y) = H(Y|X) + I(X, Y) \quad (3.14)$$

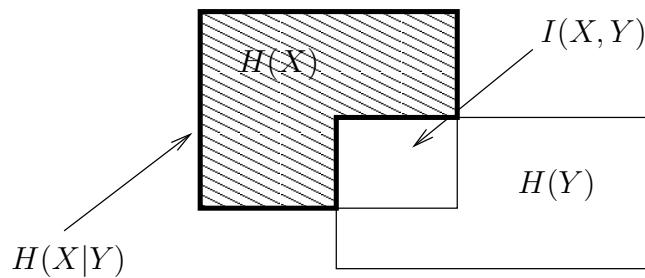


FIGURE 3.3 – Interprétation ensembliste de l'entropie conditionnelle moyenne de deux sources (diagramme de VENNE).

### 3.2.7 Exemple : Canal de communication binaire symétrique

Considérons un canal de communication binaire entre un émetteur et un récepteur. Le but est de récupérer au niveau du récepteur, l'information émise par l'émetteur. Autrement dit, l'alphabet de sortie du canal doit être le même que l'alphabet d'entrée. Dans le cas d'un canal binaire, cet alphabet est composé des bits 1 et 0. Sur un canal réel, il existe une certaine probabilité d'erreur de communication, que nous noterons  $p$ , erreur pouvant provenir d'une bande passante trop limitée, de la présence de bruit, ... La situation est schématisée à la figure 3.4.

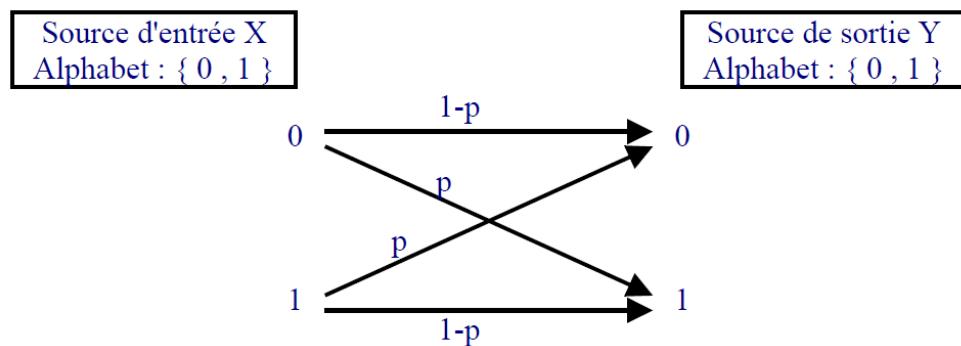


FIGURE 3.4 – Canal de communication binaire symétrique.

Nous sommes donc en présence de deux sources d'informations :

- la source  $X$  qui correspond à l'émetteur. Les symboles sont  $x_1 = 0$  et  $x_2 = 1$ . Pour simplifier, nous ferons l'hypothèse que les deux symboles sont équiprobables :  $p_X(x_1) = p_X(x_2) = \frac{1}{2}$
  - la source  $Y$  qui correspond à l'information fournie par le récepteur. Ses symboles sont  $y_1 = 0$  et  $y_2 = 1$ .

Le canal est dit **symétrique** car la probabilité de commettre une erreur dans le cas de l'émission d'un 1 est identique à la probabilité de commettre une erreur dans le cas de l'émission d'un 0. Cette probabilité est notée  $p$ .

L'entropie de la source  $X$  est simple à calculer car nous disposons des probabilités de chaque symbole. Elle est donnée par

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1 \text{ bit/symbole}$$

Celle de  $Y$  n'est pas calculable directement car nous ne disposons pas des probabilités de ses symboles. Par contre, les propriétés du canal nous fournissent les probabilités conditionnelles suivantes

$$\begin{cases} p_{Y|X}(y_1|x_1) &= 1-p \\ p_{Y|X}(y_2|x_1) &= p \\ p_{Y|X}(y_1|x_2) &= p \\ p_{Y|X}(y_2|x_2) &= 1-p \end{cases}$$

De là, en tenant compte de la relation de BAYES (3.11), nous pouvons calculer les probabilités jointes suivantes

$$\begin{cases} p_{XY}(x_1, y_1) &= \frac{1-p}{2} \\ p_{XY}(x_1, y_2) &= \frac{p}{2} \\ p_{XY}(x_2, y_1) &= \frac{p}{2} \\ p_{XY}(x_2, y_2) &= \frac{1-p}{2} \end{cases}$$

Toujours en tenant compte de la relation de BAYES, nous obtenons

$$p_Y(y_1) = p_Y(y_2) = \frac{1}{2}$$

L'entropie de la source  $Y$  est donc aussi égale à 1 bit/symbole.

Attaquons-nous à présent à l'entropie jointe des deux sources. En utilisant la formule (3.4), nous pouvons écrire

$$\begin{aligned} H(X, Y) &= \frac{1-p}{2} \log \frac{2}{1-p} + \frac{p}{2} \log \frac{2}{p} + \frac{p}{2} \log \frac{2}{p} + \frac{1-p}{2} \log \frac{2}{1-p} \\ &= (1-p) \log \frac{2}{1-p} + p \log \frac{2}{p} \\ &= (1-p)(1 - \log(1-p)) + p(1 - \log p) \\ &= 1 - (1-p) \log(1-p) - p \log p \end{aligned}$$

Nous pouvons à présent facilement calculer l'entropie conditionnelle de la source  $Y$  conditionnellement à  $X$ . Etant donné la relation (3.13), il vient

$$\begin{aligned} H(Y|X) &= H(X, Y) - H(X) \\ &= -(1-p) \log(1-p) - p \log p \end{aligned}$$

Par symétrie,  $H(X|Y)$  a la même expression

$$H(X|Y) = -(1-p) \log(1-p) - p \log p$$

Finalement, la quantité d'information mutuelle peut être calculée en utilisant (3.7) ou (3.12) :

$$I(X, Y) = 1 + (1-p) \log(1-p) + p \log p$$

Nous pouvons à présent tirer quelques conclusions :

- Lorsque  $p = 0$ , c'est-à-dire lorsqu'il n'y a aucune erreur de transmission,  $I(X, Y) = 1$  mais surtout  $H(X|Y) = 0$ , ce qui signifie que si on connaît  $Y$  (c'est-à-dire le symbole reçu), le fait de nous donner  $X$  (c'est-à-dire le symbole émis par  $X$ ) ne nous apporte aucune information supplémentaire. En effet, la transmission est parfaite et connaître  $Y$  nous suffit.
- Lorsque  $p = \frac{1}{2}$ , c'est-à-dire lorsque la transmission est complètement aléatoire, ou encore à dire que quelque soit le symbole émis par la source, on n'a pas la moindre idée du symbole que l'on va recevoir,  $I(X, Y) = 0$  et  $H(X, Y) = 2$ . Les deux sources sont complètement indépendantes et il n'y a aucune similitude entre les symboles émis par l'émetteur et les symboles reçus par le récepteur. L'information jointe des deux sources indépendantes est donc bien la somme des informations propres de chaque source, c'est-à-dire  $1+1=2$ . On pourrait aller plus loin, et même dire que dans ce cas, autant ne rien envoyer du tout au niveau de l'émetteur et générer un bit aléatoire (une chance sur deux) au niveau du récepteur.
- Lorsque  $p = 1$ , la transmission se fait avec erreur, à coup sûr, à chaque bit émis. Dans ce cas,  $I(X, Y) = 1$  et les deux sources sont à nouveau complètement semblables, le fait qu'il y a permutation du 0 et du 1 n'ayant aucune importance.

### 3.3 Codage de sources discrètes

Le codage de source est une technique visant à utiliser au mieux la capacité de stockage et/ou transmission d'un système supposé sans erreur. Par exemple, nous pouvons citer le stockage sur disque dur, la transmission numérique filaire, ... Il est donc nécessaire d'adapter les symboles émis par la source d'information que l'on considère au média visé.

Le but du codage de source est de remplacer les messages émis par une source  $X$  utilisant un alphabet  $K$ -aire  $\{x_1, x_2, \dots, x_K\}$  par des messages écrits dans un alphabet  $q$ -aire qui est celui utilisé par un canal de communication ou un système de stockage d'information. Dans la plupart des cas, on utilise un alphabet binaire ( $q = 2$ ), mais les résultats donnés ci-dessous resteront valables quelque soit  $q > 1$ . Par ailleurs, la taille de l'alphabet d'origine peut être quelconque ; dès lors, nous pourrons considérer le codage symbole par symbole, ou blocs de symboles par blocs de symboles.

Par exemple, notre alphabet est comporte 26 symboles. Le code MORSE (développé en 1837) réalisait une conversion de cet alphabet vers un alphabet quaternaire ( $q = 4$ ) : le point, le trait, l'espace court et l'espace long. Il s'agit d'un code de longueur variable qui associe la séquence la plus courte à la lettre la plus fréquente en anglais (le 'E').

Nous allons étudier les codes de longueur fixe et de longueur variable et, pour la simplicité, nous considérerons qu'un tel code associe à chaque symbole  $x_k$  ( $k = 1, 2, \dots, K$ ) de la source un **mot de code**  $q$ -aire, noté  $C_k$ , c'est-à-dire une suite de  $n_k$  symboles de l'alphabet de destination  $q$ -aire. Dans le cas des codes de longueur fixe, tous les  $n_k$  sont égaux ( $n_1 = n_2 = \dots = n_K = R$ ).

### 3.3.1 Codage avec mots de longueur fixe

Une manière simple de coder en binaire l'alphabet  $K$ -aire d'une source est d'attribuer à chaque symbole  $R$  bits. Il y a donc  $2^R$  mots de code (ou tout simplement "codes") possibles. Nous avons bien sûr la condition

$$2^R \geq K \quad (3.15)$$

sinon il n'y a pas assez de mots de code pour coder chaque symbole de l'alphabet de la source, l'égalité étant possible lorsque le nombre  $K$  de symboles de la source est une puissance de 2. Dans le cas contraire, nous aurons

$$2^{R-1} < K < 2^R \quad (3.16)$$

Bien sûr, on pourrait se contenter de  $2^R \geq K$  sans limitation supérieure sur la valeur de  $R$ . Mais cela n'aurait pas beaucoup de sens. En effet, considérons l'exemple simple d'une source de  $K = 3$  symboles  $\{a, b, c\}$ . Si on choisit de coder cet alphabet sur  $R = 8$  bits, cela respecte la condition  $2^8 = 256 \geq 3$  mais cela n'est pas du tout efficace car nous disposons de 256 mots de code possibles pour seulement 3 symboles à coder. Il y a donc un gaspillage de 253 mots de code qui ne seront pas utilisés. Par contre, si nous recherchons une valeur de  $R$  qui respecte (3.16), nous obtenons

$$2^{2-1} = 2 < 3 < 2^2 = 4$$

et donc  $R = 2$ . Nous disposons donc de 4 mots de code possibles, un seul ne sera pas utilisé, et donc beaucoup moins de gaspillage.

L'expression (3.16) permet de déterminer le nombre  $R$  de bits nécessaires au codage de l'alphabet d'une source de  $K$  symboles :

$$R = \lceil \log_2 K \rceil \quad (3.17)$$

où nous avons introduit la notation  $\lceil x \rceil$  qui représente le plus petit entier supérieur ou égal à  $x$ .

**Exemples.** Reprenons le cas de notre source de  $K = 3$  symboles  $\{a, b, c\}$ . La formule (3.17) nous fournit directement

$$R = \lceil \log_2 3 \rceil = \lceil 1,584963 \rceil = 2$$

valeur que nous avions déjà trouvée plus haut grâce à la relation (3.16). Considérons à présent une source de  $K = 16$  symboles. La formule (3.17) nous fournit alors

$$R = \lceil \log_2 16 \rceil = \lceil 4 \rceil = 4$$

La relation (3.17) est donc valable quelque soit la valeur de  $K$ , que  $K$  soit une puissance de 2 ou pas.

Nous allons à présent établir une première relation importante pour les codes de longueur fixe. Nous savons, étant donné la relation (3.15), que

$$R \geq \log_2 K$$

mais nous savons également que l'entropie d'une source discrète  $X$ , sans mémoire, et d'alphabet  $K$ -aire est bornée par  $\log_2 K$  (voir relation(3.3)) :

$$H(X) \leq \log_2 K$$

En combinant ces deux dernières relations, nous obtenons l'importante propriété suivante :

$$R \geq H(X) \quad (3.18)$$

L'égalité a lieu lorsque tous les symboles de la source sont équiprobables et lorsque  $K$  est une puissance de 2. Le nombre de bits  $R$  utilisés pour coder chaque symbole de la source  $X$  ne pourra donc jamais être inférieur à l'entropie  $H(X)$  de cette source.

Cette dernière constatation nous amène à considérer la notion d'efficacité d'un codage. Un codage est d'autant plus efficace que le nombre de mots de code possibles inutilisés est faible. L'efficacité dépend aussi de la quantité d'information moyenne de la source. Nous arrivons donc à la définition suivante.

**Définition (Efficacité d'un codage).** L'efficacité d'un codage codant les  $K$  symboles d'une source  $X$  sur  $R$  bits est définie par

$$\eta = \frac{H(X)}{R} \quad (3.19)$$

L'efficacité d'un codage est donc un nombre compris entre 0 et 1, il est généralement exprimé en %. Un codage sera donc d'autant plus efficace que le nombre de bits qu'il utilise pour coder chaque symbole de la source se rapproche de l'entropie de cette source.

**Exemple.** Considérons une source  $X$  de  $K = 24$  symboles. La relation (3.17) nous permet de calculer

$$R = \lceil \log_2 24 \rceil = \lceil 4,584963 \rceil = 5$$

Le nombre de mots de code possibles est donc égal à  $2^5 = 32$ , il y a donc 8 mots de code non utilisés. Si tous les symboles de la source  $X$  sont équiprobables, l'entropie de la source  $X$  est égale à

$$H(X) = \log_2 K = \log_2 24 = 4,584963$$

et l'efficacité du codage est alors égale à

$$\eta = \frac{H(X)}{R} = \frac{4,584963}{5} = 91,7\%$$

### 3.3.2 Codage par blocs : Extension de la source

Pour améliorer l'efficacité du codage, on peut transmettre, donc coder, les symboles non pas individuellement mais par blocs de  $J$  symboles. Cette technique est appelée **extension de la source**.

**Exemple.** Considérons une source  $X$  de  $K = 2$  symboles  $\{A, B\}$ . En rassemblant les symboles de la source  $X$ , dite **primaire**, par blocs de  $J = 2$ , on obtient une source, dite **secondaire**, de 4 symboles  $\{AA, AB, BA, BB\}$ .

Dès lors, au départ d'une source **primaire** de  $K$  symboles, nous formons une source **secondaire**, ou **étendue**, de  $K^J$  symboles. Si nous codons sur  $N$  bits chaque symbole de la source secondaire, nous avons de nouveau une relation similaire à (3.15) :

$$2^N \geq K^J \quad (3.20)$$

ce qui peut encore s'écrire

$$N \geq J \log_2 K \quad (3.21)$$

En réutilisant la notation  $\lceil \cdot \rceil$ , nous pouvons donner une relation permettant de calculer la plus petite valeur de  $N$  possible :

$$N = \lceil J \log_2 K \rceil \quad (3.22)$$

En codant chaque symbole de la source secondaire, c'est-à-dire  $J$  symboles de la source primaire, sur  $N$  bits, cela revient à coder chaque symbole de la source primaire sur

$$R = \frac{N}{J} \quad (3.23)$$

bits. Dans ce cas,  $R$  n'est plus un entier mais on peut écrire, en tenant compte de (3.21) et (3.3),

$$R = \frac{N}{J} \geq \log_2 K \geq H(X) \quad (3.24)$$

Quelque soit le codage par blocs choisi, le nombre  $R$  de bits par symbole de la source primaire  $X$  restera supérieur ou égal à l'entropie de la source  $H(X)$ . Néanmoins, on peut montrer (...) que l'efficacité du nouveau codage

$$\eta = \frac{H(X)}{R} = \frac{J H(X)}{N} \quad (3.25)$$

peut être améliorée par rapport au codage simple de longueur fixe. Dans cette dernière formule, tout se passe comme si la source étendue avait une entropie de  $J H(X)$ .

**Exemple.** Reprenons la cas de la source  $X$  de  $K = 24$  symboles. Si nous formons des blocs de  $J = 3$  symboles, le nombre de symboles de la source étendue est  $K^J = 24^3 = 13824$  et la formule (3.22) permet de calculer

$$N = \lceil J \log_2 K \rceil = \lceil 3 \log_2 24 \rceil = \lceil 13,754888 \rceil = 14$$

Le nombre de mots de code possibles de la source secondaire est donc égal à  $2^N = 2^{14} = 16384$ . Le nombre de bits par symbole de la source primaire est alors égal à

$$R = \frac{N}{J} = \frac{14}{3} = 4,666667$$

Si tous les symboles de la source primaire sont équiprobables, l'entropie de la source est  $H(X) = \log_2 K = 4,584963$  et nous avons toujours  $R \geq H(X)$ . Par contre, l'efficacité du codage est maintenant égale à

$$\eta = \frac{H(X)}{R} = \frac{4,584963}{4,666667} = 98,25\%$$

Ceci illustre bien que l'on peut augmenter l'efficacité d'un codage de longueur fixe en construisant une source étendue.

### 3.3.3 Codage avec mots de longueur variable

Lorsque tous les symboles de la source ne sont pas équiprobables, l'extension de source ne permet pas d'augmenter jusque 100 % l'efficacité du codage. Historiquement, le code de MORSE, dont on a déjà parlé plus haut, résouds ce problème. L'idée est d'utiliser un mot de code "court" pour les symboles les plus utilisés (de probabilité élevée) et de réserver un mot de code plus "long" aux symboles peu utilisés (de probabilité faible). C'est cette idée qui est reprise et formalisée dans le codage avec mots de longueur variable.

### 3.3.3.1 Codages avec et sans préfixe

Pour introduire les notions essentielles, nous allons utiliser un exemple de trois codages possibles pour une source de  $K = 4$  symboles  $x_k \{B, F, I, O\}$  de probabilités respectives  $p_k$  ( $k = 1, 2, 3, 4$ ). Ces trois codages sont repris à la table 3.1.

Symbol	Probabilité $p_k$	Code I	Code II	Code III
$I$	0,5	1	0	0
$B$	0,25	00	10	01
$F$	0,125	01	110	011
$O$	0,125	10	111	111

TABLE 3.1 – Exemple de trois codages pour une source de 4 symboles  $\{B, F, I, O\}$ .

Supposons à présent que nous cherchions à coder le message “BOF”. Nous pouvons faire les constatations suivantes :

- Avec le codage I, le message codé est 001001. Décodons ce message en utilisant la table 3.1. Qu'obtient-on ? Bien sûr, le message de départ : 00 10 01 =  $B O F$ . Mais ce décodage n'est **pas unique** ! En effet, on pourrait interpréter les message codé de la manière suivante : 00 1 00 1 =  $B I B I$ . Ceci est dû au fait que le 1, mot de code attribué au symbole  $I$ , est le début d'un autre mot de code 10, attribué au symbole  $O$ . Pour éviter cette situation, il ne faut pas qu'un mode de code soit le **préfixe** d'un autre mot de code. Les codages qui remplissent cette condition sont dits “**sans préfixe**”.
- Avec le codage III, le message codé est 01111011. Au décodage, nous pouvons voir 0 111... c'est-à-dire  $IO...$  Mais ici, nous nous rendons compte du fait que ce qui suit, c'est-à-dire soit 1, soit 10, soit 101 ne sont pas des mots de code et donc, nous pouvons revenir en arrière pour modifier l'interprétation, soit 01 111 011 et retrouver le message. Mais le message n'est pas décodable de manière **instantanée**. Ceci est aussi dû au fait que le code utilisé n'est pas un codage sans préfixe.
- Avec le codage II, le message codé est 10111110. Le décodage se fait maintenant de manière unique et sans retour en arrière : 10 111 110 =  $B O F$ . Le codage II est un codage sans préfixe et nous avons les propriétés souhaitées : décodable de manière **unique** et de manière **instantanée**.

Nous en arrivons donc à la définition suivante.

**Définition (Codage sans préfixe).** Un codage sans préfixe est, par définition, un codage dont aucun mot de code n'est le préfixe d'un mot de code. Un code sans préfixe est décodable de manière unique et instantanée.

### 3.3.3.2 Arbre d'un codage

Un codage peut être représenté de manière graphique par un arbre. Dans le cas d'un codage binaire ( $q = 2$ ), cet arbre est binaire. Les arbres sont également des représentations commodes pour décrire les algorithmes de codage et de décodage. Les arbres correspondant aux trois codages de la table 3.1 sont présentés à la figure 3.5.

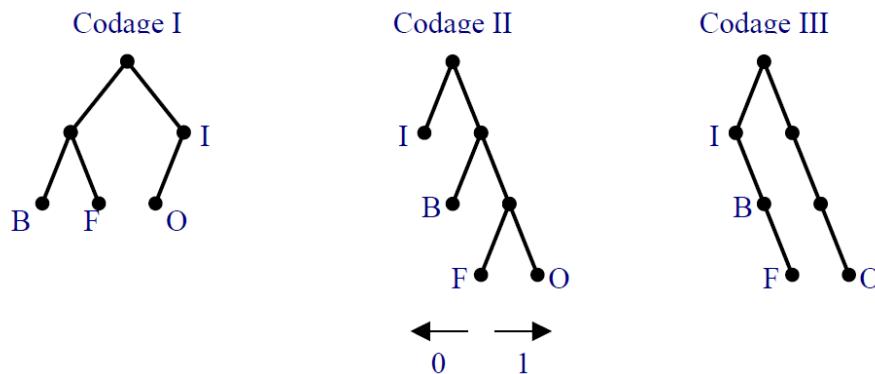


FIGURE 3.5 – Arbres correspondant aux codage de la table 3.1.

Pour ces arbres, les définitions et conventions suivantes ont été adoptées :

- Un déplacement à gauche correspond à un 0.
- Un déplacement à droite correspond à un 1.
- Chaque déplacement crée un **noeud** de l’arbre.
- Chaque noeud a un père (vers le haut), à l’exception du noeud “**racine**” de l’arbre.
- Chaque noeud peut avoir 0, 1 ou 2 fils (vers le bas).
- Le lien entre deux noeuds est une **branche**.
- Un noeud qui n’a pas de fils est une **feuille**.

L’observation de l’arbre du codage II nous permet d’introduire une nouvelle définition d’un code sans préfixe.

**Définition (Codage sans préfixe).** Un codage sans préfixe est un codage dont les symboles codés sont des feuilles d’un arbre <sup>a</sup>.

<sup>a</sup>. Toutes ces notions sont bien sûr généralisables à des arbres et codage  $q$ -aire, mais, comme nous l’avons dit plus haut, nous nous limiterons ici au cas  $q = 2$ .

### 3.3.3.3 Longueur moyenne des mots de code

Comme nous l’avons vu, les différents mots de code ne vont plus avoir la même longueur. Ce qui nous amène à introduire la longueur moyenne des mots de code. Pour une source  $X$  d’alphabet  $\{x_1, x_2, \dots, x_K\}$  de probabilités associées  $\{p_1, p_2, \dots, p_K\}$  et dont chaque symbole  $x_k$  est codé sur  $n_k$  ( $k = 1, 2, \dots, K$ ) bits, la longueur moyenne des mots de code est donnée par

$$\overline{R} = \sum_{k=1}^K n_k p_k \quad (3.26)$$

La valeur de  $\overline{R}$  n’est pas nécessairement une valeur entière.

### 3.3.3.4 Inégalité de KRAFT

Nous allons à présent établir une relation importante de la théorie des codes, celle-ci fournit une condition nécessaire et suffisante d'existance d'un codage sans préfixe, exprimée en fonction de la longueur des mots de code, et porte le nom d'inégalité de KRAFT.

Nous venons de voir qu'un codage sans préfixe peut se fabriquer à partir d'un arbre de codage et que sa condition d'obtention est que les codes soient des feuilles de l'arbre. Le schéma de la figure 3.6 va illustrer le raisonnement conduisant à l'inégalité de KRAFT :

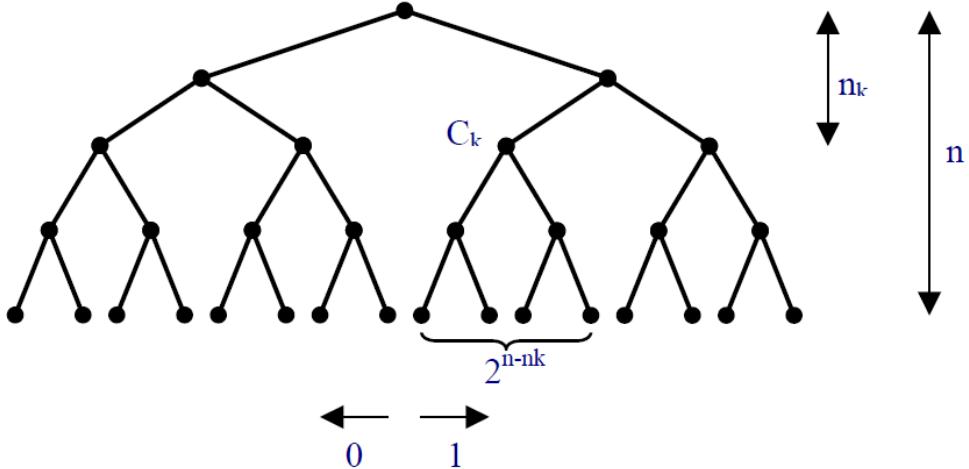


FIGURE 3.6 – Illustration de la mise en place de l'inégalité de KRAFT.

- Nous construisons un arbre binaire de longueur  $n$  (sur la figure 3.6,  $n = 4$ ). Dès lors, le nombre de mots de code possibles (nombre de feuilles possibles) est donc égal à  $2^n$ .
- A la hauteur  $n_k$  (sur la figure 3.6,  $n_k = 2$ ), nous décidons d'attribuer un noeud à un code  $C_k$ . Ce noeud devient une feuille de l'arbre et, pour obtenir un code sans préfixe, cela interdit tous les noeuds qui peuvent s'en déduire. Le nombre de noeuds interdits est égal à  $2^{n-n_k}$ .
- Si l'alphabet à coder contient  $K$  symboles  $x_k$  auxquels sont attribués des mots de code de longueur  $n_k$ , le nombre total de feuilles interdites est égal à

$$\sum_{k=1}^K 2^{n-n_k}$$

- Le nombre de feuilles interdites doit être inférieur ou égal au nombre de feuilles finales, c'est-à-dire

$$\sum_{k=1}^K 2^{n-n_k} \leq 2^n$$

En divisant chaque membre de cette dernière relation par  $2^n$ , nous obtenons l'**inégalité de KRAFT** :

$$\boxed{\sum_{k=1}^K 2^{-n_k} \leq 1} \quad (3.27)$$

Celle-ci nous donne donc une condition nécessaire et suffisante d'existence d'un code sans préfixe. Cela signifie tout d'abord que si un code est sans préfixe, il doit vérifier (3.27). Prenons l'exemple du codage II cité plus haut. Pour celui-ci, nous avons

$$2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = 1 \leq 1$$

Par contre, pour celui du codage I cité plus haut, nous avons

$$2^{-1} + 2^{-2} + 2^{-2} + 2^{-2} = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{5}{4} > 1$$

Ensuite, si on choisit un jeu de valeurs  $\{n_1, n_2, \dots, n_K\}$  qui vérifie (3.27), alors il existe un codage sans préfixe dont les mots de code, de longueurs respectives  $n_1, n_2, \dots, n_K$ , permettent de coder l'alphabet d'une source de  $K$  symboles. Par contre, l'inégalité de KRAFT ne nous donne pas la technique permettant de le créer... Nous y reviendrons.

Mais avant cela, nous allons établir le **premier théorème de SHANNON**, aussi appelé **théorème de codage de source**, qui constitue un des éléments les plus importants de la théorie de l'information.

### 3.3.3.5 Premier théorème de SHANNON : Théorème de codage de source

Comme nous allons le voir, ce théorème énonce, sous la forme d'inégalités, les conditions sur  $\overline{R}$  pour un codage sans préfixe.

Commençons par établir une première inégalité qui fournit la **limite inférieure** de la valeur de  $\overline{R}$ . Pour une source  $X$ , d'entropie  $H(X)$ , nous pouvons écrire, en utilisant les relations (3.2) et (3.26),

$$\begin{aligned} H(X) - \overline{R} &= \sum_{k=1}^K p_k \log_2 \frac{1}{p_k} - \sum_{k=1}^K p_k n_k \\ &= \sum_{k=1}^K p_k \log_2 \frac{1}{p_k} + \sum_{k=1}^K p_k \log_2 2^{-n_k} \\ &= \sum_{k=1}^K p_k \log_2 \frac{2^{-n_k}}{p_k} \\ &= \sum_{k=1}^K p_k \ln \frac{2^{-n_k}}{p_k} \log_2 e \end{aligned}$$

où nous avons utilisé le fait que  $\log_2 x = \ln x / \ln 2$ . En tenant compte du fait que  $\ln x \leq x - 1$

pour  $x > 0$  (l'égalité ayant lieu pour  $x = 1$ ), il vient ensuite

$$\begin{aligned} H(X) - \bar{R} &\leq \sum_{k=1}^K p_k \left( \frac{2^{-n_k}}{p_k} - 1 \right) \log_2 e \\ &= \left[ \sum_{k=1}^K 2^{-n_k} - \sum_{k=1}^K p_k \right] \log_2 e \\ &\leq [1 - 1] \log_2 e \\ &= 0 \end{aligned}$$

où nous avons utilisé l'inégalité de KRAFT (3.27) et le fait que  $\sum_{k=1}^K p_k = 1$ . Nous avons donc la première inégalité suivante

$$H(X) \leq \bar{R} \quad (3.28)$$

qui est à comparer à (3.18), et qui exprime que le nombre moyen de bits par symbole utilisés pour coder la source  $X$  ne pourra jamais être inférieur à l'entropie de la source  $X$ . L'égalité correspond au cas

$$p_k = \frac{1}{2^{n_k}}$$

pour tout  $k = 1, 2, \dots, K$ .

Nous allons à présent établir une **limite supérieure** que peut avoir  $\bar{R}$  si le choix des  $n_k$  est fait de manière judicieuse. Nous savons que la limite inférieure peut être théoriquement atteinte par un choix justicieux des  $n_k$  tel que

$$2^{-n_k} = p_k$$

soit encore

$$n_k = -\log_2 p_k$$

ce qui se traduit par le fait que plus un symbole  $x_k$  de la source  $X$  est probable, moins on lui attribue de bits. Cette condition n'est pas facilement réalisable car  $n_k$  est un entier. En pratique, nous devons choisir pour  $n_k$  le plus petit entier supérieur ou égal à  $-\log_2 p_k$ , c'est-à-dire

$$n_k = \lceil -\log_2 p_k \rceil$$

ce qui peut encore s'exprimer par

$$2^{-n_k} \leq p_k < 2^{-n_k+1}$$

L'inégalité de droite nous permet d'écrire

$$\log_2 p_k < -n_k + 1$$

et ensuite

$$-p_k \log_2 p_k > n_k p_k - p_k$$

En sommant ces  $K$  inégalités, nous obtenons

$$\begin{aligned} H(X) &= -\sum_{k=1}^K p_k \log_2 p_k \\ &> \sum_{k=1}^K (n_k p_k - p_k) \\ &= \sum_{k=1}^K n_k p_k - \sum_{k=1}^K p_k \\ &= \bar{R} - 1 \end{aligned}$$

qui peut encore s'écrire

$$\overline{R} < H(X) + 1 \quad (3.29)$$

Cette dernière relation ne signifie pas que  $\overline{R}$  sera inférieur à l'entropie  $H(X)$  de la source mais seulement qu'il existe au moins un codage de longueur de mot variable pour lequel (3.29) est vrai. Le rassemblement des inégalités (3.28) et (3.29) nous conduit au premier théorème de SHANNON.

**Théorème (Codage de source, SHANNON).** Soit une source discrète  $X$ , sans mémoire, de  $K$  symboles de probabilités respectives  $p_k$ . Le théorème de codage de source (ou premier théorème de SHANNON) assure que l'on peut trouver un codage avec des mots de code de longueur variable  $n_k$  tel que

$$H(X) \leq \overline{R} < H(X) + 1 \quad (3.30)$$

Un tel codage est alors appelé **codage de SHANNON**.

Bien sûr, tous les codages avec mots de longueur variable ne sont pas des codes de SHANNON. Seuls ceux vérifiant (3.29) sont des codes de SHANNON. Par contre, bien que le théorème de SHANNON nous donne une condition d'existance d'un code de SHANNON, il ne nous apporte aucune information sur la manière de construire un tel code... Nous y venons.

## 3.4 Introduction à la compression de données

Aujourd'hui, les quantités de données informatiques à gérer, transmettre, stocker dans le monde entier sont considérables. Le problème de la compression optimale de ces données est donc essentielles afin d'optimiser les ressources. Ce problème consiste à remplacer une quantité de données  $D_1$  (généralement exprimée en bits) par une autre quantité de données  $D_2$  (exprimée dans la même unité que  $D_1$ , cela à des fins de comparaisons), plus petite. Nous dirons que  $D_2$  correspond aux données compressées de  $D_1$ , et nous pouvons définir le taux de compression d'une technique d'encodage comme

$$\tau = \frac{D_1}{D_2} = \frac{\text{nombre de bits avant compression}}{\text{nombre de bits après compression}} \quad (3.31)$$

Elle mesure donc l'efficacité d'une technique de compression. Une autre manière de voir les choses est de considérer le pourcentage de la quantité de données compressée  $D_2$  par rapport à la quantité de données originales  $D_1$ .

On distingue deux grandes catégories de techniques de compression :

- les techniques de compression **sans perte** qui permettent de retrouver exactement, c'est-à-dire sans aucune perte, les données  $D_1$  à partir des données compressée  $D_2$ . Ces techniques mènent à des taux de compression de 2 à 3, dans le cas d'images médicales.
- les techniques de compression **avec perte** qui, contrairement aux précédentes, ne permettent pas de retrouver exactement les données  $D_1$  à partir des données compressées  $D_2$ . Pour certaines applications, cela n'est pas gênant. Il se fait qu'une partie de l'information théoriquement disponible n'est pas toujours perceptible. Par exemple, l'oeil humain n'est pas capable de voir les atomes sans microscope. Ainsi, il est inutile de décrire les objets au niveau atomique. Dans le cas du son, le principe consiste à supprimer, par filtrage, toutes les fréquences que l'oreille ne peut pas entendre. Les techniques de compression avec perte permettent d'atteindre des taux de compression de l'ordre de 10,

dans le cas d'images naturelles.

Dans la suite de cette introduction, nous nous limiterons aux techniques de compression **sans perte**. Parmi ces techniques, on distingue différentes catégories :

- les **codages statistiques** qui utilisent les propriétés statistiques de la source de données pour associer à chaque symbole un nombre de bits d'autant plus petit que leur probabilité (fréquence) d'utilisation augmente. Parmi ces codages, on peut citer le codage de HUFFMAN et le codage arithmétique.
- les **codages par dictionnaire**. Ces codages consistent à remplacer les chaînes de caractères rencontrées précédemment par leur adresse dans une table au fur et à mesure du codage. Le décodeur procède de façon symétrique et reconstitue le dictionnaire par le même algorithme. Parmi ces codages, on peut citer le codage de LEMPEL, ZIV et WELSH (LZW) qui est à la base de la compression Zip.
- Les **codages par répétition**. Le principe employé pour ces codages est très simple : toute suite d'octets de même valeur est remplacée par la valeur, à laquelle on associe le nombre d'occurrences de cette valeur. Parmi ces codages, on peut citer le codage RLC (Run Length Coding).

Toutefois, quelque soit la technique utilisée, il sera impossible de dépasser la limite de SHANNON énoncée dans son théorème (3.30) et qui dit que le nombre moyen de bits que l'on associera par symbole ne pourra être inférieur à l'entropie de la source. Toutes ces techniques de compression visent donc au maximum de se rapprocher de la limite imposée par le théorème de SHANNON.

Nous allons à présent aborder les techniques de compression les plus courantes. Celles-ci seront décrites de manière assez générale et il faut avoir à l'esprit qu'elles possèdent toutes des variantes et améliorations selon l'implémentation.

### 3.4.1 Codage de HUFFMAN

Il s'agit d'un codage de source avec mots de code de longueur variable. Mis au point en 1952 et basé sur les probabilités d'utilisation des symboles de la source, c'est un algorithme qui minimise le nombre moyen de bits utilisés pour le codage. Le principe de la méthode de HUFFMAN est d'associer aux symboles les plus probables le plus petit nombre de bits et aux symboles les moins probables le plus grand nombre de bits.

L'algorithme de construction des mots de code est le suivant :

1. La première étape consiste à réorganiser les symboles par ordre de probabilité décroissante. Chaque symbole est alors associé à une feuille d'un arbre en construction.
2. On relie ensuite les feuilles en créant un noeud auquel on associe la somme des probabilités des deux symboles correspondants. À chaque étape, on fusionne les 2 noeuds (ou feuilles) ayant les probabilités les plus faibles.
3. On répète ce processus jusqu'à ce qu'il ne reste plus qu'un seul noeud dont la probabilité associée vaut 1. Ce noeud correspond à la racine de l'arbre.

4. On construit la table de codage en partant de la racine de l'arbre et en descendant le long de l'arbre en associant à chaque branche un 1 ou un 0, et cela jusqu'aux feuilles représentant les symboles à coder.

Cet algorithme fournit un codage optimal, ou encore dit de SHANNON, qui respecte la double inégalité (3.30).

**Exemple.** Considérons une source  $X$  de  $K = 4$  symboles  $\{I, B, O, F\}$  et de probabilités respectives  $p_1 = 0,5$ ,  $p_2 = 0,25$ ,  $p_3 = 0,125$  et  $p_4 = 0,125$ . L'arbre construit par l'algorithme de HUFFMAN est présenté à la figure 3.7. En parcourant l'arbre ainsi construit, on obtient la table de codage fournie à la table 3.2.

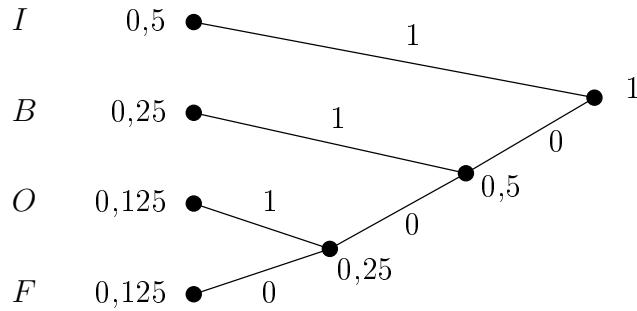


FIGURE 3.7 – Exemple de construction d'un arbre par la méthode HUFFMAN.

Symbole	Probabilité $p_k$	Mot de code $C_k$	Longueur $n_k$
$I$	0,5	1	1
$B$	0,25	01	2
$O$	0,125	001	3
$F$	0,125	000	3

TABLE 3.2 – Table de codage de l'alphabet  $\{I, B, O, F\}$ .

Le codage obtenu est bien entendu sans préfixe. Nous allons à présent vérifier dans quelle mesure il s'accorde avec le théorème de SHANNON (3.30). Le nombre moyen de bits par symbole est donné par

$$\bar{R} = \sum_{k=1}^K n_k p_k = 1,75 \text{ bits/symbole}$$

tandis que l'entropie de la source est égale

$$H(X) = - \sum_{k=1}^K p_k \log p_k = 1,75 \text{ bits/symbole}$$

et le théorème de SHANNON (3.30) est vérifié :

$$1,75 \leq 1,75 < 1,75 + 1$$

En particulier ici,  $\bar{R} = H(X)$  car le codage est tel que  $p_k = 2^{-n_k}$  pour tout  $k = 1, 2, 3, 4$ .

**Exemple 2.** Considérons un fichier texte de 1000 caractères ne comportant que les symboles  $\{E, A, S, T, U, Y\}$ . Supposons qu'une analyse statistique du fichier fournisse les fréquences

d'occurrence respectives suivantes : 48%, 21%, 12%, 8%, 6%, 5%. Ces fréquences s'apparentent aux probabilités d'utilisation des différents symboles. Nous pouvons donc construire un arbre binaire pour cet alphabet en utilisant la méthode HUFFMAN. Un exemple est fourni à la figure 3.8. La table de codage correspondante est donnée à la table 3.3.

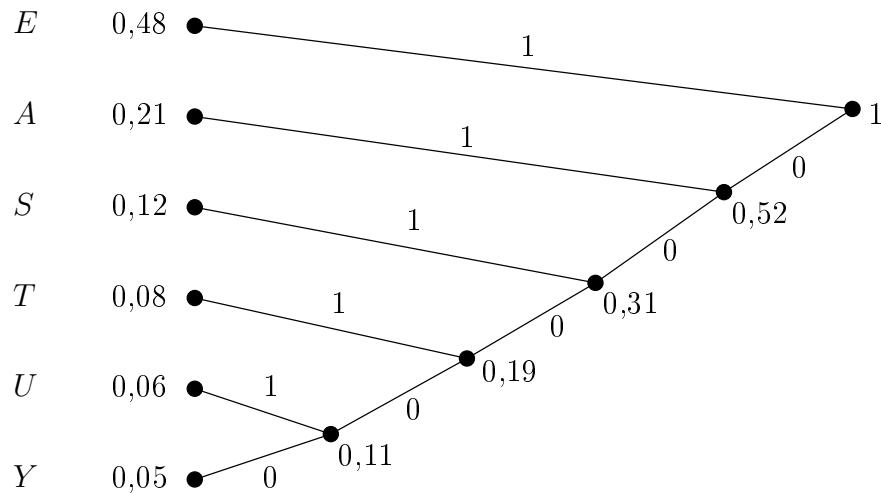


FIGURE 3.8 – Exemple de construction d'un arbre par la méthode HUFFMAN.

Symbol	Probabilité $p_k$	Mot de code $C_k$	Longueur $n_k$
$E$	0,48	1	1
$A$	0,21	01	2
$S$	0,12	001	3
$T$	0,08	0001	4
$U$	0,06	00001	5
$Y$	0,05	00000	5

TABLE 3.3 – Table de codage de l'alphabet  $\{E, A, S, T, U, Y\}$ .

Le nombre moyen de bits par symbole est donné par

$$\overline{R} = \sum_{k=1}^K n_k p_k = 2,13 \text{ bits/symbole}$$

tandis que l'entropie de la source est égale

$$H(X) = - \sum_{k=1}^K p_k \log p_k = 2,11 \text{ bits/symbole}$$

et le théorème de SHANNON (3.30) est vérifié :

$$2, 11 \leq 2, 13 < 2, 11 + 1$$

Si le fichier original est codé en ASCII (7 bits), sa taille est alors donnée par

$$D_1 = 1000 \cdot 7 = 7000 \text{ bits}$$

tandis qu'après compression, nous avons

$$D_2 = 1000 \cdot 2,13 = 2130 \text{ bits}$$

et le taux de compression est égal à

$$\tau = \frac{D_1}{D_2} = \frac{7000}{2130} = 3,29$$

ou, on peut encore dire que l'on a une compression de  $2130/7000 = 30,4\%$ .

### 3.4.2 Codage arithmétique

Le principe du codage de HUFFMAN était de remplacer chaque symbole d'un message à coder par un nombre de bits dépendant de la probabilité de ces symboles. Le codage arithmétique que nous allons aborder ici est également basé sur la connaissance des probabilités de chaque symbole de l'alphabet de la source. Il fait donc partie de la catégorie des codages statistiques. Toutefois, celui-ci établira un code correspondant au message à coder dans son intégralité, et non pas au codage de chaque symbole.

L'idée de ce codage est de trouver un sous-intervalle  $[L_c; H_c]$  de l'intervalle  $[0; 1]$  qui représente le code recherché et qui identifie le message codé de manière unique. Tout nombre réel de ce sous-intervalle, ou plus précisément l'**expansion binaire** (voir annexe 3.6.2) de tout réel de ce sous-intervalle, fournit le code binaire codant le message de départ. Soyons plus précis.

#### Algorithme de codage

Soit une source discrète  $X$  présentant  $K$  symboles  $x_k$  ( $k = 1, 2, \dots, K$ ) dont les probabilités respectives sont  $p_1, p_2, \dots, p_K$ . Et, soit  $s_1 s_2 \dots s_n$  la séquence de symboles que l'on souhaite coder, chaque  $s_i$  ( $i = 1, 2, \dots, n$ ) étant un des symboles  $x_k$  de la source. L'algorithme de codage arithmétique comporte alors les étapes suivantes :

1. On initialise un premier intervalle  $[L_c; H_c] = [0; 1]$ . La taille  $T$  de cet intervalle est  $T = H_c - L_c = 1$ . Le symbole en cours de codage est initialisé à  $s_1$ , c'est-à-dire  $i = 1$ .
2. Cet intervalle  $[L_c; H_c]$  est partitionné en  $K$  sous-intervalles  $[L_k; H_k]$  proportionnellement aux probabilités  $p_k$  des symboles de la source. Nous avons donc

$$L_k = L_c + T \sum_{j=1}^{k-1} p_j \quad \text{et} \quad H_k = L_c + T \sum_{j=1}^k p_j$$

pour tout  $k = 1, 2, \dots, K$ . A chacun de ces sous-intervalles est associé le symbole  $x_k$  correspondant.

3. On détermine le sous-intervalle  $[L_k; H_k]$  correspondant au symbole  $s_i$  ( $= x_k$ ) en cours de codage. Ensuite, on met à jour l'intervalle  $[L_c; H_c]$  de la manière suivante

$$L_c \rightarrow L_k \quad \text{et} \quad H_c \rightarrow H_k$$

et la taille  $T$  est alors mise à jour en fonction des nouvelles valeurs de  $L_c$  et  $H_c$ .

4. On répète les étapes 2 et 3 pour le symbole  $s_i$  suivant ( $i$  est incrémenté de 1) et on répète ces étapes jusqu'au dernier symbole  $s_n$  à coder.

A la fin de cette procédure, on obtient l'intervalle final  $[L_c; H_c]$  qui correspond à la séquence  $s_1 s_2 \dots s_n$  dans son intégralité. Tout nombre réel de cet intervalle fournit un code de cette séquence. Plus précisément, l'expansion binaire de tout réel de cet intervalle est un code de cette séquence. Nous reviendrons plus tard sur cette notion d'expansion binaire. Commençons plutôt par illustrer cet algorithme sur un exemple.

### Exemple

Soit une source discrète  $X$  de  $K = 5$  symboles  $\{a, b, c, d, e\}$  de probabilités respectives  $p_1 = 0,3, p_2 = 0,25, p_3 = 0,2, p_4 = 0,15$  et  $p_5 = 0,1$ . Et, soit  $bdcea$  la séquence que l'on souhaite coder. L'évolution de l'algorithme est illustré à la figure 3.9.

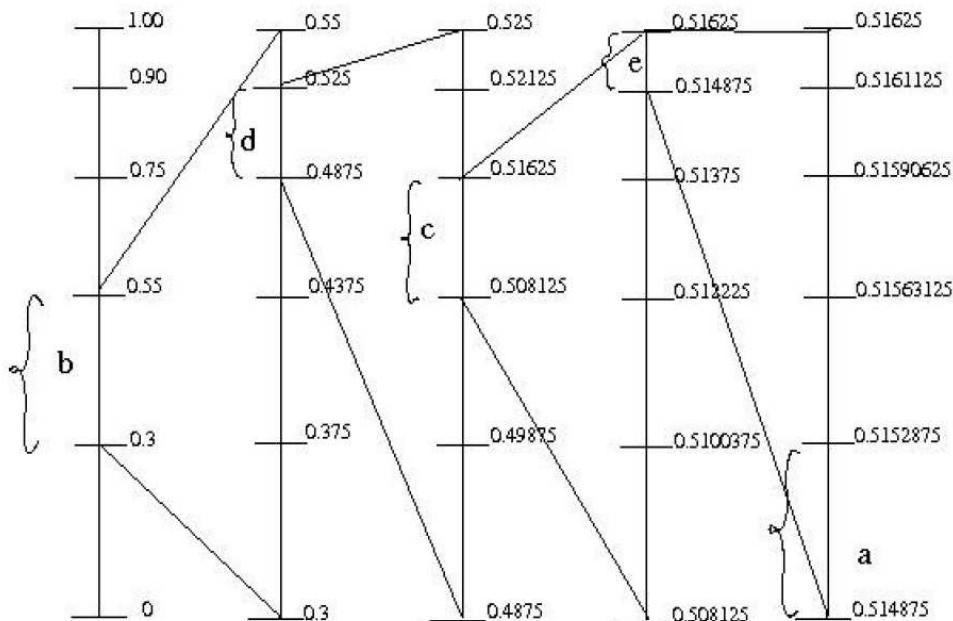


FIGURE 3.9 – Illustration de l'algorithme de codage arithmétique.

Pour coder cette séquence, l'intervalle  $[L_c; H_c] = [0; 1]$  ( $T = 1$ ) a été divisé en 5 sous-intervalles selon les probabilités  $p_k$ . Etant donné que le premier symbole de la séquence est  $b$ , c'est l'intervalle  $[L_2; H_2] = [0,3; 0,55]$  qui a été retenu. Les valeurs  $L_c, H_c$  et  $T$  ont alors été mises à jour selon

$$\begin{aligned} L_c &\rightarrow 0,3 \\ H_c &\rightarrow 0,55 \\ T &\rightarrow 0,55 - 0,3 = 0,25 \end{aligned}$$

De nouveau, l'intervalle  $[L_c; H_c] = [0,3; 0,55]$  a été divisé en 5 sous-intervalles selon les probabilités  $p_k$ . Etant donné que le second symbole de la séquence est  $d$ , c'est l'intervalle  $[L_4; H_4] = [0,4875; 0,525]$  qui a été retenu. Les valeurs  $L_c, H_c$  et  $T$  ont alors été mises à jour selon

$$\begin{aligned} L_c &\rightarrow 0,4875 \\ H_c &\rightarrow 0,525 \\ T &\rightarrow 0,525 - 0,4875 = 0,0375 \end{aligned}$$

On continue ainsi jusqu'au dernier symbole  $a$  de la séquence à coder. Ceci nous fournit l'intervalle final  $[L_c; H_c] = [0, 514875 ; 0, 5152875]$ . N'importe quel nombre réel de cet intervalle code la séquence  $bdcea$ . Voyons maintenant comment représenter ce nombre réel sous la forme d'un code binaire.

### Calcul du code binaire issus du codage arithmétique

Comme nous l'avons vu plus haut, nous recherchons l'extension binaire d'un nombre réel, représenté par  $0, \alpha_1\alpha_2\dots\alpha_m$ , qui se trouve dans l'intervalle final  $[L_c, H_c]$  fourni par l'algorithme décrit plus haut. La séquence  $\alpha_1\alpha_2\dots\alpha_m$  correspondra au code binaire recherché.

Pour illustrer nos propos, repartons de l'exemple, commencé plus haut, pour lequel  $[L_c; H_c] = [0, 514875 ; 0, 5152875]$ . La table 3.4 illustre la recherche d'un nombre réel compris dans cet intervalle. La technique consiste à ajouter successivement un bit supplémentaire  $\alpha_i$  à l'expansion binaire jusqu'à ce que le nombre réel correspondant soit compris dans l'intervalle.

$i$	$2^{-i}$	si $\alpha_i = 0$	si $\alpha_i = 1$	$\alpha_i$ choisi	Valeur actuelle
1	0,5	0	0,5	1	0,5
2	0,25	0,5	0,75	0	0,5
3	0,125	0,5	0,625	0	0,5
4	0,0625	0,5	0,5625	0	0,5
5	0,03125	0,5	0,53125	0	0,5
6	0,015625	0,5	0,515625	0	0,5
7	0,0078125	0,5	0,5078125	1	0,5078125
8	0,00390625	0,5078125	0,51171875	1	0,51171875
9	0,001953125	0,51171875	0,513671875	1	0,513671875
10	0,0009765625	0,513671875	0,5146484375	1	0,5146484375
11	0,00048828125	0,5146484375	0,51513671875	1	0,51513671875

TABLE 3.4 – Recherche de l'expansion binaire d'un nombre réel compris dans l'intervalle  $[0, 514875 ; 0, 5152875]$ .

Nous obtenons ainsi le nombre  $0,51513671875 = 0,10000011111$  compris dans l'intervalle considéré. Le codage binaire ainsi généré pour la séquence de 5 symboles  $bdcea$  est

$$10000011111$$

et comporte 11 bits. Nous avons donc un nombre moyen de bits par symbole, **pour cette séquence**, égal à

$$\overline{R} = \frac{11}{5} = 2,2 \text{ bits/symbole}$$

### Algorithme de décodage

La première chose à faire pour décoder le message codé  $\alpha_1\alpha_2\dots\alpha_m$  est de calculer, au moyen de la formule (3.33), le nombre réel issus de l'extension binaire

$$r_c = 0, \alpha_1\alpha_2\dots\alpha_m$$

Ensuite, il faut appliquer les étapes suivantes :

1. On initialise un premier intervalle  $[L_c; H_c] = [0; 1]$ . La taille  $T$  de cet intervalle est  $T = H_c - L_c = 1$ . Le symbole en cours de décodage est initialisé à  $s_1$ , c'est-à-dire  $i = 1$ .
2. Cet intervalle  $[L_c; H_c]$  est partitionné en  $K$  sous-intervalles  $[L_k; H_k]$  proportionnellement aux probabilités  $p_k$  des symboles de la source. Nous avons donc

$$L_k = L_c + T \sum_{j=1}^{k-1} p_j \quad \text{et} \quad H_k = L_c + T \sum_{j=1}^k p_j$$

pour tout  $k = 1, 2, \dots, K$ . A chacun de ces sous-intervalles est associé le symbole  $x_k$  correspondant.

3. On détermine le sous-intervalle  $[L_k; H_k]$  dans lequel se trouve la valeur  $r_c$ . Ce sous-intervalle détermine le symbole  $x_k$  correspondant au symbole  $s_i$  en cours de décodage. Ensuite, on met à jour l'intervalle  $[L_c; H_c]$  de la manière suivante

$$L_c \rightarrow L_k \quad \text{et} \quad H_c \rightarrow H_k$$

et la taille  $T$  est alors mise à jour en fonction des nouvelles valeurs de  $L_c$  et  $H_c$ .

4. On répète les étapes 2 et 3 pour le symbole  $s_i$  suivant ( $i$  est incrémenté de 1) et on répète ces étapes jusqu'au dernier symbole  $s_n$  à décoder.

### Exemple

Nous allons donc décoder la séquence 10000011111 codée plus haut. Pour cela, nous devons connaître les symboles de la source, ainsi que leur probabilités respectives. On commence par calculer le nombre

$$r_c = 0,10000011111 = 0,51513671875$$

La construction des intervalles et sous-intervalles se fait exactement comme lors de l'encodage. On détermine ainsi que  $r_c$  se trouve dans l'intervalle  $[L_2; H_2] = [0,3 ; 0,55]$ . Le premier symbole  $s_1$  de la séquence décodée est donc égal à  $b$ . La mise à jour de l'intervalle  $[L_c; H_c]$  conduit donc à  $[L_c; H_c] = [0,3 ; 0,55]$ . Après construction des nouveaux sous-intervalles, on observe que  $r_c$  se trouve dans l'intervalle  $[0,4875 ; 0,525]$  et on en déduit que le second symbole  $s_2$  de la séquence décodée est  $d$ . On poursuit ainsi jusqu'au décodage complet du message.

### Remarques

On peut remarquer que, dans le codage arithmétique, plus le message à coder est long, plus la longueur des intervalles considérés se réduit. Il faut donc travailler en arithmétique de précision croissante, rapidement prohibitive et pouvant atteindre la précision machine. De plus, il faut attendre que le message soit entièrement codé avant de pouvoir émettre le premier bit de son code. Ces différents défauts ont mené au développement de variantes de cette technique, comme le codage arithmétique par intervalles, mais dont l'étude sort du cadre de cette introduction.

### 3.4.3 Codage par dictionnaire : méthode de LEMPEL-ZIV

Les algorithmes de codage par dictionnaire sont des méthodes, qui, n'ayant aucune statistique sur la source de données, vont se constituer à la volée un dictionnaire où figurent

les groupes de mots qui se trouvent répétés dans le document à compresser. Le fait que la production soit spécifique aux données à traiter font de ces algorithmes un procédé capable d'adaptation à un grand nombre de besoins.

L'algorithme de base a été développé en 1977-1978 par Abraham LEMPEL et Jacob ZIV et est l'algorithme dit LZ77 et LZ78. Une amélioration de ces algorithmes a été proposée en 1984 par Welsh, ce qui donne l'algorithme LZW. Ce dernier fournit des taux de compression de 30% à 40% et il est à la base de nombreuses utilisations :

- La grande majorité des algorithmes de compression : GZIP, PKZIP, WINZIP, ...
- Les formats GIF (Graphic Interchange Format) et TIFF (Tagged Image File Format) de compression d'images,
- Les fichiers audio MOD,
- La compression de données pour la transmission sur modem norme V42 bis.

C'est l'algorithme LZW qui est décrit ci-dessous.

### Algorithme de compression

Le principe de l'algorithme de codage est le suivant :

1. On dispose d'une table de correspondance (le dictionnaire) initialisée avec les codes ASCII numérotés de 0 à 255<sup>2</sup>.
2. Sont ajoutés à cette table 2 caractères de contrôles (256 et 257) réservés. L'utilité de ces deux caractères sera expliquée plus loin.
3. Au fur et à mesure de l'apparition de blocs de caractères, ceux-ci sont ajoutés au dictionnaire et un code (valeur supérieure à 257) leur est attribué. En même temps que ces chaînes sont ajoutées au dictionnaire, le premier caractère est envoyé en sortie du codeur. Chaque fois qu'une chaîne déjà rencontrée est lue, la chaîne la plus longue déjà rencontrée est déterminée, et le code correspondant à cette chaîne est envoyée sur la sortie.
4. Tous les éléments du message sont codés sur le même nombre de bits : 9bits, 10 bits, ... Ce nombre détermine la taille du dictionnaire. Par exemple, sur 10 bits, le dictionnaire comporte 1024 entrées dont les 258 premières sont réservées par le code ASCII de base et les deux caractères de contrôle.

La figure 3.10 illustre cet algorithme de compression. Sur cette figure, *c* représente le caractère lu à l'entrée de l'encodeur, *w* un mot constitué de plusieurs caractères et *w+c* la concaténation de *w* et *c*.

### Exemple

Nous allons à présent illustrer cet algorithme sur la chaîne de 24 caractères suivantes :

TOBEORNOTTOBEORTOBEORNOT

---

2. Pour information, la table des 256 codes ASCII est fournie à l'annexe 3.6.3

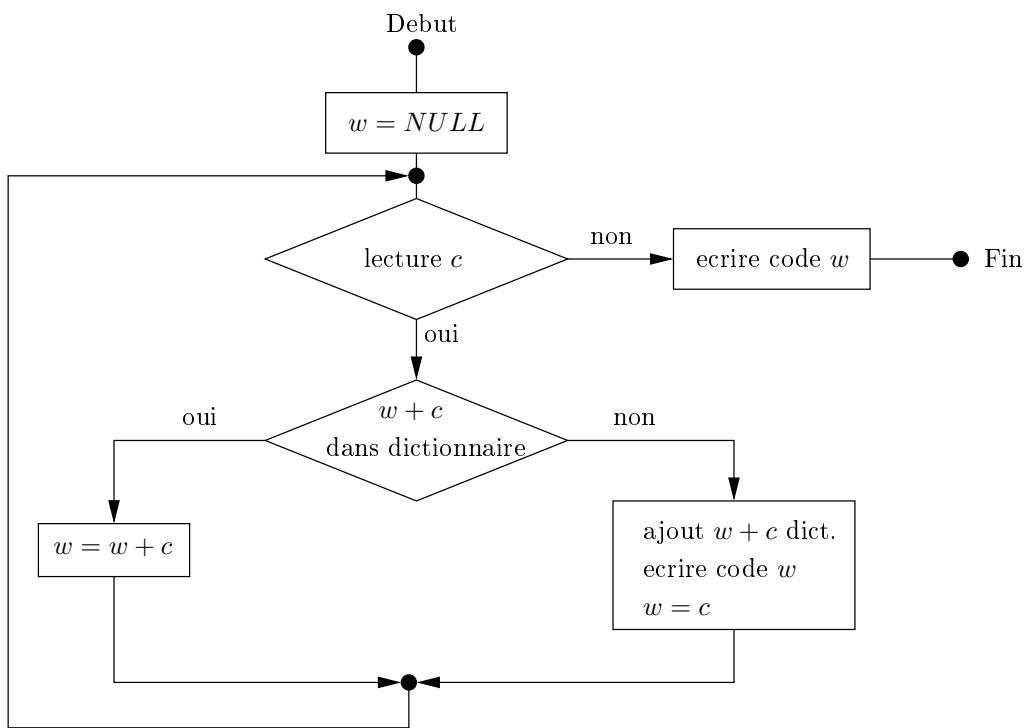


FIGURE 3.10 – Algorithme de compression LZW.

La table 3.5 montre l'avancement de l'algorithme au fur et à mesure de la compression de cette chaîne. Après la compression, nous obtenons la séquence de codes sur 9 bits suivantes :

TOBEORNOT<258><260><262><267><261><263><265>

Elle nécessite  $16 \times 9 = 144$  bits de stockage au lieu des  $24 \times 8 = 192$  bits de la chaîne originale. Cela correspond à un taux de compression de

$$\tau = \frac{192}{144} = 1,333$$

ou encore de 75 %, sur cet exemple adapté mais il est évident que l'efficacité de cet algorithme sera d'autant plus grande que les données à compresser seront de taille importante.

### Algorithme de décompression

Un des avantages de l'algorithme LZW est qu'il ne nécessite pas la transmission de la table de codage. En effet, l'algorithme de décompression a seulement besoin du texte compressé en entrée. En effet, il reconstruit une table chaînes de caractères / code (le dictionnaire) identique à mesure qu'il régénère le texte original.

La figure 3.11 fournit l'organigramme de l'algorithme de décompression. Sur celle-ci,  $c$  représente le caractère lu à l'entrée du décodeur,  $t$  et  $w$  des mots constitués de plusieurs caractères et  $w + t[0]$  la concaténation de  $w$  et du premier caractère de la chaîne  $t$ .

### Exemple

La table 3.6 présente le résultat de l'algorithme de décompression sur la séquence précédemment compressée dans l'exemple ci-dessus.

$w$	entrée $c$	$w + c$	sortie	ajout au dictionnaire
	T	T		
T	O	TO	T	TO=<258>
O	B	OB	O	OB=<259>
B	E	BE	B	BE=<260>
E	O	EO	E	EO=<261>
O	R	OR	O	OR=<262>
R	N	RN	R	RN=<263>
N	O	NO	N	NO=<264>
O	T	OT	O	OT=<265>
T	T	TT	T	TT=<266>
T	O	TO		
TO	B	TOB	<258>	TOB=<267>
B	E	BE		
BE	O	BEO	<260>	BEO=<268>
O	R	OR		
OR	T	ORT	<262>	ORT=<269>
T	O	TO		
TO	B	TOB		
TOB	E	TOBE	<267>	TOBE=<270>
E	O	EO		
EO	R	EOR	<261>	EOR=<271>
R	N	RN		
RN	O	RNO	<263>	RNO=<272>
O	T	OT		
OT			<265>	

TABLE 3.5 – Illustration de l'algorithme de compression LZW.

### Codes spéciaux

Au fur et à mesure de la compression, le nombre d'entrées dans le dictionnaire augmente et la limite du nombre de bits peut être atteinte. Le codeur peut alors augmenter le nombre de bits par code. Par exemple, avec 9 bits, nous disposons de 512 - 256 (codes ASCII de base) - 2 (codes spéciaux) = 254 entrées libres dans le dictionnaire. La première augmentation du nombre de bits sera nécessaire à l'ajout de la 255ème entrée. Le décodeur procèdera de même lorsqu'il ajoutera une 255ème entrée au dictionnaire. L'augmentation se reproduit à chaque passage des puissances de 2.

Afin d'éviter au dictionnaire de grandir démesurément, il peut être prévu un nombre d'entrées maximum dans le dictionnaire. Lorsque ce nombre est atteint, le codeur insère un code spécial et vide complètement le dictionnaire. Le processus de codage recommence alors comme pour le premier caractère. Le décodeur purgera son dictionnaire lorsqu'il rencontrera ce code spécial. Ce code spécial est la valeur <256> pour la compression LZW des images TIFF, et est, comme on l'a introduit plus haut, exclu des entrées "normales" des dictionnaires.

D'autres codes spéciaux peuvent exister selon les implémentations. Par exemple, pour signaler la fin du flux de données (code <257> pour la compression LZW des images TIFF), pour effectuer un nettoyage partiel du dictionnaire, etc...

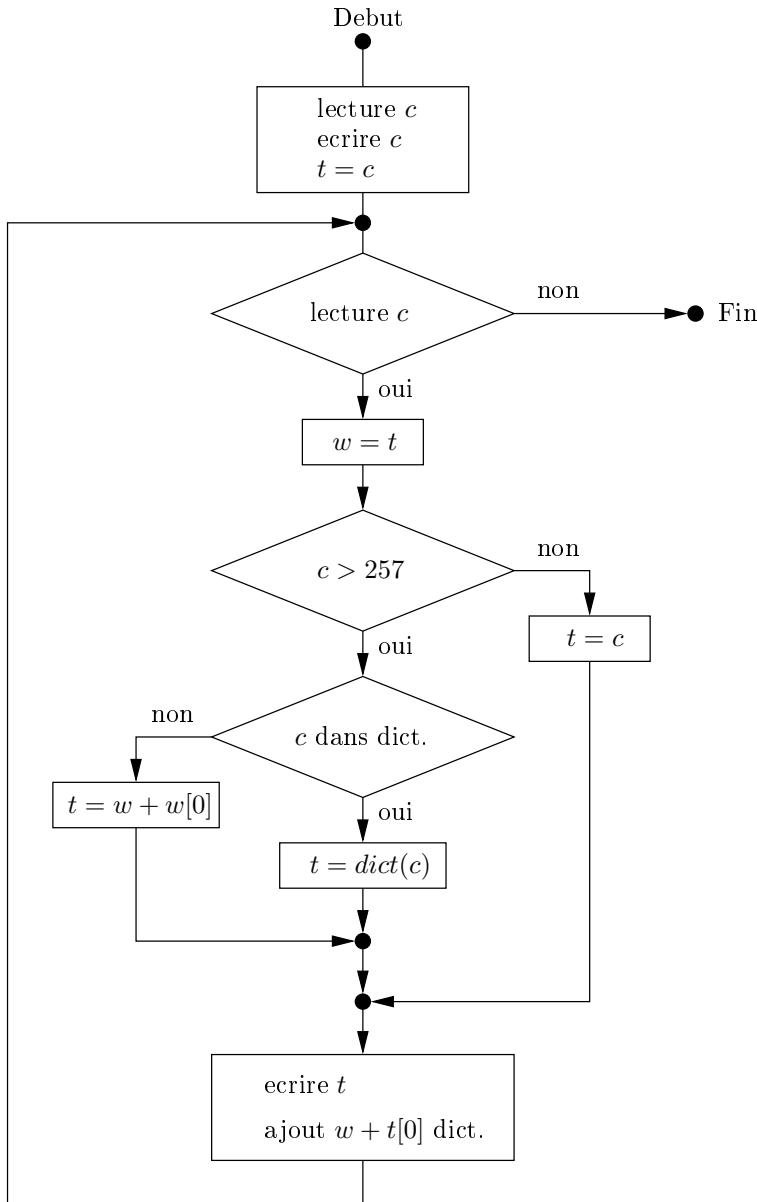


FIGURE 3.11 – Algorithme de décompression LZW.

### 3.4.4 Codage par répétition : méthode RLE (RLC)

Le plus typique et le plus simple de ces algorithmes est la méthode RLE (Run Length Encoding) ou encore RLC (Run Length Coding). Elle consiste à repérer une donnée qui a des apparitions consécutives fréquentes. Dans ce cas, elle sera remplacée par deux indications :

- Un chiffre qui indique le nombre de répétitions.
- La donnée elle-même.

#### Exemples.

Elle → E2le

donnée → do2née

Ces exemples montrent que, dans le cas du texte, la méthode risque d'être peu efficace. Pour qu'elle le devienne, il faut que les données concernées aient si possible un grand nombre de

entrée $c$	$w$	$t$	$w + t[0]$	sortie	ajout au dictionnaire
T		T		T	
O	T	O	TO	O	TO=<258>
B	O	B	OB	B	OB=<259>
E	B	E	BE	E	BE=<260>
O	E	O	EO	O	EO=<261>
R	O	R	OR	R	OR=<262>
N	R	N	RN	N	RN=<263>
O	N	O	NO	O	NO=<264>
T	O	T	OT	T	OT=<265>
<258>	T	TO	TT	TO	TT=<266>
<260>	TO	BE	TOB	BE	TOB=<267>
<262>	BE	OR	BEO	OR	BEO=<268>
<267>	OR	TOB	ORT	TOB	ORT=<269>
<261>	TOB	EO	TOBE	EO	TOBE=<270>
<263>	EO	RN	EOR	RN	EOR=<271>
<265>	RN	OT	RNO	OT	RNO=<272>

TABLE 3.6 – Illustration de l’algorithme de décompression LZW.

répétitions successives. Les données qui conviennent à ce type d’algorithme sont les images où il y a de grandes répétitions entre pixels voisins. Cela est applicable pour des images binaires mais aussi des images en niveaux de gris (et même couleur) si elles sont séparées en différents plans de bits.

Prenons le cas simple d’une image binaire où nous avons la convention suivante : 1 représente un pixel blanc et 0 représente un pixel noir. La succession suivante de pixels

11111111000001111110000

sera représentée par

81 50 61 40

Nous pouvons économiser quelques bits supplémentaires sachant qu’il s’agit d’une alternance de 1 et de 0, il suffit de préciser la nature du premier bit de la liste, soit dans l’exemple

81 5 6 4

Dans le cas des couleurs, celles-ci sont représentées par un nombre entier (souvent compris entre 0 et 255 pour chaque composante). Afin de ne pas confondre le nombre de répétitions avec le code de la couleur, il est nécessaire d’ajouter un caractère séparateur entre les différentes zones. Le caractère séparateur usuel pour RLE est le “#” et un autre caractère séparateur entre le nombre d’occurrences et le code de couleur lui-même qui est en général l’espace. Par exemple, la suite de pixels en niveau de gris suivante

8 8 8 8 8 8 8 24 24 24 24 24 24 24 67 67 67 67

sera codée par

#8 8#7 24#4 67#

De grande simplicité, cet algorithme est toujours utilisé dans

- Le format d'images PCX.
- Les télécopies (fax) norme CCITT groupe 3 et 4.

### 3.5 Exercices

1. Soit la table de contingences de 2 sources de données  $X$  et  $Y$  fournie à la table 3.7.

	$y_1$	$y_2$
$x_1$	$\frac{1}{3}$	$\frac{1}{3}$
$x_2$	0	$\frac{1}{3}$

TABLE 3.7 – Table de contingences de l'exercice 1.

Calculer

- (a)  $H(X)$  et  $H(Y)$
  - (b)  $H(X, Y)$
  - (c)  $H(X|Y)$  et  $H(Y|X)$
  - (d)  $I(X, Y)$
  - (e) Dessiner un diagramme de VENNE qui résume la situation.
2. Soit la source de données  $X$  dont les  $K = 7$  symboles possibles  $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$  ont des probabilités respectives  $p_1 = 0,49$ ,  $p_2 = 0,26$ ,  $p_3 = 0,12$ ,  $p_4 = 0,04$ ,  $p_5 = 0,04$ ,  $p_6 = 0,03$  et  $p_7 = 0,02$ .
    - (a) Calculer l'entropie  $H(X)$  de la source de données.
    - (b) Trouver un codage de HUFFMAN adapté et calculer la longueur moyenne correspondante.
    - (c) Calculer l'efficacité du codage obtenu.
    - (d) Comparer cette longueur moyenne aux bornes prédites par le théorème de SHANNON.
  3. Lesquels parmi les codes suivants ne peuvent pas être un codage de HUFFMAN ? Justifier.
    - (a)  $\{0, 10, 11\}$
    - (b)  $\{00, 01, 10, 110\}$
    - (c)  $\{01, 10\}$
  4. Soit la source discrète  $X$  de  $K = 5$  symboles  $\{-2, -1, 0, +1, +2\}$  qui définit une série de 5 vecteurs de mouvements verticaux possibles (vecteurs utilisés en codage vidéo). La probabilité de production de la source est  $\{0, 1; 0, 2; 0, 4; 0, 2; 0, 1\}$ . En utilisant le codage arithmétique, coder la séquence  $\{0, -1, 0, 2\}$ .

5. Soit la chaîne de caractères suivante

### LES PAGES D'IMAGES D'ORAGES

- (a) Compresser cette chaîne à l'aide de l'algorithme LZW. Décrire le processus à l'aide d'une table.
- (b) Sachant que chaque symbole du message de départ est codé sur 8 bits, calculer le taux de compression obtenu.
- (c) Décompresser les données compressées, en décrivant le processus à l'aide d'une table.

## 3.6 Annexes

### 3.6.1 Maximisation de l'entropie d'une source de $K$ symboles

Le problème est donc de trouver les valeurs  $p_k$  ( $k = 1, 2, \dots, K$ ) qui maximise l'expression

$$H(S) = - \sum_{k=1}^K p_k \log p_k$$

sous la contrainte  $p_1 + p_2 + \dots + p_K = 1$ . Pour cela commençons par démontrer le lemme suivant.

**Lemme (Inégalité de GIBBS).** Soient  $q_k$ , avec  $0 \leq q_k \leq 1$  pour  $k = 1, 2, \dots, K$  tels que  $q_1 + q_2 + \dots + q_K = 1$ . Alors

$$\sum_{k=1}^K p_k \log \left( \frac{p_k}{q_k} \right) \geq 0$$

**Preuve.** Si on tient compte que  $\ln x \leq x - 1$  pour tout  $x \in ]0, +\infty[$ , nous pouvons écrire

$$\begin{aligned} - \sum_{k=1}^K p_k \log \left( \frac{q_k}{p_k} \right) &= - \frac{1}{\ln 2} \sum_{k=1}^K p_k \ln \left( \frac{q_k}{p_k} \right) \\ &\geq - \frac{1}{\ln 2} \sum_{k=1}^K p_k \left( \frac{q_k}{p_k} - 1 \right) \\ &= - \frac{1}{\ln 2} \sum_{k=1}^K (q_k - p_k) \\ &= 0 \end{aligned}$$

d'où le résultat. ■

A partir de ce lemme, si nous choisissons  $q_k = \frac{1}{K}$  pour tout  $k = 1, 2, \dots, K$ , nous obtenons

$$\begin{aligned} \sum_{k=1}^K p_k \log (K p_k) \geq 0 &\iff \sum_{k=1}^K p_k \log K + \sum_{k=1}^K p_k \log p_k \geq 0 \\ &\iff - \sum_{k=1}^K p_k \log p_k \leq \log K \end{aligned}$$

Donc, quels que soient les  $p_k$ , l'entropie  $H(S)$  ne pourra jamais dépasser  $\log K$ , l'égalité ayant lieu lorsque  $p_k = \frac{1}{K}$  pour tout  $k = 1, 2, \dots, K$ , c'est-à-dire lorsque tous les symboles sont équiprobables.

### 3.6.2 Expansion binaire d'un nombre réel

Nous savons comment un nombre entier est codé en binaire. Par contre, cela devient moins clair lorsqu'il s'agit d'un nombre réel. Par exemple, prenons le nombre binaire 101 qui est égal, en base 10, à

$$101 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 5$$

En toute généralité, nous pouvons écrire

$$a_n a_{n-1} \dots a_1 a_0 = a_n \times 2^n + a_{n-1} \times 2^{n-1} + \dots + a_1 \times 2^1 + a_0 \times 2^0$$

où les  $a_i$  valent 1 ou 0. Nous pouvons même généraliser cette notion aux cas des nombres binaires avec virgules en utilisant des puissances négatives de 2 :

$$a_n \dots a_1 a_0, \alpha_1 \alpha_2 \dots \alpha_m = a_n \times 2^n + \dots + a_1 \times 2^1 + a_0 \times 2^0 + \alpha_1 \times 2^{-1} + \alpha_2 \times 2^{-2} + \dots + \alpha_m \times 2^{-m} \quad (3.32)$$

qui porte le nom d'expansion binaire d'un nombre réel, car pouvant représenter un nombre réel. Voici un exemple simple :

$$101,011 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3} = 5,375$$

Dans le cas des nombres réels compris entre 0 et 1, nous pouvons utiliser

$$0, \alpha_1 \alpha_2 \dots \alpha_m = \alpha_1 \times 2^{-1} + \alpha_2 \times 2^{-2} + \dots + \alpha_m \times 2^{-m} \quad (3.33)$$

### 3.6.3 Table ASCII

Les 128 premiers caractères de la table des codes ASCII (American Standard Code for Information Interchange) de base sont fournis à la figure 3.12. La figure 3.13 présente, en plus, les 128 codes ASCII étendus.

## 3.7 Références

1. Théorie de l'information - Codage de source. G. BINET. Université de Caen.
2. Mathématiques Appliquées. R. FOURNEAU. Haute Ecole de la Province de Liège.
3. Principes généraux de codage entropique d'une source. J. MVOGO NGONO.
4. Théorie de l'information et du codage. L. WEHENKEL. Université de Liège.
5. Télécommunications et ordinateurs. M. VAN DROOGENBROECK. Université de Liège
6. Site Wikipédia.

# ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEG. ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[END OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	{	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	-	127	7F	[DEL]

FIGURE 3.12 – Table des 128 premiers codes ASCII de base.

Table des codes ASCII

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	00	Null	32	20	Space	64	40	8	96	60	`
1	01	Start of heading	33	21	!	65	41	A	97	61	a
2	02	Start of text	34	22	"	66	42	B	98	62	b
3	03	End of text	35	23	#	67	43	C	99	63	c
4	04	End of transmit	36	24	\$	68	44	D	100	64	d
5	05	Enquiry	37	25	%	69	45	E	101	65	e
6	06	Acknowledge	38	26	&	70	46	F	102	66	f
7	07	Audible bell	39	27	'	71	47	G	103	67	g
8	08	Backspace	40	28	(	72	48	H	104	68	h
9	09	Horizontal tab	41	29	)	73	49	I	105	69	i
10	0A	Line feed	42	2A	*	74	4A	J	106	6A	j
11	0B	Vertical tab	43	2B	+	75	4B	K	107	6B	k
12	0C	Form feed	44	2C	,	76	4C	L	108	6C	l
13	0D	Carriage return	45	2D	-	77	4D	M	109	6D	m
14	0E	Shift out	46	2E	.	78	4E	N	110	6E	n
15	0F	Shift in	47	2F	/	79	4F	O	111	6F	o
16	10	Data link escape	48	30	0	80	50	P	112	70	p
17	11	Device control 1	49	31	1	81	51	Q	113	71	q
18	12	Device control 2	50	32	2	82	52	R	114	72	r
19	13	Device control 3	51	33	3	83	53	S	115	73	s
20	14	Device control 4	52	34	4	84	54	T	116	74	t
21	15	Neg. acknowledge	53	35	5	85	55	U	117	75	u
22	16	Synchronous idle	54	36	6	86	56	V	118	76	v
23	17	End trans. block	55	37	7	87	57	W	119	77	w
24	18	Cancel	56	38	8	88	58	X	120	78	x
25	19	End of medium	57	39	9	89	59	Y	121	79	y
26	1A	Substitution	58	3A	:	90	5A	Z	122	7A	z
27	1B	Escape	59	3B	;	91	5B	{	123	7B	{
28	1C	File separator	60	3C	<	92	5C	\	124	7C	
29	1D	Group separator	61	3D	=	93	5D	]	125	7D	}
30	1E	Record separator	62	3E	>	94	5E	^	126	7E	~
31	1F	Unit separator	63	3F	?	95	5F	-	127	7F	[DEL]

D'après [http://www.uninova.pt/~atm/SLII/table\\_ascii.htm](http://www.uninova.pt/~atm/SLII/table_ascii.htm)

FIGURE 3.13 – Table des 256 codes ASCII.



# Troisième partie

## Traitement d'images



# Chapitre 4

## Analyse et traitement d'images

### 4.1 Introduction

Une image est avant tout un signal bidimensionnel (2D). Le traitement d'images entre donc dans une catégorie plus vaste qui est le traitement du signal en général. Dès lors, il est normal de retrouver des similitudes entre le traitement du signal 1D et le traitement d'images (ou traitement du signal 2D) : convolution, filtrage, transformées, ... Néanmoins, décrire une image n'est pas chose facile. Plusieurs approches existent :

- considérer une image comme une fonction bidimensionnelle  $f(x, y)$  où  $x$  et  $y$  sont les deux paramètres du signal, représentant les coordonnées d'un point (on dit plutôt pixel) de l'image,
- considérer une image comme une matrice  $f(m, n)$  ( $m = 0, \dots, M-1$  et  $n = 0, \dots, N-1$ ) contenant  $M \times N$  valeurs numériques,
- ou encore considérer une image comme un ensemble d'objets (ou de formes) présents sur un arrière plan.

De plus, que représente exactement la valeur de  $f$ ? Cela dépend en plus du type d'image considéré... Commençons donc par une classification simplifiée d'images les plus rencontrées dans la pratique.

#### 4.1.1 Type d'images

Dans le contexte de cette introduction, nous nous limiterons à 3 grandes catégories d'images (voir figure 4.1) :

- les images en niveaux de gris pour lesquelles la fonction ou la matrice  $f$  représente l'intensité lumineuse ou encore la *luminance*.
- les images couleurs, sur lesquelles nous revenons ci-dessous,
- les images binaires, souvent représentées avec deux couleurs uniques ; en général blanc et noir mais ce n'est pas une obligation. Ce type d'image est généralement obtenu après traitement d'une image en niveau de gris ou en couleurs dans le but d'en extraire de l'information pertinente (recherche de formes ou d'objets, segmentation, contrôle industriel, ...).

La représentation des images couleurs est toutefois plus complexe que les deux autres. Il est en effet nécessaire de caractériser l'information "couleur". Pour cela, il existe différents systèmes de couleurs comme par exemple

- le système RVB (ou RGB en anglais) qui est le plus commun et qui est un système additif de couleurs utilisé pour la représentation des couleurs sur écran. Dans ce système, chaque couleur est représentée par trois composantes : le rouge (R), le vert (V) et le bleu

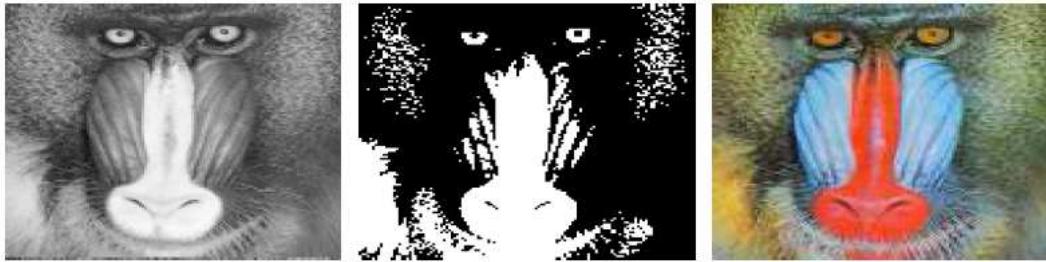


FIGURE 4.1 – Différents types d'image : En niveaux de gris, binaire et en couleurs.

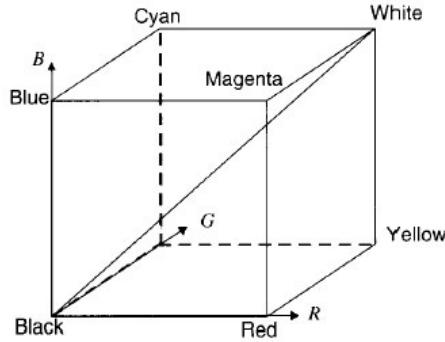


FIGURE 4.2 – Addition des composantes Rouge, Vert, bleu pour reproduire une couleur  $G$ .

(B). Une couleur  $G$  est alors obtenue en additionnant les trois composantes R,V,B qui la composent. En termes simplifiés, on ajoute des composantes au noir (origine du système) pour obtenir une couleur. Voir figure 4.2.

- le système CMJ (ou CMY en anglais) qui est un système soustractif de couleurs utilisé dans les systèmes d'impression. Dans ce système, chaque couleur est représentée par trois composantes : le cyan (C), la magenta (M) et le jaune (J). De manière imagée, pour obtenir une couleur dans ce système, on part du blanc et on retire les composantes nécessaires.
- les systèmes YIQ, YUV,  $YC_bC_r$ , ...

Remarquons que pour chaque système de couleurs, il faut 3 composantes pour représenter une couleur. Il existe des transformations matricielles permettant de passer d'un système à un autre. Mais revenons à notre image  $f$ . Pour un pixel  $(x, y)$ ,  $f$  doit donc contenir les trois informations liées à la couleur correspondante. On peut en effet décomposer une image couleur  $f(x, y)$  en trois images, on dira plutôt *plans de couleurs* :

$$f_R(x, y), f_V(x, y), f_B(x, y)$$

si on considère la décomposition RVB à laquelle nous nous limiterons dans ces notes. Une image couleur est dès lors représentée par trois fonctions 2D. La figure 4.3 illustre la décomposition RVB d'une image couleurs en ses trois plans de couleurs.

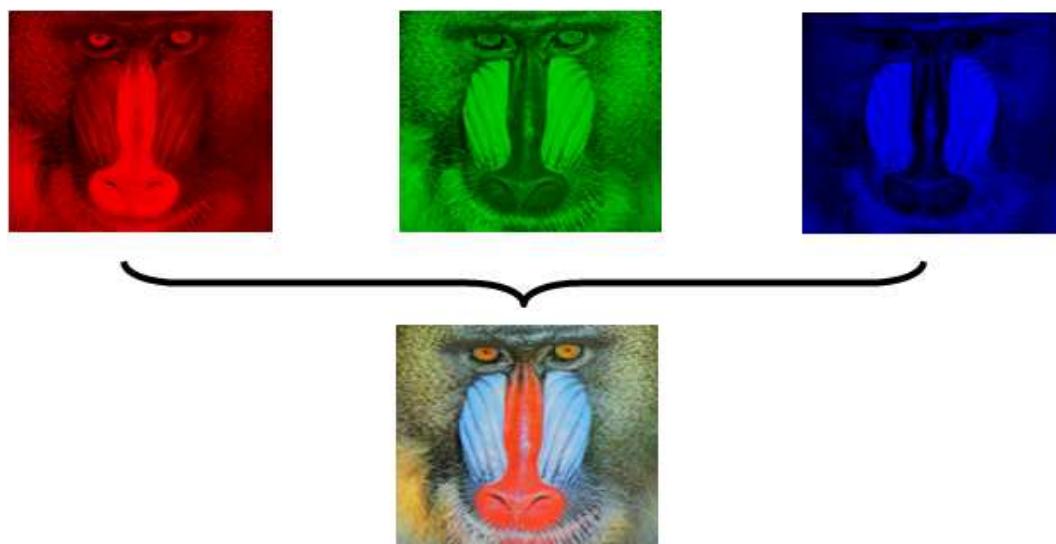
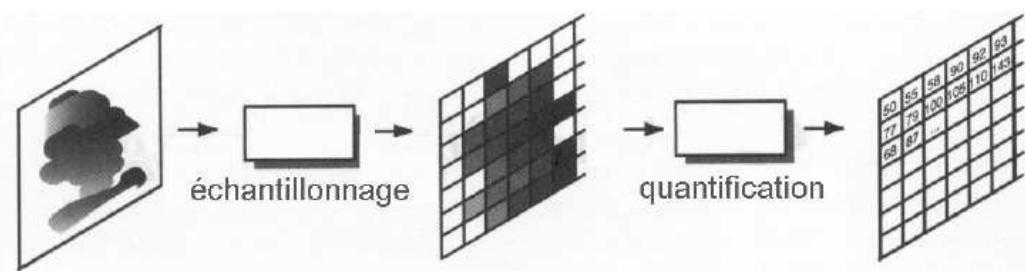


FIGURE 4.3 – Décomposition d'une image couleurs en ses plans de couleurs RVB.



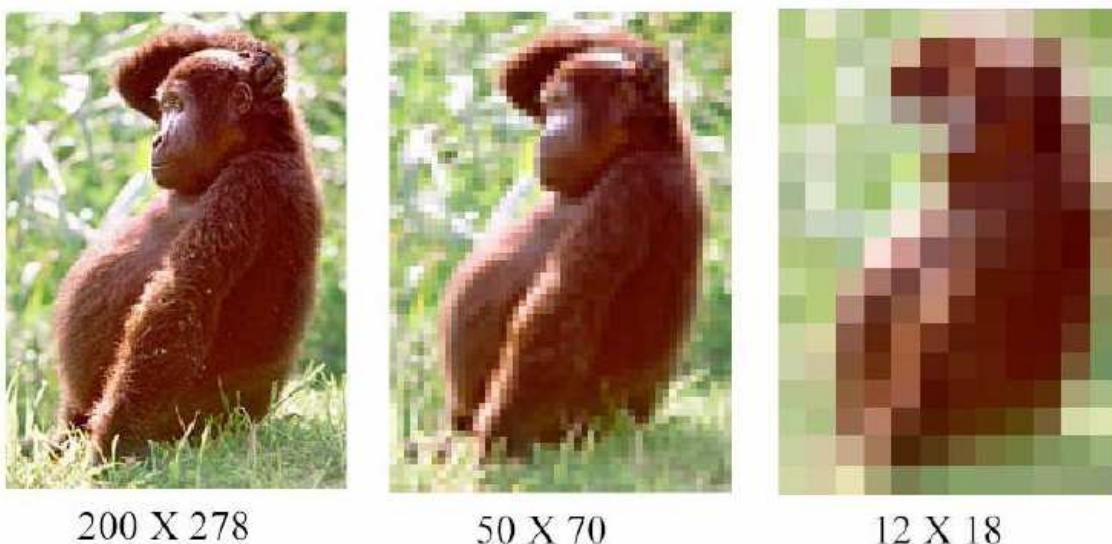


FIGURE 4.5 – Résolution spatiale.

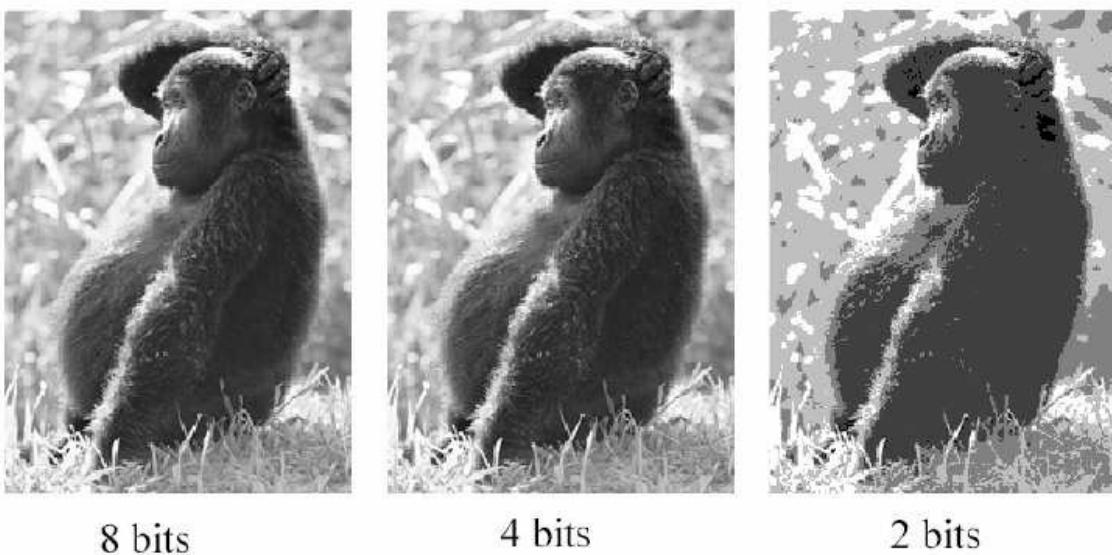


FIGURE 4.6 – Résolution des tons de gris.

### Résolution des tons de gris (ou de couleurs)

La résolution des tons de gris est directement liée à la plus petite nuance de gris discernable dans l'image. La figure 4.6 illustre la même image pour trois résolutions de niveaux de gris différentes. Cette notion est intimement liée à celle de *codage* des niveaux de gris. Plus on utilisera de bits pour représenter un niveau de gris, meilleure sera la résolution des tons de gris. Par exemple, si on code une couleur sur 8 bits, on pourra ainsi représenter 256 niveaux de gris différents, ce qui est déjà suffisant pour une visualisation satisfaisante par l'oeil humain.

Dans le cas des images couleurs, il faudra coder les trois composantes RVB. Par exemple, lorsque l'on parle d'une image couleurs 32 bits, cela signifie que l'on code chaque composante RVB sur 8 bits, ce qui permet de représenter  $256 \times 256 \times 256 = 16.777.216$  couleurs.

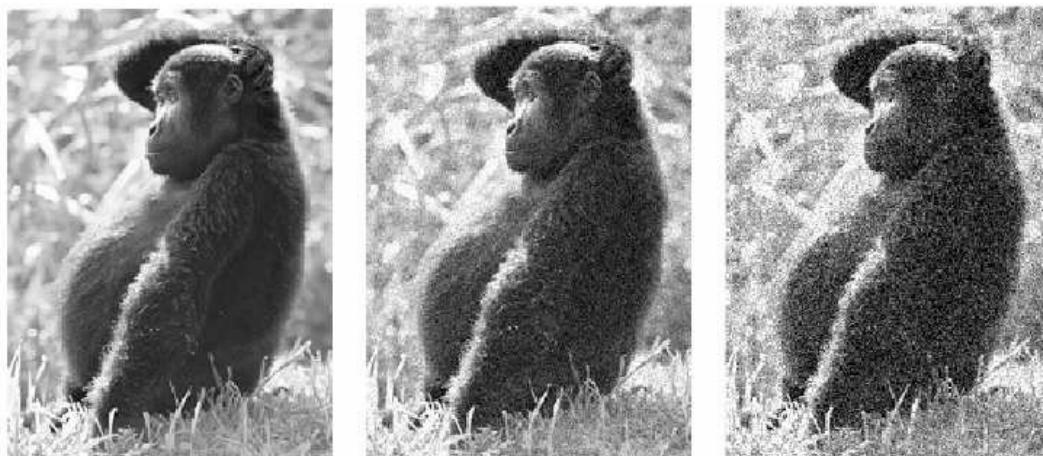


FIGURE 4.7 – Bruit dans les images.

### Bruit dans les images

Les systèmes d'acquisition numérique d'images entachent généralement les images obtenues de bruit et cela d'autant plus que les conditions d'éclairage sont mauvaises. La figure 4.7 illustre la même image avec des niveaux de bruit différents.

#### 4.1.3 Traitement d'images

Les applications et les techniques de traitement d'images sont nombreuses et variées. En voici une liste des plus fréquentes :

- Filtrage dans le but d'améliorer la qualité en supprimant le bruit, de supprimer certaines composantes indésirables, mettre en évidence les contours d'objets présents dans l'image... Ces techniques fournissent en sortie une image du même type que celle d'entrée mais avec les modifications ou altérations souhaitées. Elles peuvent être linéaires (convolution, FOURIER, ...) mais également non-linéaires comme celles basées sur la morphologie mathématique.
- Segmentation et reconnaissance de formes qui sont des domaines liés à la vision automatisée par ordinateur (ou robotique) et au contrôle industriel. Dans ce cas, le résultat du traitement peut être une image du même type que celle d'entrée (exemple : suppression du fond d'une image en laissant le personnage intact à des fins filmographiques) mais également un simple résultat numérique (exemple : comptage de particules présentes dans une image). Les techniques utilisées combinent souvent plusieurs opérations : transformation de l'image en niveaux de gris (ou couleurs) en images binaires, puis application de techniques morphologiques de traitement d'images binaires.
- Restauration et rehaussement afin de rendre à une image un aspect visuellement acceptable
- Contrôle de qualité par analyse des propriétés de l'image
- Suppression d'une inhomogénéité d'éclairage
- Reconnaissance de caractères

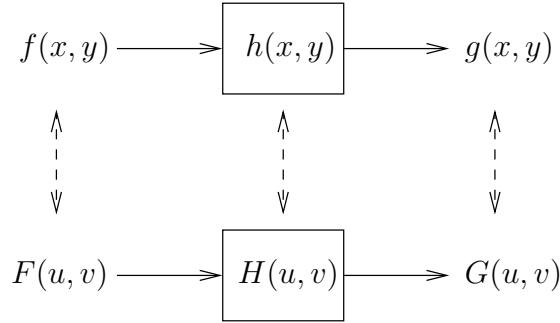


FIGURE 4.8 – Filtrage linéaire d'une image  $f(x,y)$ .

— ...

Dans cette introduction au traitement d'images, nous nous limiterons aux techniques et applications les plus courantes.

## 4.2 Traitements linéaires

Le filtrage linéaire d'une image (ou signal 2D) est la généralisation à deux dimensions du filtrage des signaux temporels que nous avons déjà étudié. À ce niveau, la variable temporelle  $t$  est remplacée par les deux variables spatiales  $x$  et  $y$ . La variable fréquentielle  $f$  est remplacée par les variables  $u$  et  $v$ . Dans ce contexte, la transformée de FOURIER constitue toujours un outil extrêmement intéressant. Pour son étude, nous renvoyons le lecteur à l'annexe consacrée à la transformée de FOURIER 2D.

Le filtrage linéaire d'une image peut être représenté à l'aide de la figure 4.8. Un filtre linéaire est donc toujours caractérisé par sa *réponse impulsionale*  $h(x,y)$  ou sa *fonction de transfert*  $H(u,v)$ .

Le filtrage de l'image peut donc se réaliser de deux manières différentes. La première possibilité est de le réaliser dans le domaine spatial en utilisant la convolution :

$$g(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\alpha, \beta) h(x - \alpha, y - \beta) d\alpha d\beta \quad (4.1)$$

où  $g(x,y)$  est l'image filtrée. La seconde possibilité est de passer par la transformée de FOURIER  $F(u,v)$  de l'image de départ  $f(x,y)$ . La transformée de FOURIER de l'image filtrée  $G(u,v)$  s'obtient alors par simple multiplication (dans le domaine fréquentiel) :

$$G(u,v) = H(u,v) F(u,v) \quad (4.2)$$

Toutefois, les images que l'on rencontre le plus souvent sont discrètes, c'est-à-dire qu'elles sont fournies sous la forme d'une matrice  $f(m,n)$  de pixels avec  $m = 0, \dots, M-1$  et  $n = 0, \dots, N-1$ . Il est donc nécessaire de recourir à la transformée de FOURIER *discrète* ou à la convolution *discrète*.

### 4.2.1 Transformée de FOURIER discrète et convolution discrète

Considérons une image donnée sous sa forme matricielle  $f(m,n)$  connue pour  $m = 0, \dots, M-1$  et  $n = 0, \dots, N-1$ .



FIGURE 4.9 – L'image *Lena* et le module de sa transformée de FOURIER discrète.

**Définition [Transformée de FOURIER 2D discrète] (DFT pour *Discrete Fourier Transform*)**. La transformée de FOURIER discrète de l'image  $f(m, n)$  est définie par

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \exp \left[ -j2\pi \left( \frac{mu}{M} + \frac{nv}{N} \right) \right] \quad (4.3)$$

avec  $u = 0, \dots, M - 1$  et  $v = 0, \dots, N - 1$ .

Il est possible de retrouver l'image de départ par transformée de FOURIER discrète inverse :

$$f(m, n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \exp \left[ j2\pi \left( \frac{mu}{M} + \frac{nv}{N} \right) \right] \quad (4.4)$$

avec  $m = 0, \dots, M - 1$  et  $n = 0, \dots, N - 1$ .

### Propriétés de périodicité intrinsèque de la DFT

La périodicité est une propriété importante de la transformée de FOURIER discrète. La transformée est définie par une matrice d'éléments  $F(u, v)$  pour  $u = 0, \dots, M - 1$  et  $v = 0, \dots, N - 1$ . Si on permet aux indices  $u$  et  $v$  de prendre d'autres valeurs, on obtient une transformée périodique et une image périodique :

$$\begin{aligned} F(u, -v) &= F(u, N - v) & F(-u, v) &= F(M - u, v) \\ f(-m, n) &= f(M - m, n) & f(m, -n) &= f(m, N - n) \end{aligned}$$

Plus généralement,

$$F(aM + u, bN + v) = F(u, v) \quad f(aM + m, bN + n) = f(m, n)$$

où  $a$  et  $b$  sont entiers. Ces propriétés ne sont pas étonnantes car elles résultent de l'échantillonnage de l'image analogique dans les deux plans.

La figure 4.9 montre une image ainsi que le module de sa transformée de FOURIER.

Une remarque est à faire au sujet de la visualisation de la transformée de FOURIER discrète. La composante fréquentielle  $F(0, 0)$ , appelée composante continue ou DC, se trouve en haut et à gauche de l'image alors qu'on a l'habitude de voir cette composante située au milieu de l'image. En utilisant les propriétés de périodicité, il est possible d'observer le module de la transformée de FOURIER d'une manière plus conventionnelle. Voir figure 4.10. Le spectre modifié par décalage de l'image Lena est donné à la figure 4.11.

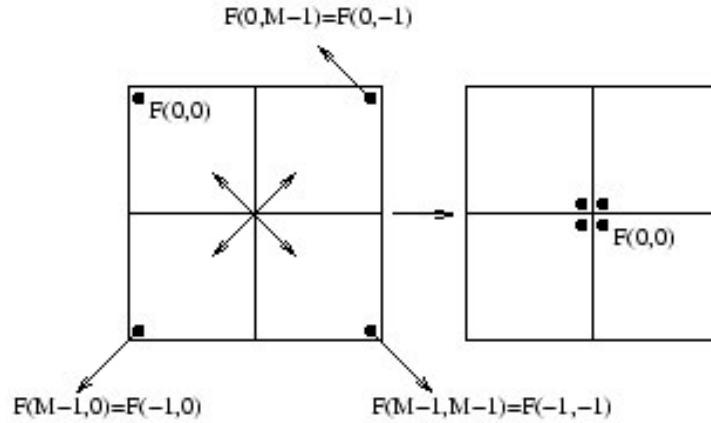


FIGURE 4.10 – Décalage du spectre pour arriver à centrer l'origine.

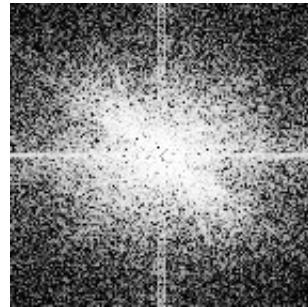


FIGURE 4.11 – Visualisation du spectre de l'image Lena après centrage de l'origine.

### Convolution discrète

La convolution discrète d'une image  $f(m, n)$  par la réponse impulsionnelle (discrète nécessairement)  $h$  est définie par

$$g(m, n) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f(k, l) h(m - k, n - l) \quad (4.5)$$

avec  $m = 0, \dots, M - 1$  et  $n = 0, \dots, N - 1$ . La réponse impulsionnelle  $h(i, j)$  doit donc être connue pour  $i = -(M - 1), \dots, (M - 1)$  et  $j = -(N - 1), \dots, (N - 1)$ .

Cette définition nous montre que la valeur d'un pixel de l'image filtrée est égale à une somme de tous les pixels de l'image originale pondérés par les coefficients  $h(i, j)$ . Chaque pixel de l'image filtrée dépend donc de *tous* les pixels de l'image originale. On dit alors que le traitement réalisé sur l'image est *global*. Dans le cas où la plupart des coefficients  $h(i, j)$  sont nuls et donc que la valeur d'un pixel de l'image filtrée ne dépend que de quelques pixels de l'image originale (et généralement situés dans le voisinage du pixel traité), on parle de *traitement local* de l'image.

#### 4.2.2 Traitement global

##### Filtre idéal

La notion de filtre idéal des images est similaire à celle rencontrée pour les signaux 1D.

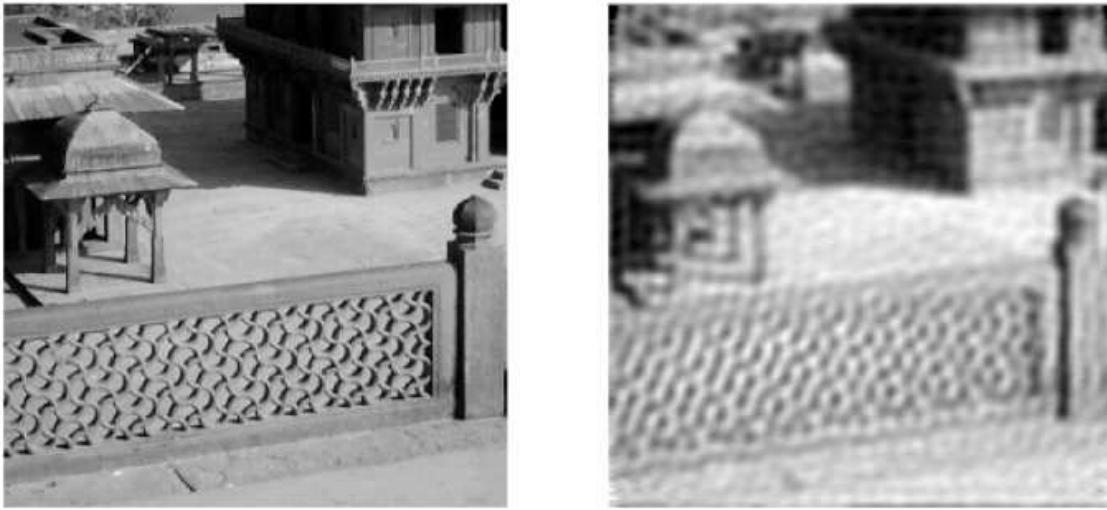


FIGURE 4.12 – Filtrage passe-bas d'une image : (Gauche) Image originale. (Droite) Image filtrée.

**Définition [Filtre idéal].** Un filtre est idéal si sa fonction de transfert est telle que

$$H(u, v) = 1 \text{ ou } 0 \quad \text{pour tout } (u, v) \quad (4.6)$$

Dans le cas d'un signal 1D, on distingue habituellement 3 catégories de filtres idéaux en fonction de leur gabarit le long de l'axe des fréquences. L'extension à des signaux 2D est simple si l'on utilise des filtres à symétrie circulaire.

### Filtre passe-bas idéal

Un filtre passe-bas idéal est caractérisé par le fait que seuls les coefficients  $H(u, v)$  à proximité de l'origine sont non nuls. Physiquement, un filtre passe-bas à pour effet d'atténuer les variations rapides d'intensité de l'image pouvant aller jusqu'à faire apparaître une certaine impression de flou dans l'image filtrée.

La figure 4.12 montre image de taille  $256 \times 256$  et l'image filtrée au moyen d'un filtre à symétrie circulaire de fréquence de coupure  $R_0 = 30$  pixels. Ce filtre a un effet moyenneur sur le niveau de luminance. En raison de la suppression des composantes à haute fréquence, les transitions se retrouvent adoucies.

Le filtre passe-bas idéal circulaire est défini par une fonction de transfert de la forme

$$H(u, v) = \begin{cases} 1 & \text{si } \sqrt{u^2 + v^2} \leq R_0 \\ 0 & \text{si } \sqrt{u^2 + v^2} > R_0 \end{cases} \quad (4.7)$$

Les composantes fréquentielles de l'image correspondant aux couples  $(u, v)$  situées à l'intérieur du disque de rayon  $R_0$ , dites basses fréquences, ne subissent aucune modification tandis que les autres composantes fréquentielles, dites hautes fréquences, sont complètement supprimées. La réponse impulsionale du filtre est obtenue par transformée de FOURIER inverse de la fonction de transfert :

$$h(x, y) = R_0 \frac{J_1(2\pi R_0 \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}} \quad (4.8)$$

### Filtre passe-haut idéal

Ce filtre supprime les composantes basse fréquence de l'image, tandis qu'il laisse intact les composantes haute fréquence. Visuellement, un filtre passe-haut a pour effet de supprimer la composante continue de l'image et de ne garder que les variations rapides d'intensité dans l'image filtrée. Le filtre passe-haut idéal est défini par la fonction de transfert suivante :

$$H(u, v) = \begin{cases} 1 & \text{si } \sqrt{u^2 + v^2} \geq R_0 \\ 0 & \text{si } \sqrt{u^2 + v^2} < R_0 \end{cases} \quad (4.9)$$

Les basses fréquences dont la fréquence radiale  $\sqrt{u^2 + v^2}$  est inférieure à  $R_0$  sont complètement rejetées tandis que les hautes fréquences restent inchangées.

### Filtre passe-bande idéal

Un filtre passe-bande idéal est un filtre qui supprime les basses et les hautes fréquences de l'image. Seule une plage de fréquences n'est pas modifiée. Le filtre passe-bande idéal est défini par la fonction de transfert suivante :

$$H(u, v) = \begin{cases} 1 & \text{si } R_0 \leq \sqrt{u^2 + v^2} \leq R_1 \\ 0 & \text{sinon} \end{cases} \quad (4.10)$$

Seules les composantes fréquentielles dont la fréquence radiale est comprise entre  $R_0$  et  $R_1$  sont conservées. Toutes les autres composantes fréquentielles sont rejetées.

Le figure 4.13 compare les effets d'un filtrage passe-bas, d'un filtrage passe-bande et d'un filtrage passe-haut. Pour la visualisation des spectres, nous avons adopté la convention de *vidéo inverse*; elle consiste simplement à inverser l'échelle de luminance.

### Forme de la fenêtre

Le problème majeur dans le choix d'un filtre est la forme de la fenêtre. Les trois filtres idéaux décrits ci-dessus ont tous une fenêtre circulaire. Ceci n'est pas une obligation. En effet, on pourrait choisir une fenêtre de forme rectangulaire afin de faciliter l'implémentation. Ou encore, on pourrait construire des filtres idéaux hybrides comme par exemple un filtre idéal qui serait passe-bas dans la direction  $x$  ( $u$ ) et passe-haut dans la direction  $y$  ( $v$ ). La forme de la fenêtre serait alors une bande verticale englobant l'axe  $v$ .

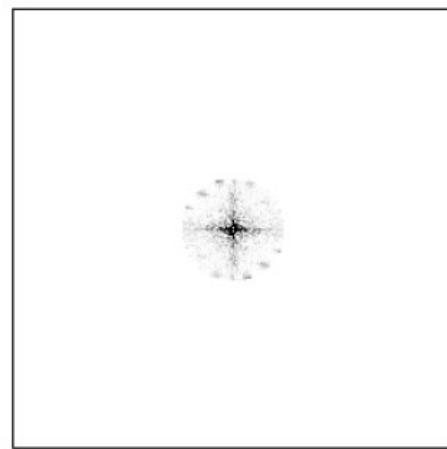
On remarquera que les filtres idéaux ont une fonction de transfert limitée spectralement mais une réponse impulsionnelle infinie. Cela signifie que, dans le plan spatial, une information locale va être fortement étalée. On parle bien de traitement global de l'image.

### Filtre non-idéal

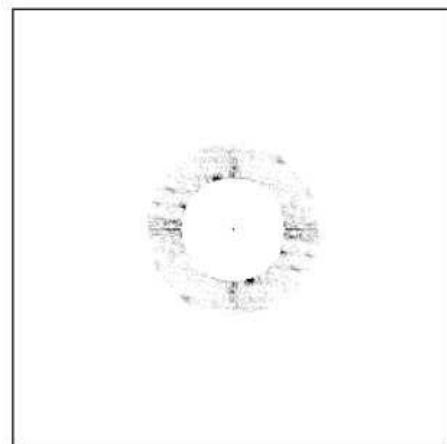
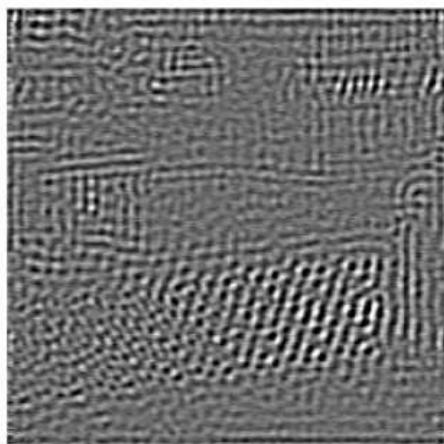
Un filtre non-idéal est un filtre dont la fonction de transfert

$$H(u, v) = ||H(u, v)|| e^{-j2\pi\theta(u, v)} \quad (4.11)$$

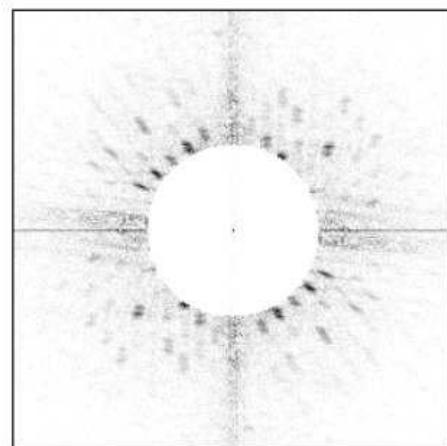
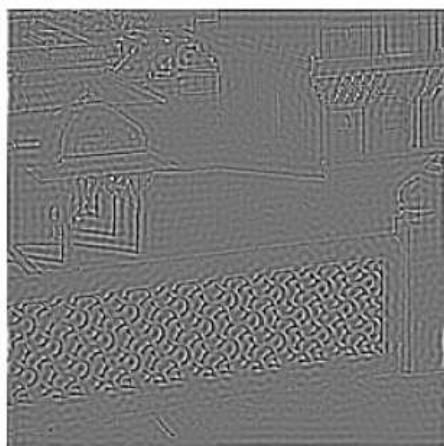
ne vaut pas simplement 1 ou 0 mais dont le module et la phase vérifient les propriétés de symétrie nécessaires pour que la réponse impulsionnelle  $h(x, y)$  soit réelle. Pour exemple, nous citons ci-après le filtre de BUTTERWORTH.



Application d'un filtre passe-bas circulaire de fréquence de coupure  $R_0 = 30$  pixels.



Application d'un filtre passe-bande circulaire conservant les fréquences  $[30, 50]$  pixels.



Application d'un passe-haut circulaire de fréquence de coupure  $R_0 = 50$  pixels.

FIGURE 4.13 – Images filtrées au moyen de filtres idéaux circulaires et spectres correspondants.

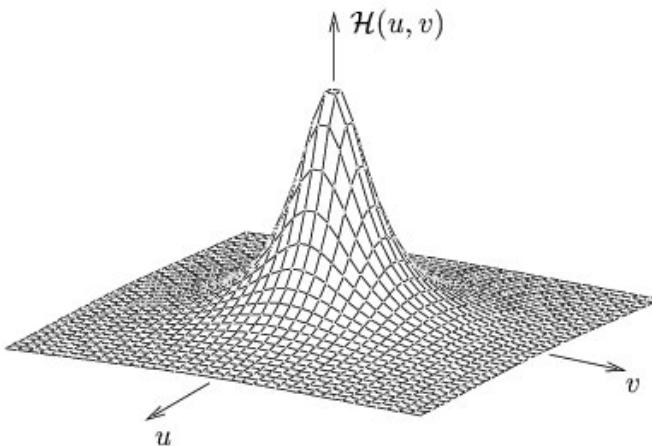


FIGURE 4.14 – Fonction de transfert du filtre passe-bas de BUTTERWORTH pour  $n = 1$ .

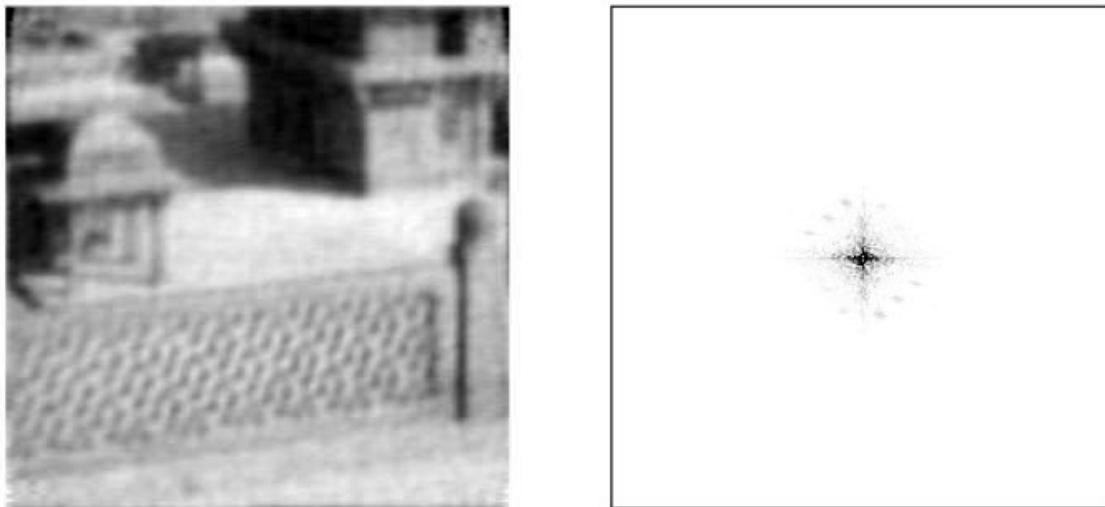


FIGURE 4.15 – Image filtrée par un filtre passe-bas de BUTTERWORTH d'ordre 1 ( $R_0 = 30$  pixels).

### Filtre passe-bas non-idéal

Un exemple de filtre passe-bas non-idéal est le filtre passe-bas de BUTTERWORTH d'ordre  $n$  défini par la fonction de transfert suivante

$$H(u, v) = \frac{1}{1 + \left(\frac{\sqrt{u^2+v^2}}{R_0}\right)^{2n}} \quad (4.12)$$

On remarque ici que toutes les composantes fréquentielles, hormis l'origine, subissent une atténuation d'autant plus grande que le couple  $(u, v)$  est éloigné de l'origine. Plus l'ordre  $n$  du filtre est élevé, plus l'atténuation des hautes fréquences est importante. La fonction de transfert du filtre passe-bas de BUTTERWORTH pour  $n = 1$  est représentée à la figure 4.14. La figure 4.15 montre une image filtrée par le filtre passe-bas de BUTTERWORTH.

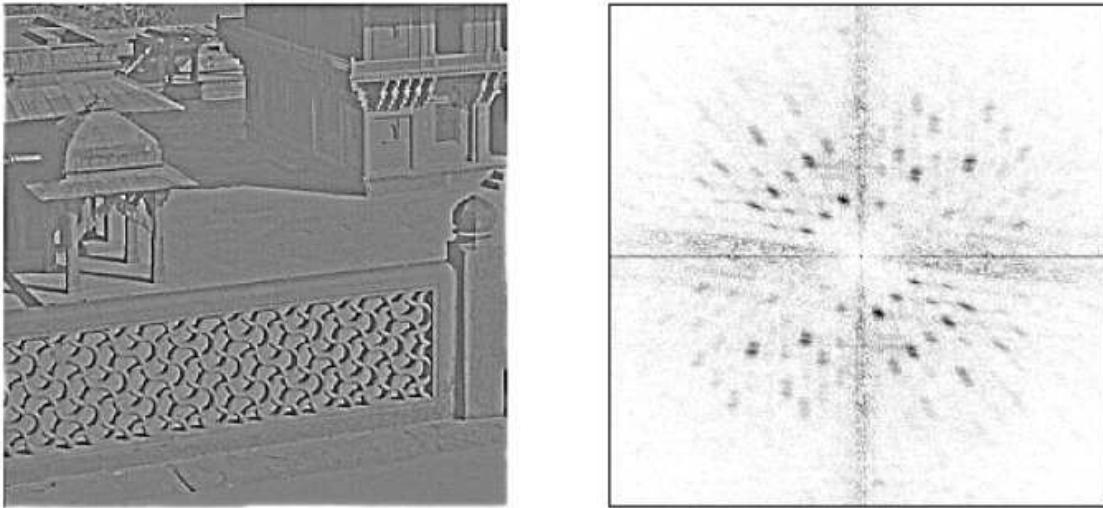


FIGURE 4.16 – Image filtrée par un filtre passe-haut de BUTTERWORTH d’ordre 1 ( $R_0 = 50$  pixels).

### Filtre passe-haut

Un exemple de filtre passe-haut non-idéal est celui de BUTTERWORTH d’ordre  $n$  dont la fonction de transfert est donnée par

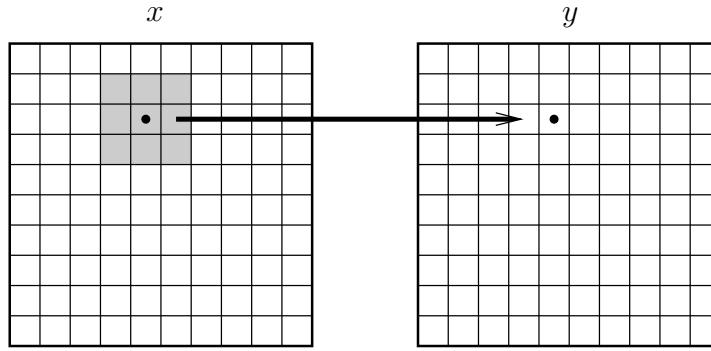
$$H(u, v) = \frac{1}{1 + \left(\frac{R_0}{\sqrt{u^2 + v^2}}\right)^{2n}} \quad (4.13)$$

Comme pour le cas du filtre passe-bas, toutes les fréquences sont atténuées et cela d’autant plus que  $\sqrt{u^2 + v^2}$  est petit par rapport à  $R_0$ . De plus,  $n$  fixe la pente de la transition entre basses et hautes fréquences. La figure 4.16 montre une image filtrée par le filtre passe-haut de BUTTERWORTH.

### 4.2.3 Traitement local : masques de convolution

Le problème d’un traitement global de l’image est qu’il n’est pas adapté aux usages industriels qui nécessitent des temps de traitement très courts. En effet, comme nous l’avons vu plus haut, chaque pixel de l’image filtrée dépend de tous les pixels de l’image originale. L’image filtrée s’obtient soit par une convolution assez lourde, vu le nombre important de coefficients, soit par transformée de FOURIER, multiplication par la fonction de transfert du filtre et transformée de FOURIER inverse. Dans les deux cas, les opérations à effectuer peuvent prendre un temps trop long.

Un filtre local est un filtre pour lequel la réponse impulsionnelle  $h(i, j)$  est nulle pour un très grand nombre de coefficients. Dès lors, la valeur d’un pixel de l’image filtrée ne dépend plus que d’un petit nombre (en général 9 ou 25) de pixels proches du pixel traité, ce qui rend le traitement local et nettement plus rapide. Dans ce contexte, la réponse impulsionnelle porte plutôt le nom de *masque de convolution* ou encore de *noyau du filtre*. Dans la plupart des cas, le masque de convolution a une forme carrée de taille  $3 \times 3$ ,  $5 \times 5$  pixels et peut se représenter

FIGURE 4.17 – Application d'un masque de convolution de taille  $3 \times 3$ .

par une matrice carrée de la forme

$$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \text{ ou } \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} \\ h_{21} & h_{22} & h_{23} & h_{24} & h_{25} \\ h_{31} & h_{32} & h_{33} & h_{34} & h_{35} \\ h_{41} & h_{42} & h_{43} & h_{44} & h_{45} \\ h_{51} & h_{52} & h_{53} & h_{54} & h_{55} \end{bmatrix} \quad (4.14)$$

L'origine  $h(0, 0)$  de la réponse impulsionnelle étant le coefficient  $h_{22}$  dans le cas d'un masque  $3 \times 3$  ou le coefficient  $h_{33}$  dans le cas d'un masque  $5 \times 5$ . On pourrait imaginer des masques de taille  $7 \times 7$ ,  $9 \times 9$  mais leur comportement serait alors de moins en moins local.

L'application d'un filtre local se réalise alors simplement par une sommation de produits, c'est-à-dire la sommation des pixels de l'image originale pondérés par les coefficients  $h_{11}, h_{12}, \dots$ . Dans le cas d'un masque  $3 \times 3$ , nous avons

$$\begin{aligned} y(m, n) = & h_{11} x(m-1, n-1) + h_{12} x(m, n-1) + h_{13} x(m+1, n-1) \\ & + h_{21} x(m-1, n) + h_{22} x(m, n) + h_{23} x(m+1, n) \\ & + h_{31} x(m-1, n+1) + h_{32} x(m, n+1) + h_{33} x(m+1, n+1) \end{aligned} \quad (4.15)$$

On remarque bien qu'un pixel  $y(m, n)$  de l'image filtrée  $y$  ne dépend que de, au plus, 9 pixels de l'image originale  $x$ . La figure 4.17 illustre la situation.

La formule (4.15) s'applique sans problème pour tous les pixels situés au sein de l'image, c'est-à-dire pour  $m = 1, \dots, M-2$  et  $n = 2, \dots, N-2$ . Mais qu'en est-il des bords de l'image, c'est-à-dire des pixels pour lesquels  $m = 0, m = M-1, n = 0$  et  $n = N-1$ ? En effet, pour ces pixels, la valeur de l'image filtrée dépend de pixels situés en-dehors de l'image originale... Plusieurs possibilités existent :

- considérer ces pixels "extérieurs" comme étant égaux à la valeur 0. Ce n'est sans doute pas la meilleure solution...
- considérer que ces pixels "extérieurs" prolongent l'image. On ajoute donc lors du calcul un bord, épais d'un pixel, à l'image originale avec des valeurs identiques aux pixels adjacents.
- "Miroiriser" l'image autour de ses bords.

Néanmoins, il n'y a pas de solution miracle...

La plupart du temps, on s'arrangera pour que la somme des coefficients du masque de convolution soit égale à 1 afin de ne pas modifier la dynamique de l'image. Nous citons ci-après quelques exemples de filtres locaux simples. D'autres seront étudiés ultérieurement pour l'extraction de caractéristiques de l'image.

### Filtre Moyenne

Un premier exemple de filtre local simple est le filtre moyenne dont le masque de convolution est donné par

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{ou} \quad \frac{1}{25} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Ce filtre

- remplace la valeur d'un pixel par la moyenne des valeurs de tous ses pixels voisins, y compris lui-même ;
- permet de lisser l'image (“smoothing”)
- réduit le bruit présent dans l'image
- “réduit” les détails de l'image
- rend floue l'image (“blur edge”)
- a typiquement le comportement d'un filtre passe-bas. Ceci pourrait être montré en calculant la transformée de FOURIER de la relation (4.15).

## 4.3 Traitements non-linéaires

Les traitements non-linéaires de signaux ne peuvent pas se mettre sous la forme d'un produit de convolution. Dès lors, la transformée de FOURIER ne sera ici d'aucune utilité. De plus, la plupart des méthodes non-linéaires sont locales et sont donc particulièrement appréciées pour l'usage industriel de part leur rapidité de traitement mais également par la qualité des résultats de traitement obtenus. En effet, nous verrons que de nombreux filtres non-linéaires donnent des résultats préférés à ceux fournis par les filtres linéaires “équivalents”.

Les méthodes non-linéaires que nous envisagerons ici sont basées sur la morphologie mathématique. Cette branche des mathématiques est fondamentalement géométrique. Elle consiste à comparer les objets d'une image à analyser à un autre objet de forme connue appelé *élément structurant*. Cette théorie repose sur des notions ensemblistes que nous rappellerons brièvement. Néanmoins, elle s'appliquera aussi bien à des images binaires qu'à des images en niveaux de gris.

### 4.3.1 Images binaires

#### 4.3.1.1 Rappels sur la théorie des ensembles

Les ensembles seront notés par des majuscules  $A, B, \dots$  et les éléments qu'ils contiennent par les minuscules  $a, b, \dots$  L'ensemble vide sera noté  $\emptyset$ .

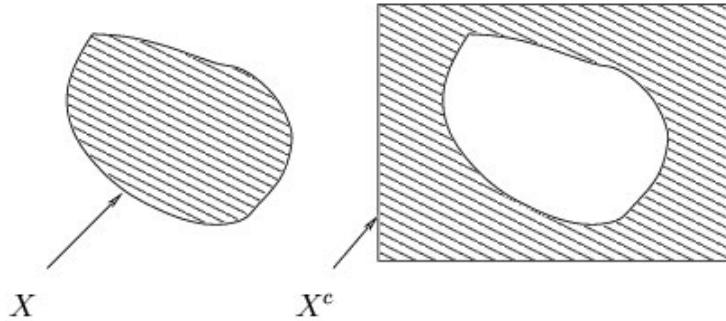


FIGURE 4.18 – Complémentaire d'un ensemble.

### Égalité d'ensembles

Deux ensembles sont égaux s'ils sont formés des mêmes éléments :

$$X = Y \Leftrightarrow (x \in X \Rightarrow x \in Y) \text{ et } (x \in Y \Rightarrow x \in X)$$

### Inclusion

$X$  est inclus dans  $Y$  si tous les éléments de  $X$  appartiennent à  $Y$  :

$$X \subseteq Y \Leftrightarrow (x \in X \Rightarrow x \in Y)$$

### Intersection

L'intersection de deux ensembles  $X$  et  $Y$  est l'ensemble des éléments qui appartiennent aux deux :

$$X \cap Y = \{x : x \in X \text{ et } x \in Y\}$$

### Union

L'union de deux ensembles est constituée des éléments appartenant à l'un ou à l'autre, c'est-à-dire

$$X \cup Y = \{x : x \in X \text{ ou } x \in Y\}$$

### Différence

Étant donnés  $X$  et  $Y$ , la différence de  $X$  par  $Y$ , notée  $X - Y$  ou  $X \setminus Y$  est l'ensemble des éléments de  $X$  qui n'appartiennent pas à  $Y$  :

$$X - Y = \{x : x \in X \text{ et } x \notin Y\}$$

### Complémentaire

Soit un ensemble  $X$  contenu dans un ensemble  $\varepsilon$  servant de référentiel (en bref l'encadrement de l'image). Le complémentaire de  $X$  dans  $\varepsilon$  est l'ensemble  $X^c$  fourni par

$$X^c = \{x : x \in \varepsilon \text{ et } x \notin X\}$$

La figure 4.18 illustre la notion de complémentaire.



FIGURE 4.19 – Translation d'un ensemble par un élément  $b$ .

### Translaté

Le translaté d'un ensemble  $X$  par un point (ou vecteur)  $b$  vaut

$$X_b = \{z : z = x + b, x \in X\}$$

Comme le montre la figure 4.19, la translation d'un ensemble consiste à déplacer l'ensemble dans le référentiel.

#### 4.3.1.2 Transformations morphologiques élémentaires

##### Érosion

Pour définir de manière intuitive l'opération d'érosion, situons-nous dans le plan de l'espace euclidien  $\mathbb{R}^2$  partiellement occupé par un ensemble  $X$ . Prenons un élément structurant  $B$  représentant une figure géométrique simple, par exemple un disque (ou un carré). Cet élément  $B_z$  est repéré par son centre et placé en  $z$  dans l'espace  $\mathbb{R}^2$ . Il est ensuite déplacé de telle sorte que son centre occupe successivement toutes les positions de l'espace. Pour chacune de ces positions, la question suivante est posée : l'ensemble  $B$  est-il entièrement inclus dans l'ensemble  $X$  ( $B_z \subseteq X$ ) pour cette position de  $z$ ? L'ensemble des  $z$  fournissant une réponse positive forme un nouvel ensemble appelé érosion de  $X$  par  $b$  et noté  $X \ominus B$ . Il vaut

$$X \ominus B = \{z : B_z \subseteq X\}$$

On peut montrer que la définition suivante, de forme plus algébrique, fournit le même résultat :

$$X \ominus B = \bigcap_{b \in B} X_{-b}$$

Ce qui signifie que la transformation par érosion d'un ensemble  $X$  par  $B$  s'obtient en translatant  $X$  par l'opposé de chaque élément de  $B$  et en ne conservant que les points appartenant à toutes les translations de  $X$ . La figure 4.20 illustre cette formulation algébrique. Avec cette interprétation, il est clair que si la taille de  $B$  dépasse celle de  $X$ , l'ensemble  $X$  érodé par  $B$  est vide.

L'opération d'érosion est illustrée à la figure 4.21. Les effets habituels de l'érosion sont :

- la séparation des objets à l'endroit des étranglements,
- le rétrécissement des objets de taille supérieure à  $B$ , et
- la disparition des petits objets (dont la taille est inférieure à  $B$ ).

##### Dilatation

La dilatation se définit de manière analogue à l'érosion. En prenant le même élément structurant  $B$ , pour chaque position  $z$  du centre de  $B$  la question est : l'ensemble  $B_z$  touche-t-il  $X$

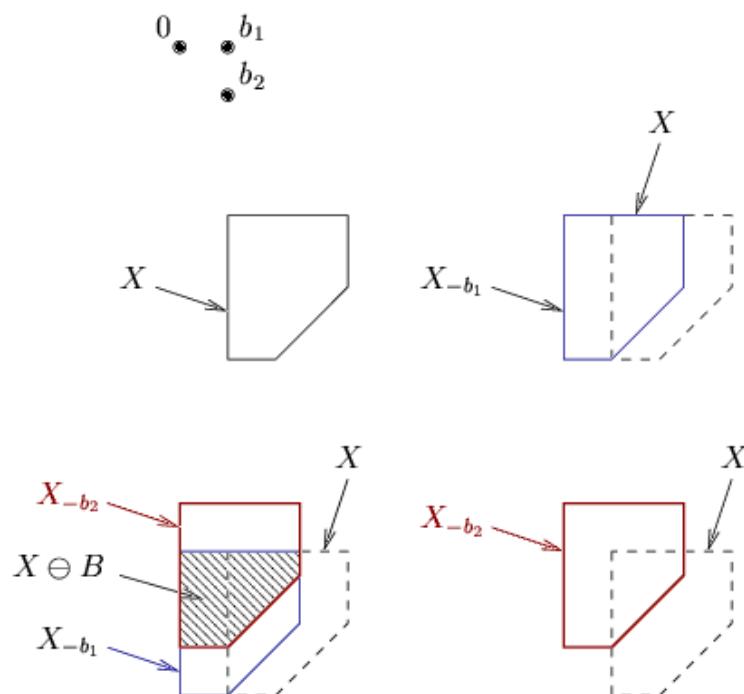
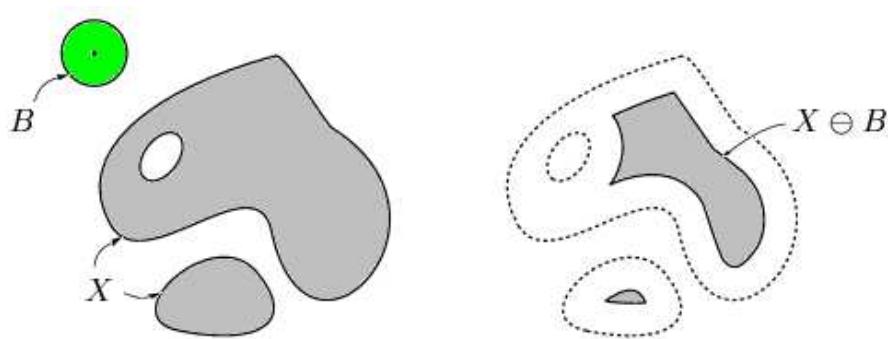


FIGURE 4.20 – Interprétation algébrique de l'érosion.

FIGURE 4.21 – Érosion de  $X$  par un disque  $B$ . L'origine de l'élément structurant est représentée par un point noir.

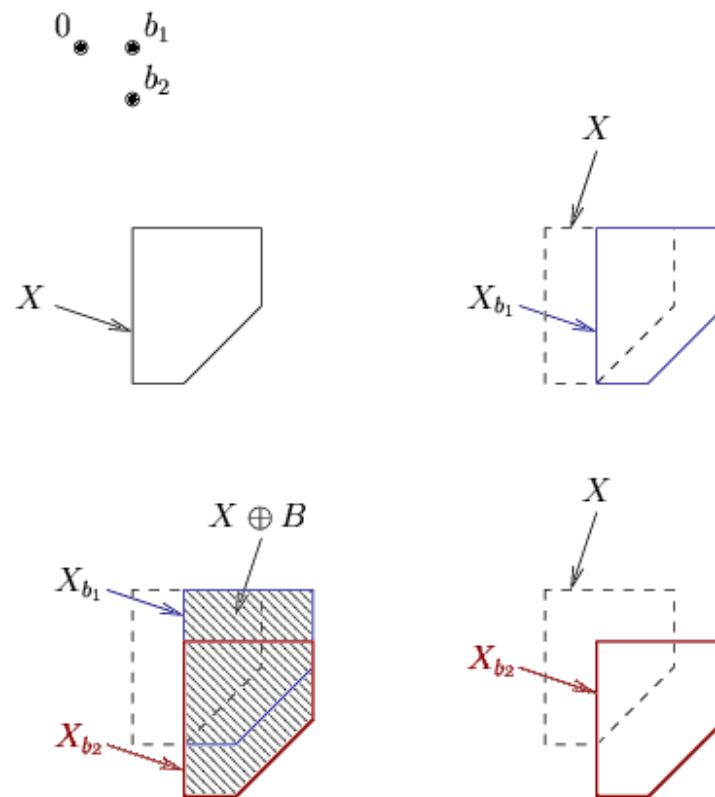
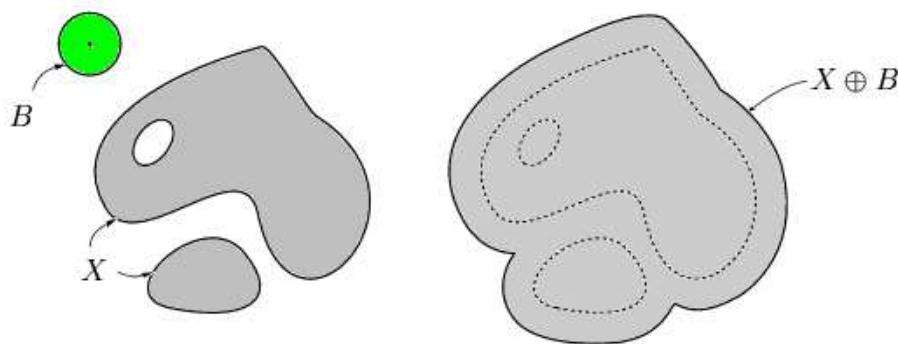


FIGURE 4.22 – Illustration de la définition algébrique de la dilatation.

FIGURE 4.23 – Dilatation de  $X$  par un disque  $B$ .

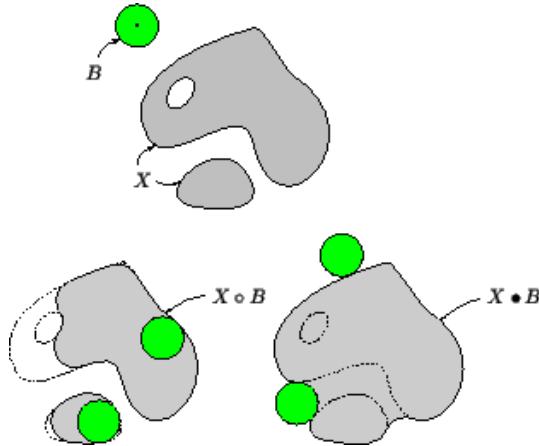
$(X \cap B_z \neq \emptyset)$ ? L'ensemble des points  $z$  correspondant à une réponse positive forme le nouvel ensemble  $X \oplus B$ :

$$X \oplus B = \bigcup_{x \in X} B_x = \bigcup_{b \in B} X_b = \{x + b : x \in X, b \in B\}$$

Le principe de la dilatation, à savoir l'union des translatés de  $X$  par les éléments de  $B$ , est illustré à la figure 4.22.

La figure 4.23 montre la dilatation par un disque du même ensemble  $X$  que celui traité lors de l'érosion. Dans cet exemple, les deux composantes connexes<sup>2</sup> sont réunies dans le dilaté.

2. Un ensemble  $X$  est connexe si  $\forall x_1, x_2 \in X$ , il existe un chemin reliant  $x_1$  à  $x_2$  totalement inclus dans  $X$ .

FIGURE 4.24 – Ouverture et fermeture de  $X$  par un disque  $B$ .

### Ouverture

L'opération obtenue par la succession d'une érosion et d'une dilatation est l'ouverture morphologique. Elle se note

$$X \circ B = (X \ominus B) \oplus B$$

En général, l'ensemble traité diffère de l'ensemble de départ : l'ensemble ouvert est plus régulier et moins riche en détails que l'ensemble initial. La transformation par ouverture adoucit les contours. L'ouverture peut donc jouer le rôle d'un filtre.

### Fermeture

En inversant l'ordre des opérations utilisées pour définir l'ouverture, nous obtenons une nouvelle opération appelée fermeture, c'est-à-dire la transformation

$$X \bullet B = (X \oplus B) \ominus B$$

L'ouverture et la fermeture sont illustrées par la figure 4.24.

### Interprétation de l'ouverture et de la fermeture

L'interprétation la plus commode de l'opération d'ouverture est illustrée par la propriété que voici :

$$X \circ B = \bigcup \{B_z : B_z \subseteq X\}$$

Ainsi, l'ouverture d'une figure par un élément structurant  $B$  est l'ensemble des points recouverts lors du déplacement de  $B$  à l'intérieur de la figure. Une propriété similaire vaut pour la fermeture ; mais cette fois, l'élément structurant parcourt le complémentaire de la figure.

### Notes sur l'implémentation de la dilatation et de l'érosion

Considérons un élément structurant  $B$  carré de taille  $N \times N$ . L'opération de dilatation d'un ensemble  $X$  de pixels est une opération aisée dans une architecture actuelle. Le nombre d'opérations nécessaires au calcul “brutal” de la dilatation  $X \oplus B$  est proportionnel à  $N^2$ , le facteur de proportionnalité dépendant du nombre de pixels de l'ensemble  $X$ . Il est possible de diminuer le nombre d'opérations en se basant sur la propriété d'associativité de la dilatation :

$$X \oplus B = X \oplus (B_1 \oplus B_2) = (X \oplus B_1) \oplus B_2$$

Dès lors, si l'élément structurant  $B$  peut se décomposer en la dilatation de deux éléments structurants plus simple  $B_1$  et  $B_2$ , la dilatation de  $X$  par  $B$  peut être réalisée par la dilatation successive de  $X$  par  $B_1$ , suivie de la dilatation de l'ensemble obtenu par  $B_2$ . Dans le cas de notre élément structurant carré de taille  $N \times N$ , on pourrait le décomposer en deux éléments structurants de taille  $1 \times N$  (ligne horizontale de longueur  $N$ ) et de taille  $N \times 1$  (ligne verticale de longueur  $N$ ). Ceci conduirait alors à un nombre d'opérations proportionnel à  $2N$  au lieu de  $N^2$ , d'où l'intérêt.

Une propriété similaire peut être utilisée pour réduire le nombre d'opérations nécessaires à une érosion par un élément structurant  $B = B_1 \oplus B_2$ . Il s'agit de

$$X \ominus B = X \ominus (B_1 \oplus B_2) = (X \ominus B_1) \ominus B_2$$

L'érosion de  $X$  par  $B$  peut alors être réalisée par l'érosion successive de  $X$  par  $B_1$ , suivie de l'érosion de l'ensemble obtenu par  $B_2$ .

#### 4.3.1.3 Transformations morphologiques complexes

À partir des opérations élémentaires d'érosion, de dilatation, d'ouverture et de fermeture, il est possible de construire de nouveaux opérateurs morphologiques plus évolués et dont l'utilité sera précisée plus loin dans cet exposé.

#### Dilatation géodésique

Une dilatation géodésique fait toujours intervenir deux images. La première image est le dilaté par un élément structurant élémentaire adapté à la trame ; il s'agit d'un carré de taille  $3 \times 3$  pour une trame carrée. Quant à la seconde image, elle limite l'extension de la dilatation de la première.

La dilatation géodésique de taille 1 de l'ensemble  $X$  conditionnellement à  $Y$ , notée  $D_Y^{(1)}(X)$ , est définie comme l'intersection du dilaté de taille 1 et de  $Y$  :

$$\boxed{\forall X \subseteq Y : D_Y^{(1)}(X) = (X \oplus B) \cap Y} \quad (4.16)$$

où  $B$  est l'élément le plus simple adapté à la trame. Suite à l'apparition de l'intersection dans la définition,  $D_Y^{(1)}(X)$  est toujours inclus ou égal à  $Y$ . On dit aussi que  $Y$  sert de *masque géodésique*. La figure 4.25 illustre le principe de la dilatation géodésique de taille 1 sur un ensemble dans une trame carrée.

La dilatation géodésique de taille  $n$  de l'ensemble  $X$  conditionnellement à  $Y$ , notée  $D_Y^{(n)}(X)$ , est définie comme une succession du dilaté géodésique de taille 1 :

$$\boxed{\forall X \subseteq Y : D_Y^{(n)}(X) = \underbrace{D_Y^{(1)}(D_Y^{(1)}(\dots D_Y^{(1)}(X)))}_{n \text{ fois}}}$$

où  $B$  est l'élément le plus simple adapté à la trame.

#### Érosion géodésique

L'érosion géodésique de taille  $n$  de l'ensemble  $X$  conditionnellement à  $Y$ , notée  $(X \ominus B)^{(n)}$ , est définie comme une succession de l'érodé géodésique de taille 1 et de  $Y$  :

$$(X \ominus B)^{(n)} = \underbrace{((X \ominus B)^{(1)} \ominus B)^{(1)} \dots}_{n \text{ fois}}$$

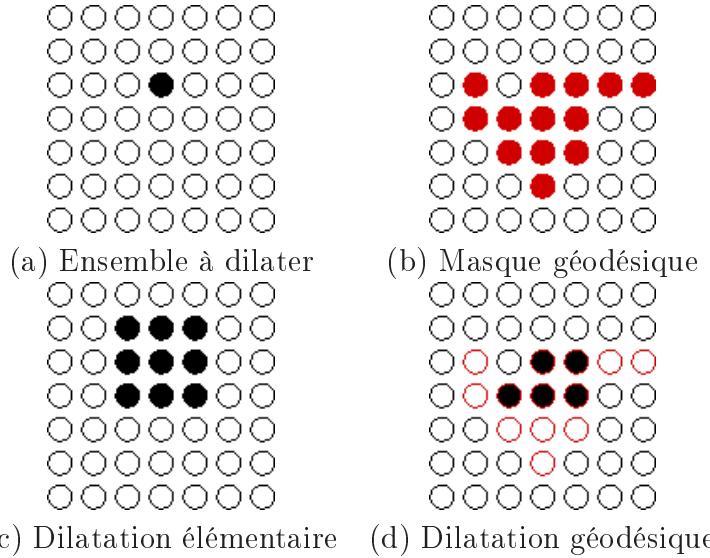


FIGURE 4.25 – Dilatation géodésique de taille 1 d'un ensemble.

où l'érodé géodésique de taille 1 est égal à

$$\forall X \supseteq Y : (X \ominus B)^{(1)} = (X \ominus B) \cup Y \quad (4.17)$$

### Reconstruction géodésique

Les opérations géodésiques sont rarement utilisées en tant que telles. C'est plutôt l'ensemble vers lequel convergent ces opérations lorsque  $n$  croît qui permet de résoudre certains problèmes pratiques, comme par exemple l'extraction de particules pré-sélectionnées dans une image. En fait, l'érosion et la dilatation géodésique convergent après un certain nombre d'itérations. Le procédé de reconstruction géodésique tire profit de ce principe.

La reconstruction de  $X$  conditionnellement à  $Y$  est la dilatation géodésique de  $X$  jusqu'à convergence. Soit  $i$  la valeur à partir de laquelle la limite de convergence est atteinte, la reconstruction de  $X$  est définie par

$$R_Y(X) = D_Y^{(i)}(X) \text{ avec } D_Y^{(i+1)}(X) = D_Y^{(i)}(X) \quad (4.18)$$

Comme l'indique son nom, cette opération permet de reconstruire des détails perdus éventuellement lors d'une autre opération. La figure 4.26 représente les 3 étapes mises en oeuvre pour extraire des particules choisies dans une image. L'image (a) est le signal de départ. Dans l'image (b), on sélectionne certaines particules en insérant quelques points dans celles-ci. La reconstruction conduit alors à l'image (c) où ne sont reconstituées que les particules sélectionnées ; les autres ont été gommées.

### 4.3.2 Images en niveaux de gris

Il est possible de généraliser les notions d'érosion et de dilatation que nous avons vues plus haut aux images en niveaux de gris. Pour cela, nous devons tout d'abord généraliser les notions ensemblistes aux fonctions. Nous verrons donc une image comme une fonction qui, à tout point du plan, fait correspondre une valeur.

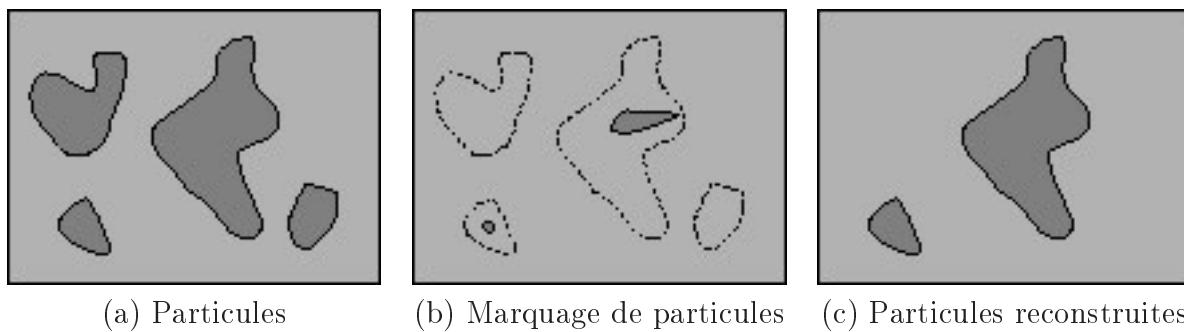


FIGURE 4.26 – Extraction de particules pré-sélectionnées dans une image.

#### 4.3.2.1 De la notion d'ensemble à celle de fonction

Pour la morphologie mathématique en niveaux de gris, les opérations de base ne sont plus l'union et l'intersection mais le *supremum*  $\vee$  et l'*infimum*  $\wedge$ . Pour aborder les autres notions, il faut d'abord définir la notion d'ordre entre fonctions, tout comme nous l'avions fait pour les ensembles en utilisant la notion d'inclusion.

#### Ordre entre fonctions (inclusion)

Soient deux fonctions  $f$  et  $g$ . La fonction  $f$  est inférieure à  $g$  se note :

$$f \leq g \text{ si } f(x) \leq g(x) \text{ pour tout } x \quad (4.19)$$

La relation “est plus grand”, notée  $\geq$ , est définie de la même manière.

#### Infimum et Supremum (intersection et union)

Soient deux fonctions  $f$  et  $g$ . L'infimum et le supremum se ramènent en fait à la notion de minimum et de maximum :

$$(f \vee g)(x) = \max(f(x), g(x)) \quad (4.20)$$

$$(f \wedge g)(x) = \min(f(x), g(x)) \quad (4.21)$$

#### Translaté

La translation de la fonction  $f$  par  $b$  est la fonction  $f_b$  définie par

$$f_b(x) = f(x - b) \quad (4.22)$$

#### 4.3.2.2 Transformations morphologiques élémentaires

Il est possible de remplacer la notion d'élément structurant par celle de fonction quelconque dite structurante. Néanmoins, cela ne conduit guère à des implémentations satisfaisantes. Nous nous limiterons donc au cas d'éléments structurant représentés par une fonction  $B$  caractérisée par un support fini que nous noterons  $D$ . Ce support peut être dans les cas les plus simples un carré ou un disque. Ce support est à mettre en rapport direct avec les masques de convolution que nous avons étudiés plus haut.

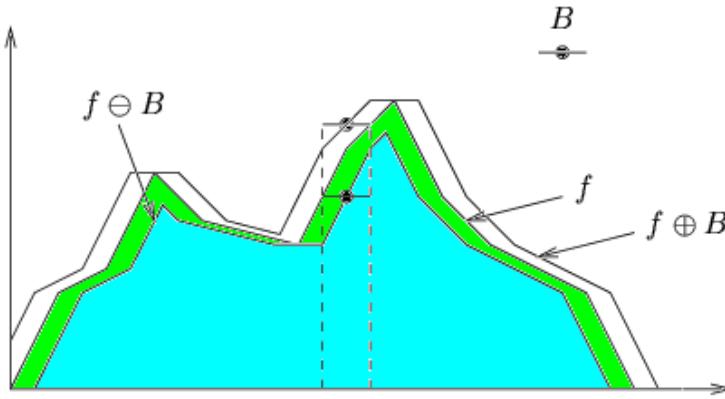


FIGURE 4.27 – Érosion et dilatation d'une fonction.

### Érosion et dilatation

Avec les notions introduites ci-dessus, il est maintenant possible de donner les définitions des opérateurs d'érosion et de dilatation d'une fonction  $f$  par un élément structurant  $B$  :

$$f \ominus B = \bigwedge_{h \in D} f_{-h}(x) \quad (4.23)$$

$$f \oplus B = \bigvee_{h \in D} f_h(x) \quad (4.24)$$

Ces opérations sont illustrées à la figure 4.27. Avec des éléments structurants de ce type, une dilatation se calcule comme l'enveloppe supérieure de tous les translatés de la fonction par les éléments  $h$  de l'élément structurant  $B$ . Quant à l'érosion, elle devient équivalente à la recherche du minimum sur le support de  $B$ .

### Ouverture et fermeture

Comme dans le cas des images binaires, l'ouverture  $f \circ B$  et la fermeture  $f \bullet B$  résultent de la mise en cascade de l'érosion et de la dilatation :

$$f \circ B = (f \ominus B) \oplus B \quad (4.25)$$

$$f \bullet B = (f \oplus B) \ominus B \quad (4.26)$$

Les figures 4.28 et 4.29 montrent le résultat d'une ouverture et d'une fermeture sur une fonction unidimensionnelle.

Par analogie avec le traitement d'ensembles, l'ouverture sur des images produit un effet de filtrage comme le montre la figure 4.30 où sont rassemblées plusieurs images dont l'image érodée, l'image dilatée et l'image ouverte, obtenues toutes avec un élément structurant carré. Puisqu'elle recherche un minimum, l'érosion assombrit l'image alors qu'à l'inverse, la dilatation éclaircit l'image. L'ouverture produit une image qui ne conserve que les parties claires ayant la taille et la forme de l'élément structurant ; les autres sont proprement gommées, c'est-à-dire noircies.

#### 4.3.2.3 Filtrage non-linéaire

Dans cette introduction au traitement d'images, nous nous limiterons au filtre de rang et au filtre médian qui est un cas particulier de ce dernier et qui est couramment utilisé pour la suppression du bruit dans les images.

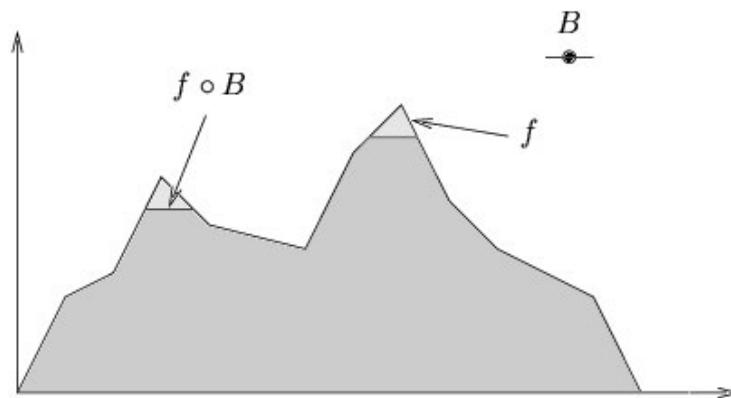


FIGURE 4.28 – Ouverture d'une fonction.

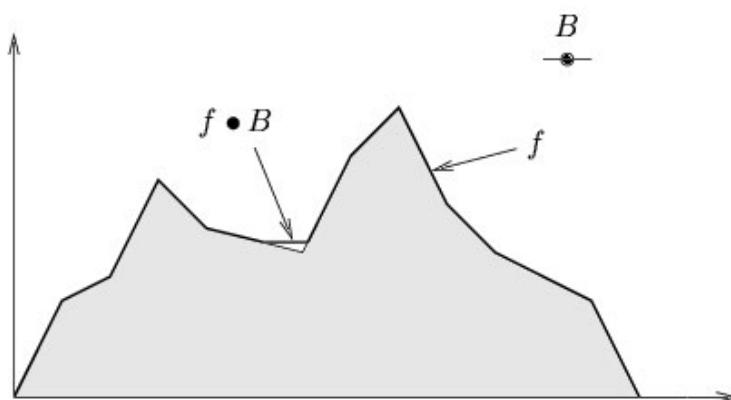


FIGURE 4.29 – Fermeture d'une fonction.



(a) Image originale



(b) Érosion par un carré



(c) Dilatation par un carré



(d) Ouverture par un carré

FIGURE 4.30 – Illustration des opérations élémentaires sur une image en niveaux de gris.

### Filtre de rang

L'érosion ou la dilatation correspondent à la sélection des valeurs extrêmes puisqu'il s'agit de déterminer l'infimum ou le supremum. Ce type d'opérations est donc très sensible au bruit, particulièrement à un bruit impulsif. On songe donc tout naturellement à considérer d'autres valeurs que ces valeurs extrêmes dans l'espoir de diminuer la sensibilité au bruit ; c'est la notion de *filtre de rang* qui s'en dégage.

Considérons un élément structurant carré  $B$  de taille  $n \times n$ . Une fois positionné sur un pixel de  $(x, y)$  de l'image, l'élément structurant englobe  $\sharp(B) = n^2$  pixels que l'on peut classer par ordre décroissant :

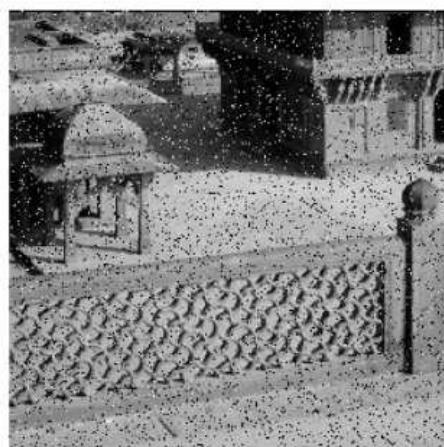
$$f_1 \geq f_2 \geq f_3 \geq \dots \geq f_{\sharp(B)-1} \geq f_{\sharp(B)}$$

Le filtre de rang d'ordre  $k$  remplace la valeur du pixel  $(x, y)$  par la  $k$ -ième valeur de la série ainsi obtenue. L'opération d'érosion correspond alors au filtre de rang d'ordre  $k = \sharp(B)$  tandis que la dilatation correspond au filtre de rang d'ordre  $k = 1$ .

### Filtre médian

Si  $\sharp(B)$  est impair, le choix  $k = \frac{1}{2}(\sharp(B) + 1)$  conduit à la définition d'un cas particulier de filtre de rang. Il s'agit du filtre médian car il sélectionne la médiane des valeurs de la série construite plus haut. Le filtrage médian est une technique de filtrage non-linéaire couramment utilisée en pratique. Il s'avère particulièrement efficace pour juguler les effets d'un bruit impulsif. Sa caractéristique essentielle est sa capacité à conserver des transitions fortes tout en supprimant une partie importante du bruit.

La figure 4.31 compare quelques opérations de filtrage. De même, la figure 4.32 montre l'effet d'un changement de la taille de la fenêtre sur la qualité de l'image filtrée.



(a) Image originale bruitée

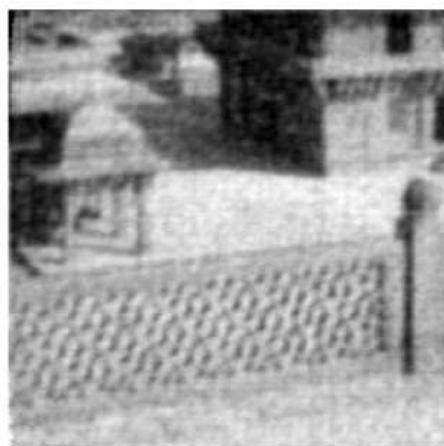
(b) Ouverture par un carré  $5 \times 5$ (c) Passe-bas BUTTERWORTH ( $R_0 = 50$ )(d) Médian par un carré  $5 \times 5$ 

FIGURE 4.31 – Illustration des opérations élémentaires sur une image en niveaux de gris.



(a) Image originale

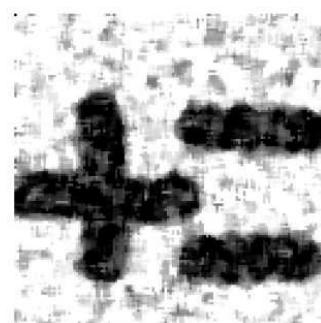
(b) Médian  $3 \times 3$ (c) Médian  $5 \times 5$ 

FIGURE 4.32 – Effet de la taille de l'élément structurant sur le filtrage.

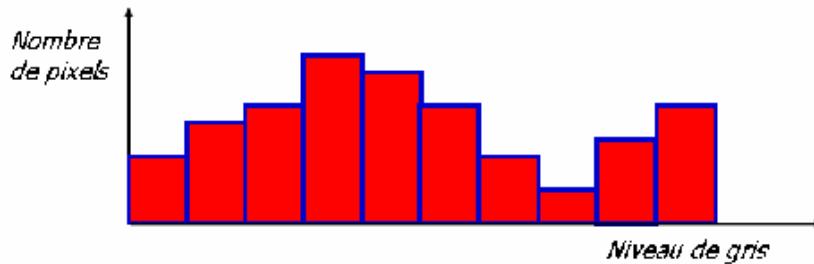


FIGURE 4.33 – Définition de l'histogramme d'une image.

## 4.4 Traitement spécifique : Rehaussement et restauration

L'acquisition d'une image s'accompagne souvent d'une distorsion ou d'une certaine dégradation. On peut songer à un éclairage trop faible ou non-uniforme, à des poussières sur une lentille, etc... Il n'y a dès lors pas d'autres solutions que de chercher à corriger les défauts par un procédé algorithmique. Dans ce contexte, on distingue deux familles de procédés de correction de défauts :

- le rehaussement, qui consiste à donner à l'image un aspect visuellement correct, et
- la restauration qui vise à rétablir la valeur exacte des pixels, c'est-à-dire des pixels de l'image qui aurait été obtenue en l'absence des conditions perturbatrices.

Dans cette introduction, nous parlerons essentiellement du rehaussement car la restauration nécessite de connaître les sources précises de la dégradation. Néanmoins, avant d'aborder différentes techniques, il nous faut introduire quelques définitions.

### 4.4.1 Définitions

Plusieurs termes apparaissent fréquemment lorsque l'on s'intéresse à la qualité d'une image : la *luminance* (ou luminosité, ou encore la brillance), le *contraste* mais aussi la *dynamique* d'une image. Un outil intéressant à l'analyse d'une image est l'*histogramme* des niveaux de gris.

#### Histogramme

L'histogramme d'une image représente la distribution des niveaux de gris (ou de couleurs). Il s'agit simplement d'une fonction que nous noterons *hist* dont la valeur *hist(k)* représente le nombre de pixels de l'image ayant la valeur *k*. Cette fonction est généralement représentée sous la forme d'un diagramme en bâtonnets, comme illustré à la figure 4.33.

#### Dynamique

Comme dans le cas des signaux 1D, la dynamique d'une image *f* est définie comme l'intervalle des valeurs comprises entre le minimum et le maximum de l'image :

$$[\min(f), \max(f)] \quad (4.27)$$

#### Luminance (ou brillance)

La luminance d'une image *f(m, n)* est définie comme la moyenne de tous les pixels de l'image

$$L = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \quad (4.28)$$

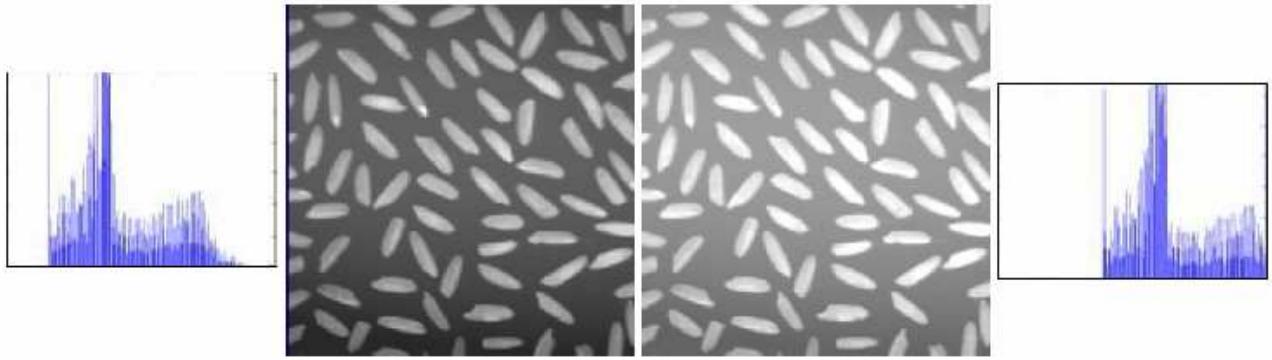


FIGURE 4.34 – Deux images de luminosité différente.

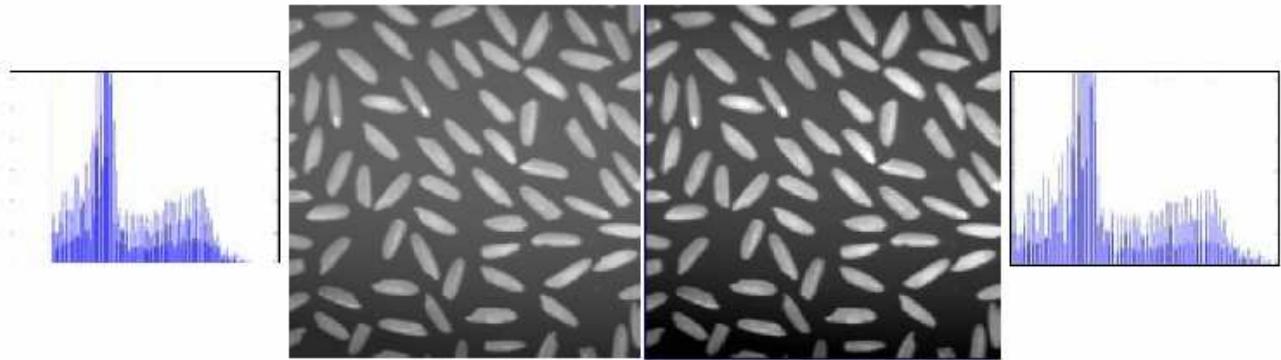


FIGURE 4.35 – Deux images de contraste différent.

La figure 4.34 illustre deux images de luminosité différente, ainsi que leur histogrammes respectifs.

### Contraste

Le contraste d'une image peut être défini de plusieurs façons :

- l'écart-type des variations des niveaux de gris :

$$C_1 = \sqrt{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (f(m, n) - L)^2} \quad (4.29)$$

- la variation entre niveaux de gris minimum et maximum :

$$C_2 = \frac{\max(f) - \min(f)}{\max(f) + \min(f)} \quad (4.30)$$

La figure 4.35 illustre deux images de contraste différent ainsi que leur histogrammes respectifs.

Le rehaussement d'une image (que nous verrons ici comme une amélioration du contraste) consiste à appliquer une fonction particulière à toutes les valeurs d'intensité  $I$  de l'image. Le type de rehaussement dépend essentiellement de la forme de la fonction à appliquée. On impose généralement que cette fonction soit croissante, de sorte que toute relation d'ordre entre pixels traduite par leur valeur soit maintenue, mais elle peut être de forme quelconque. Nous en donnons quelques exemples ci-dessous. Ces différentes techniques sont connues sous l'appellation de “manipulation de l'histogramme”.

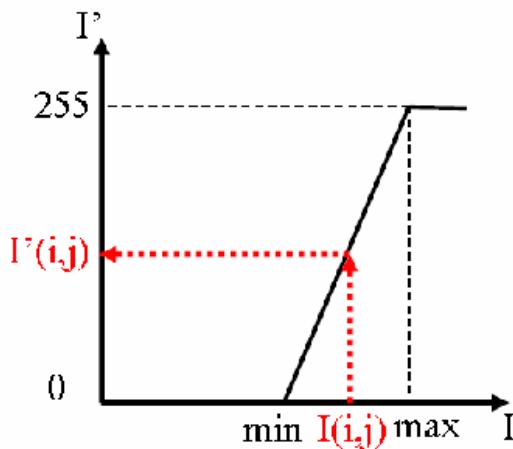


FIGURE 4.36 – Transformation linéaire de l'histogramme.

## 4.4.2 Amélioration du contraste par transformation de l'histogramme

### 4.4.2.1 Transformation linéaire

Cette transformation consiste à appliquer une fonction *rampe* aux valeurs d'intensité  $I$  de l'image. Chaque valeur  $I$  est ainsi transformée en une valeur  $I'$  selon une loi linéaire. Dans ce qui suit, nous noterons  $\min = \min(f)$ ,  $\max = \max(f)$ ,  $I(i, j)$  l'intensité du pixel  $f(i, j)$  avant transformation et  $I'(i, j)$  l'intensité du même pixel après transformation. La figure 4.36 illustre la loi linéaire considérée ici. La courbe ainsi représentée est appelée *courbe tonale*.

L'effet de cette transformation est de dilater au maximum la dynamique de l'image originale, ce qui s'exprime par

$$\frac{\max - \min}{I(i, j) - \min} = \frac{255 - 0}{I'(i, j) - 0}$$

pour une image ayant 256 niveaux de gris compris entre 0 et 255, et enfin conduit à la loi de transformation suivante :

$$I'(i, j) = 255 \frac{I(i, j) - \min}{\max - \min} \quad (4.31)$$

où

$$\frac{I(i, j) - \min}{\max - \min} \in [0, 1]$$

La figure 4.37 illustre l'effet de cette transformation sur une image. Au départ de cette transformation, il est possible d'imaginer d'autres transformations du même type.

### 4.4.2.2 Transformation linéaire avec saturation

Cette transformation consiste à choisir deux niveaux de saturation  $S_{\min}$  et  $S_{\max}$  en-dessous et au-dessus desquels les niveaux d'intensité  $I$  de l'image sont tous respectivement mis à 0 ou à 255. Ceci a pour effet d'accentuer une zone particulière de la dynamique de l'image :  $[S_{\min}, S_{\max}]$ . Cette transformation est caractérisée par la loi suivante :

$$I'(i, j) = \begin{cases} 255 \frac{I(i, j) - S_{\min}}{S_{\max} - S_{\min}} & \text{si } S_{\min} \leq I(i, j) \leq S_{\max} \\ 0 & \text{si } I(i, j) < S_{\min} \\ 255 & \text{si } S_{\max} < I(i, j) \end{cases} \quad (4.32)$$

La figure 4.38 illustre l'effet de cette transformation sur une image. De nouveau, la dynamique de l'image sortie est dilatée au maximum.

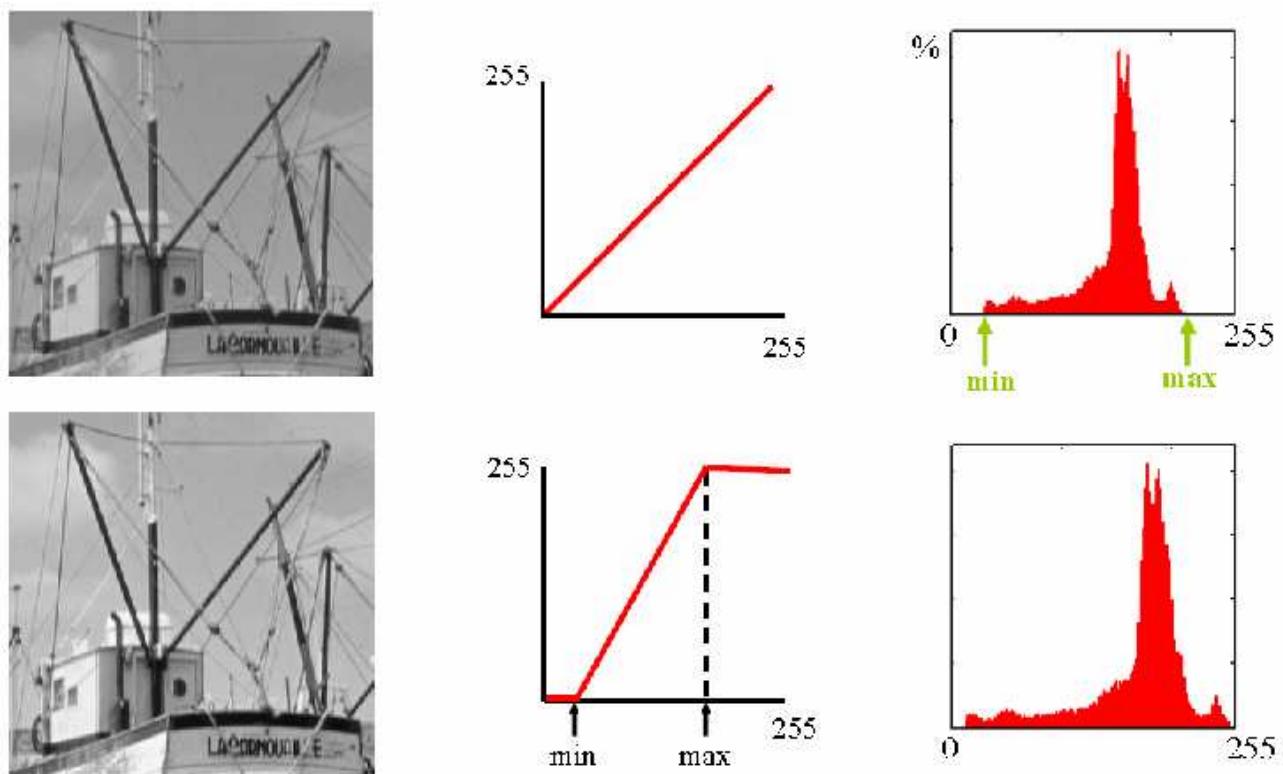


FIGURE 4.37 – Effet d'une transformation linéaire de l'histogramme.

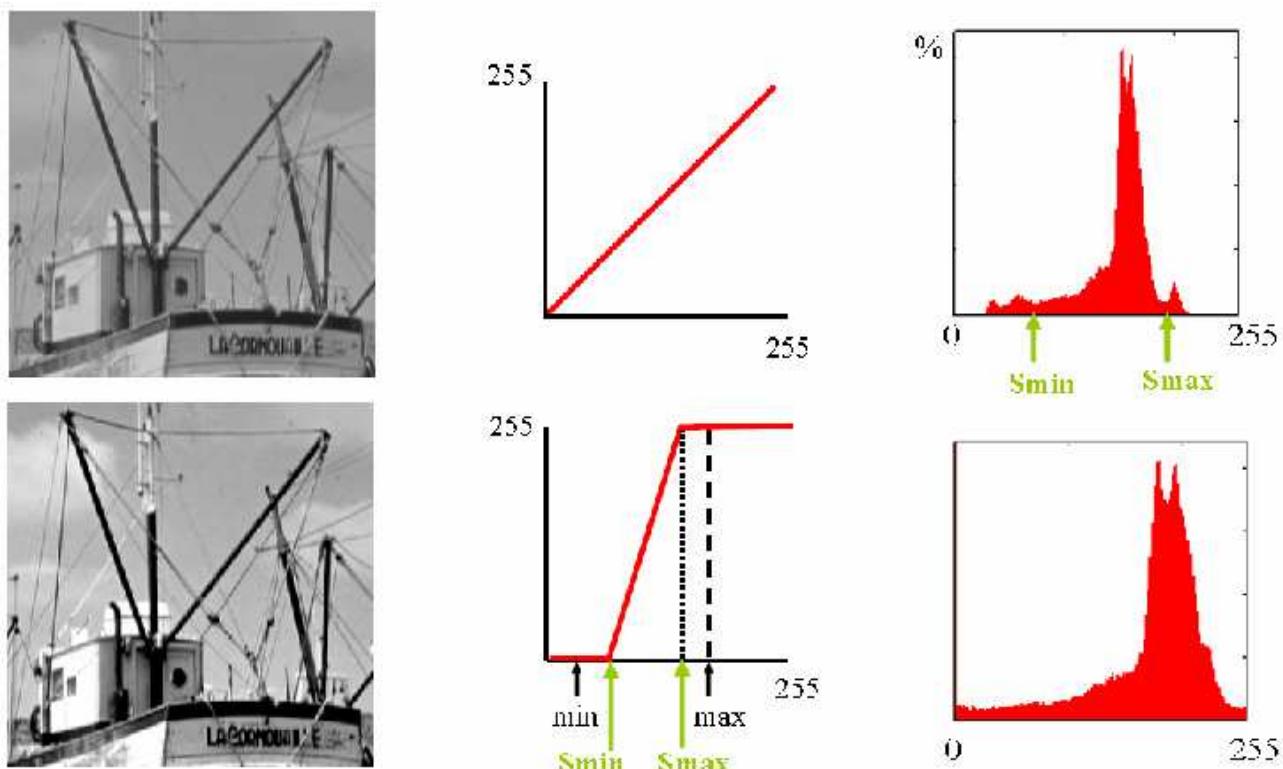


FIGURE 4.38 – Effet d'une transformation linéaire avec saturation.

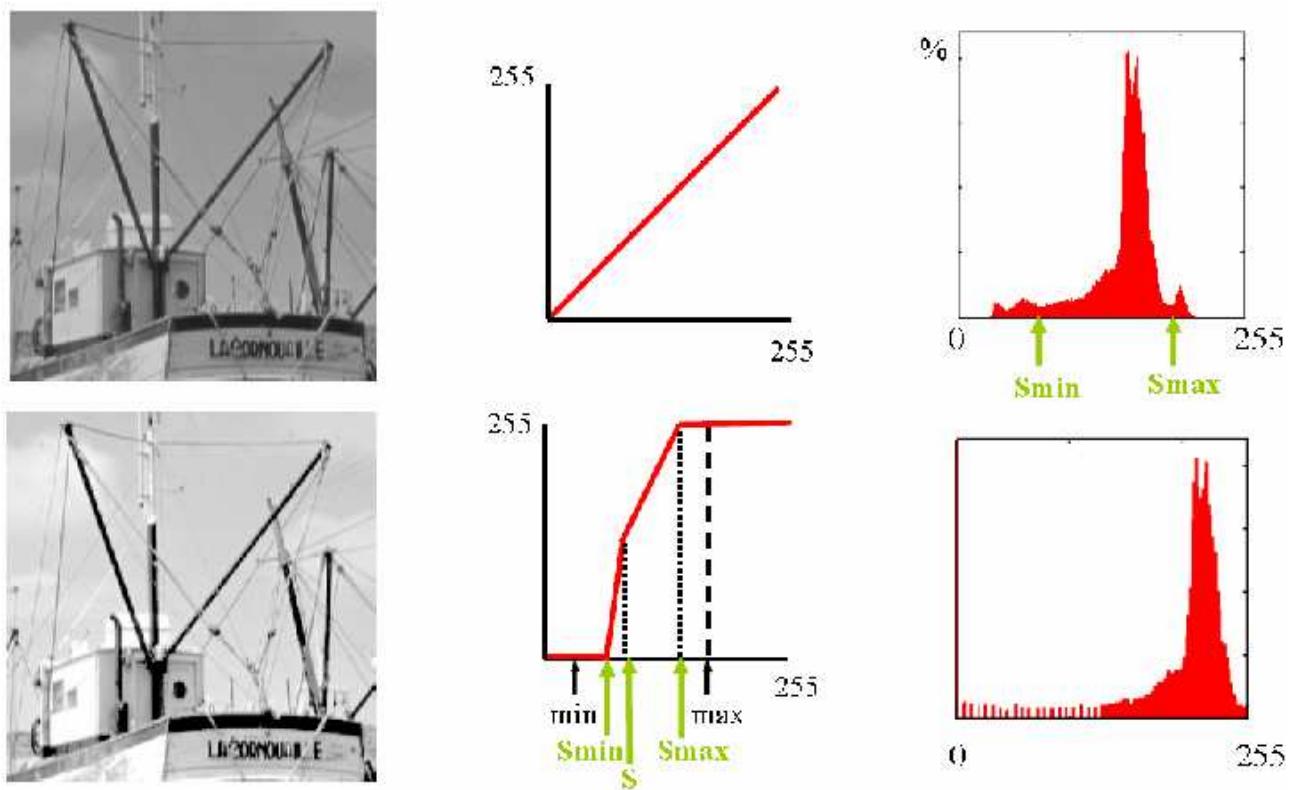


FIGURE 4.39 – Effet d'une transformation linéaire par morceaux.

#### 4.4.2.3 Transformation linéaire par morceaux

À la transformation précédente, il est possible d'ajouter un (ou plusieurs) paramètre supplémentaire  $S$  coupant la rampe de base en deux rampes de pente différente. Cela permet d'accentuer différemment plusieurs zones de la dynamique de l'image de départ. Cette transformation est illustrée à la figure 4.39.

#### 4.4.2.4 Transformation non-linéaire

On peut également imaginer d'utiliser comme loi de transformation une fonction non-linéaire. Par exemple, nous citons ici la *correction Gamma*<sup>3</sup> caractérisée par la loi

$$I'(i, j) = 255 \left( \frac{I(i, j)}{255} \right)^{\frac{1}{\gamma}} \quad (4.33)$$

où  $\gamma$  est compris dans l'intervalle  $[1, 3 ; 3, 0]$  pour compenser les effets dûs à la physique des composants, ou  $[1/2, 1/3]$  pour compenser les effets dûs à la perception visuelle humaine. L'effet de cette transformation est illustrée à la figure 4.40.

#### 4.4.2.5 Autre fonction possible... Le négatif d'une photo

La transformation envisagée ici n'est pas à proprement parler une amélioration du contraste mais plutôt une inversion complète de la dynamique de l'image afin de faire ressortir ce que l'on appelle le "négatif d'une photo". Néanmoins, cette transformation rentre dans la catégorie

3. Cette correction permet de compenser les effets non-linéaires de la reproduction de l'intensité lumineuse. Ce facteur Gamma s'explique par divers phénomènes physiques relevant non seulement de la physique (écrans à tube cathodique, écrans à cristaux liquides (LCD), photographie, acquisition video, impression) mais aussi de la perception visuelle humaine.

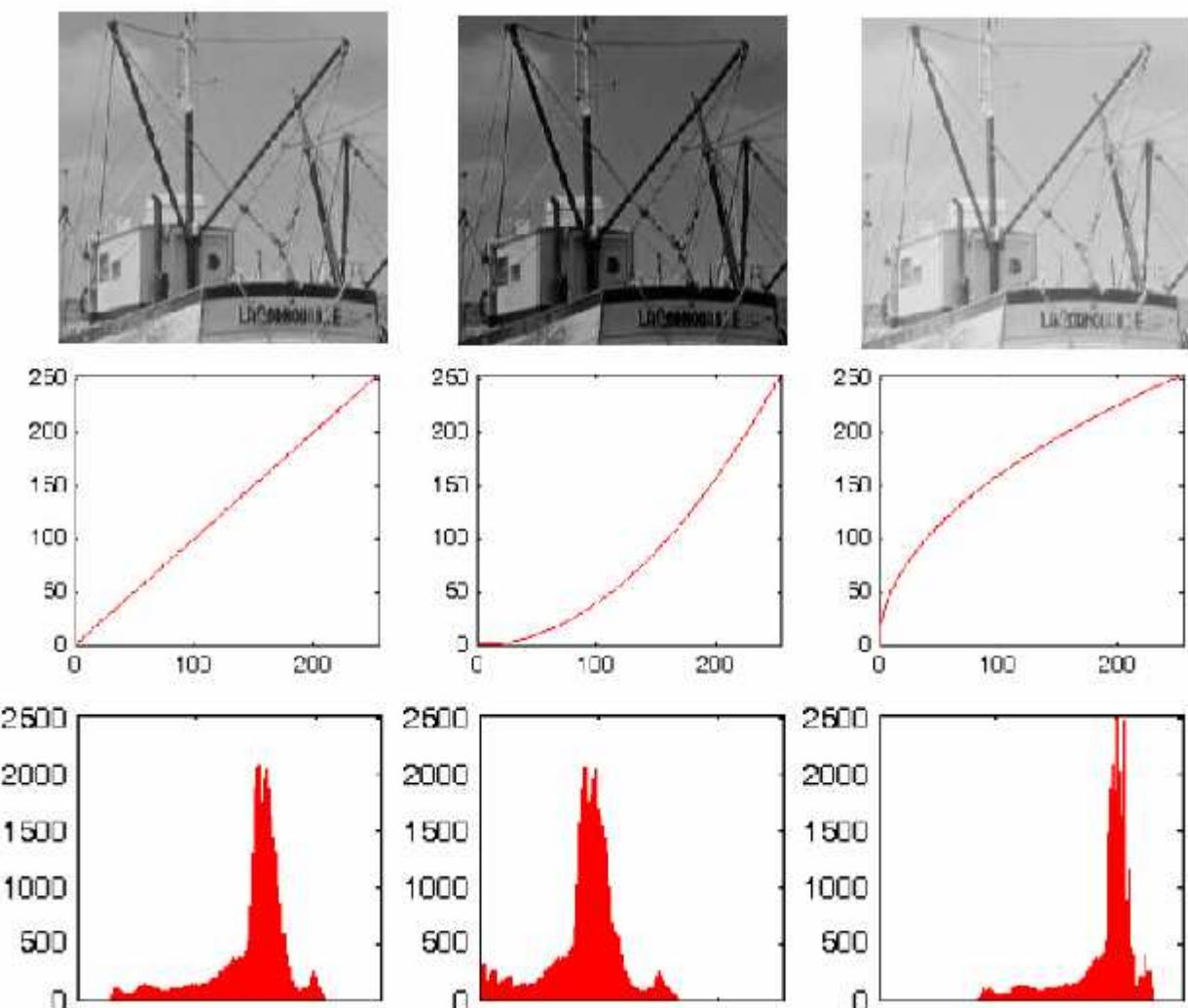


FIGURE 4.40 – Effet de la correction Gamma sur une image en 256 niveaux de gris. (À gauche) image originale. (Au milieu)  $\gamma < 1$ , effet assombrissant. (À droite)  $\gamma > 1$ , effet éclaircissant.

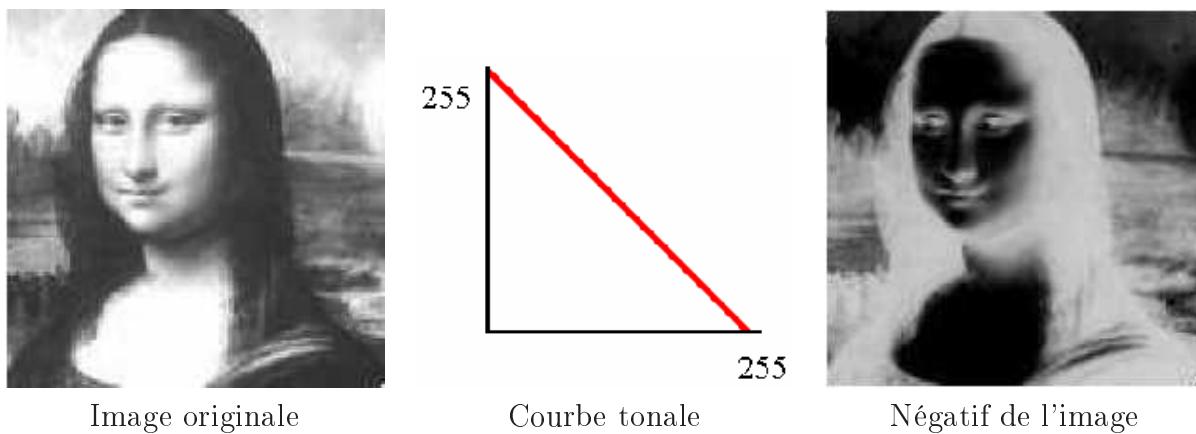


FIGURE 4.41 – Négatif d'une image.

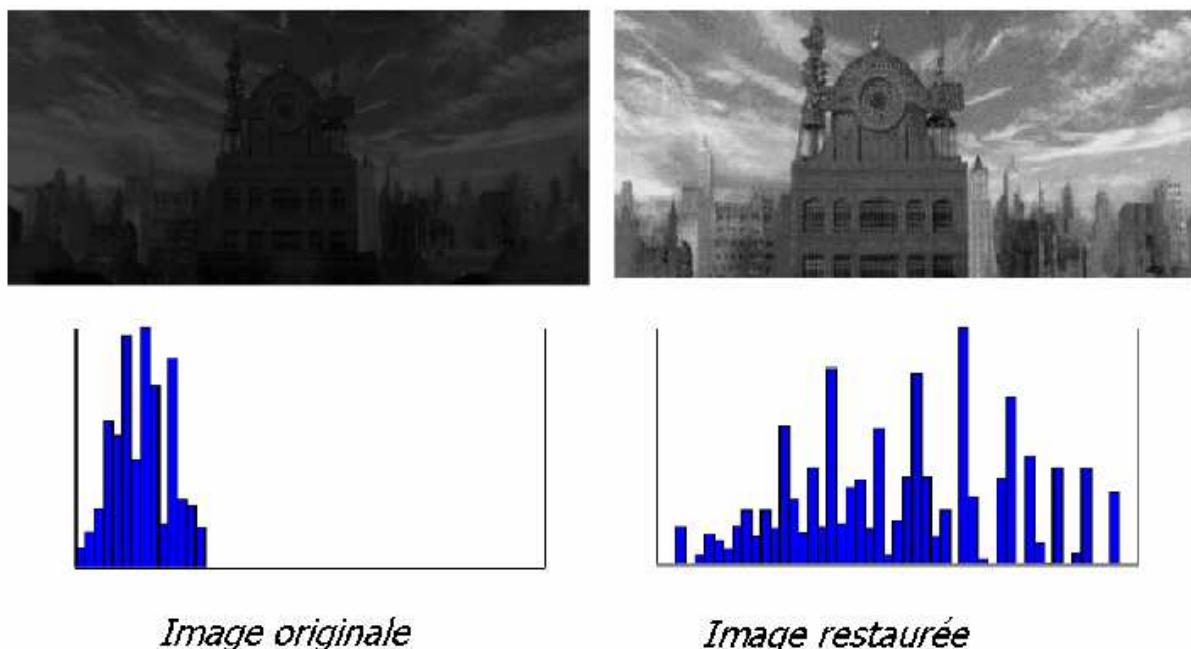


FIGURE 4.42 – Effet bénéfique d'une transformation de l'histogramme.

des opérations de manipulation de l'histogramme. La loi associée à cette transformation est simplement

$$I'(i, j) = 255 - I(i, j) \quad (4.34)$$

Celle-ci est illustrée à la figure 4.41.

#### 4.4.3 Égalisation de l'histogramme

Les techniques de transformation de l'histogramme vues plus haut donnent souvent de bons résultats. Elles dilatent la dynamique de l'image d'entrée afin que l'image de sortie présente une dynamique maximale. Un bon résultat d'une de ces techniques est illustré à la figure 4.42. Dans cet exemple, l'image originale est très sombre et son histogramme est concentré sur les faibles valeurs de niveaux de gris. L'effet de la dilatation de la dynamique de l'image originale est donc très net.

Par contre, dans certains cas, les transformations décrites plus haut ne s'avèrent pas très efficaces. La figure 4.43 illustre une image pour laquelle ces techniques ne donnent pas de bons

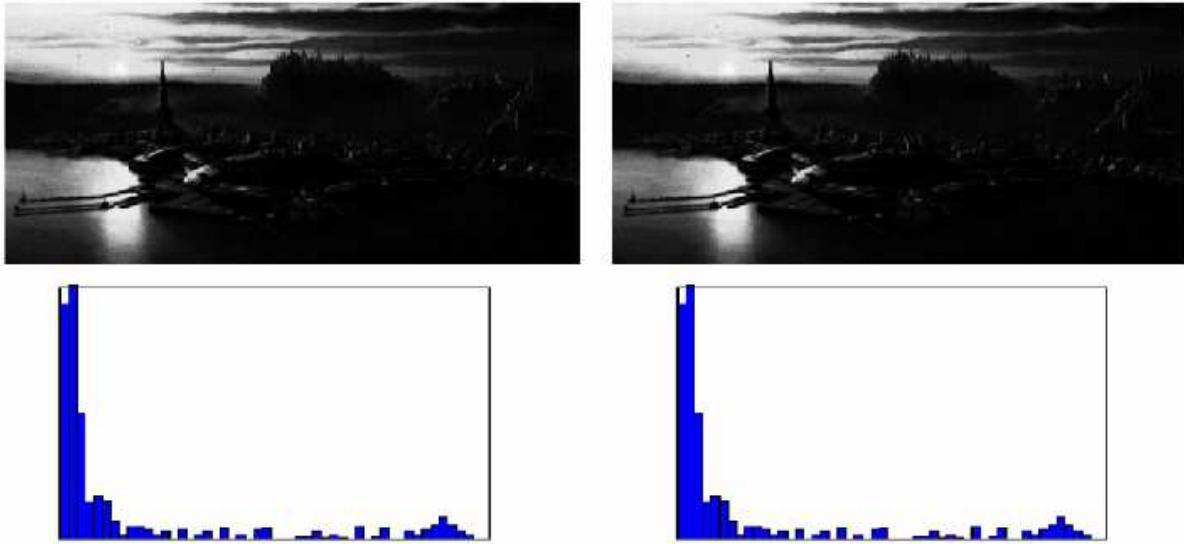


FIGURE 4.43 – Aucun effet bénéfique dans le cas où l'histogramme initial occupe déjà toute la plage de dynamique.

résultats. Il s'agit d'une image sombre mais présentant des zones claires. Dès lors, la dynamique de l'image originale est déjà très large et il est donc difficile de la dilater plus encore. L'histogramme, bien que très large, présente une forte concentration dans les faibles valeurs de niveaux de gris. Dans ce cas, on peut recourir à une autre technique qui porte le nom d'égalisation de l'histogramme.

Dans une image réelle, tous les niveaux de gris ne sont pas présents avec une même occurrence. Il en résulte donc des disparités dans l'histogramme, comme nous venons de le voir avec l'image de la figure 4.43. L'égalisation de l'histogramme est une méthode courante de rehaussement. Elle vise à assurer une distribution homogène des valeurs dans la totalité de la plage dynamique des valeurs possibles de niveaux de gris. Il s'agit donc d'une distorsion de l'échelle des valeurs.

La loi de la transformation de l'égalisation de l'histogramme est calculée à partir de l'histogramme même de l'image originale. Elle varie donc pour chaque image. Plusieurs étapes sont nécessaires pour réaliser l'égalisation d'une image  $f(m, n)$  ( $m = 0, \dots, M-1$  et  $n = 0, \dots, N-1$ ) :

1. Calcul de l'histogramme de l'image originale  $hist(k)$ ,  $k \in [0, 255]$
2. Normalisation de l'histogramme (assimilation d'une probabilité ou fréquence d'occurrence d'un niveau de gris dans l'image) :

$$hist_n(k) = \frac{hist(k)}{NM} \quad \text{avec } k \in [0, 255]$$

3. Calcul de l'histogramme des fréquences cumulées :

$$C(k) = \sum_{l=0}^k hist_n(l) \quad \text{avec } k \in [0, 255]$$

4. Transformation des niveaux de gris en utilisant la loi

$$I'(i, j) = 255 C(I(i, j))$$

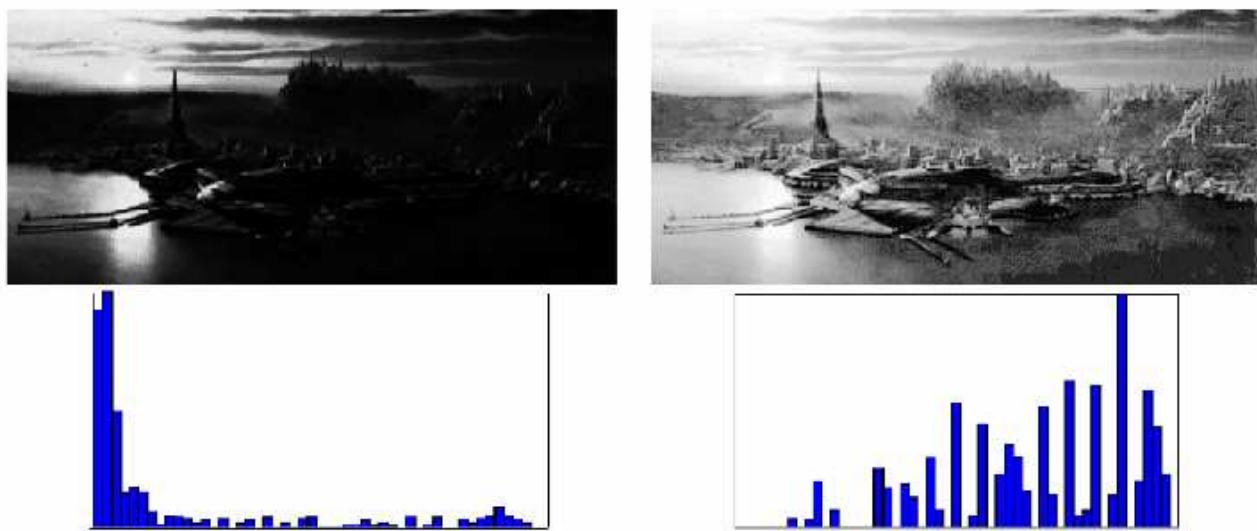


FIGURE 4.44 – Égalisation de l'histogramme d'une image en niveaux de gris.

La figure 4.44 illustre l'application de l'algorithme d'égalisation de l'histogramme sur une image en niveau de gris. L'image résultante est nettement plus contrastée. Par contre, son histogramme n'est pas vraiment uniforme... En fait, cet algorithme vise à rendre l'histogramme cumulatif de l'image résultante le plus linéaire possible, comme en témoigne la figure 4.45 où sont affichés les histogramme et histogramme cumulé d'une image avant et après égalisation.

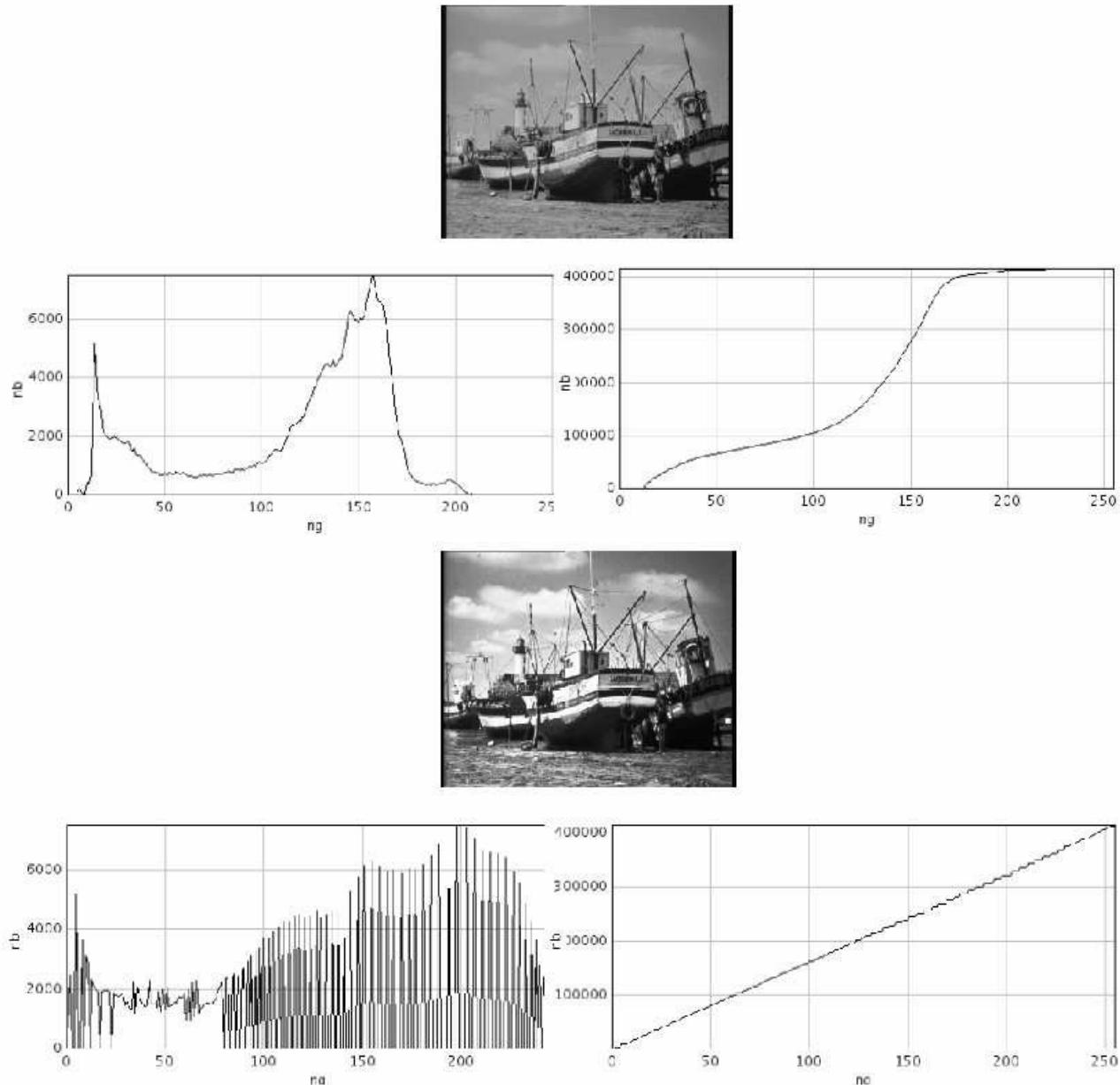


FIGURE 4.45 – Histogramme et histogramme cumulé d'une image avant et après égalisation.

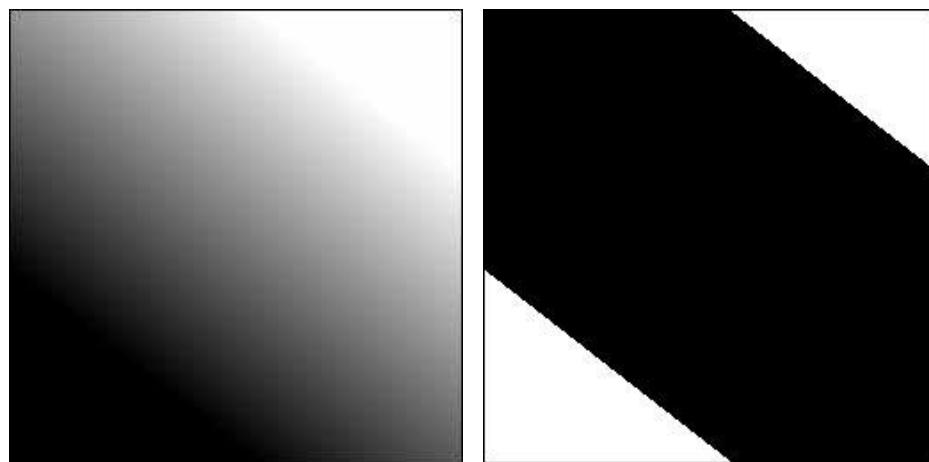


FIGURE 4.46 – Une image en dégradé et ses contours (en noir).

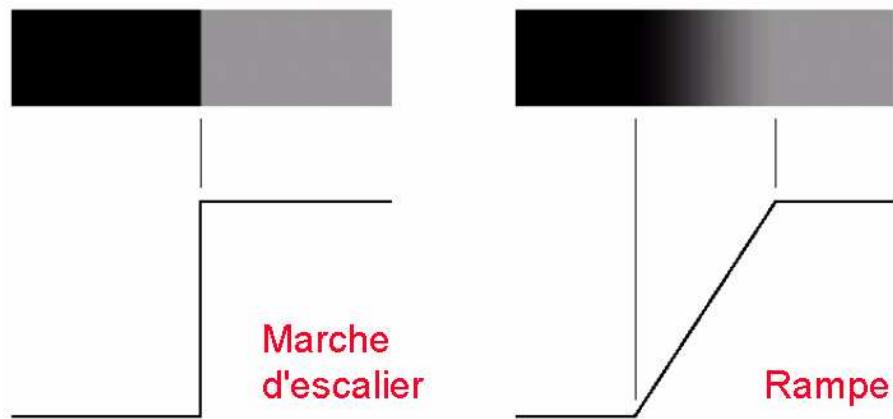


FIGURE 4.47 – Types de contours.

## 4.5 Traitement spécifique : Détection de contours

La détection de contours d'objets présents dans une image est utilisée à différentes occasions, par exemple lors de la segmentation d'une image que nous aborderons un peu plus loin dans ces notes. Intuitivement, un contour est une transition marquée entre deux régions ayant chacune une luminosité distincte. Un contour est donc fondamentalement une transition haute fréquence mais son contenu spectral est très large. C'est là tout le problème de la détection de contours. Enfin, il se peut qu'une image contienne des contours alors que l'oeil ne les perçoit pas. Cet effet est illustré à la figure 4.46. La figure 4.47 illustre dans le même ordre d'idée différents types de contours.

### 4.5.1 Opérateurs linéaires basés sur le calcul des dérivées

Le but d'un opérateur d'extraction de contours est de fournir, en réponse à son application sur une image  $I_1$ , une image  $I_2$  à fortes variations aux différents endroits où  $I_1$  présente des contours. Un opérateur répondant à ce critère est, sans conteste, l'opérateur différentiel ou dérivée. Deux choix s'offrent à nous :

- Si l'on se place dans le contexte d'une fonction continue  $f(x)$  à une dimension, sa dérivée  $f'(x)$  est maximale (en valeur absolue) aux différents endroits où  $f(x)$  présente de fortes variations ; cette valeur maximale étant d'autant plus grande que la variation de  $f$  est importante sur une courte distance. Ceci est illustré à la figure 4.48.

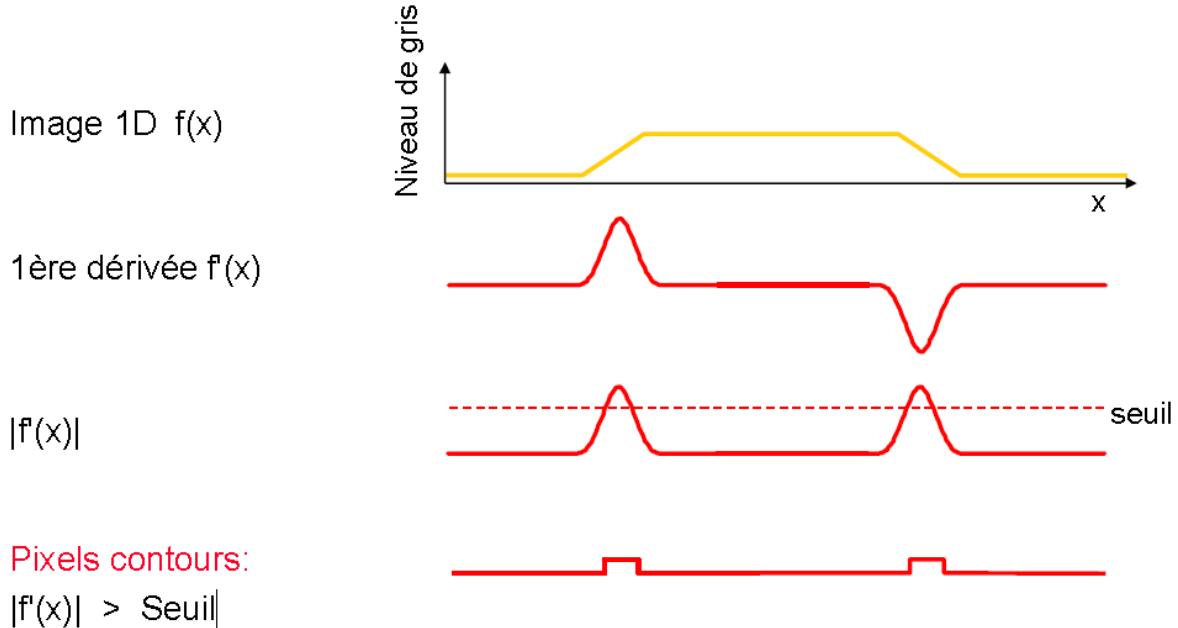


FIGURE 4.48 – Illustration de la dérivée première pour la détection de contours.

2. La dérivée seconde  $f''(x)$  présente, quant à elle, un passage par zéro aux différents endroits où  $f(x)$  contient de fortes variations. Ceci est illustré à la figure 4.49.

L'opérateur dérivée (première ou seconde) est linéaire et peut être vu comme un filtre. Nous verrons plus loin que le filtre dérivée est un filtre de type passe-haut. L'application de l'opérateur dérivée impliquera donc un très net effet accentuateur des hautes fréquences.

Le problème majeur que l'on rencontre en pratique est qu'une image n'est pas une fonction continue. Il n'est donc pas possible d'en calculer la dérivée exacte ; on peut tout juste l'approximer par différentes formules. D'autre part, l'effet accentuateur des hautes fréquences entraîne une amplification du bruit. On est donc face à un double problème :

- approximer au mieux la dérivée et
- éviter une amplification excessive du bruit. L'effet d'amplification du bruit sur la détection des bords est illustré à la figure 4.50.

#### 4.5.1.1 Opérateur de dérivée première et Gradient

Par commodité, nous considérons une image caractérisée par une fonction  $f(x, y)$  bidimensionnelle. Il faut, dans ce cas, indiquer dans quelle direction on calcule la dérivée. Les principaux choix sont les directions horizontales  $\vec{e}_x$  ou verticale  $\vec{e}_y$  mais on pourrait également considérer une direction quelconque  $\vec{\theta}$  en se ramenant à une combinaison linéaire des dérivées suivant les axes principaux - il s'agit alors de dérivée directionnelle.

Considérons la dérivée partielle de la fonction  $f(x, y)$  par rapport à  $x$ . La transformée de FOURIER de cet opérateur vaut

$$\frac{\partial f}{\partial x}(x, y) \rightleftharpoons j2\pi u F(u, v) \quad (4.35)$$

Dériver par rapport à  $x$  équivaut donc à multiplier la transformée de FOURIER de  $f$  par la

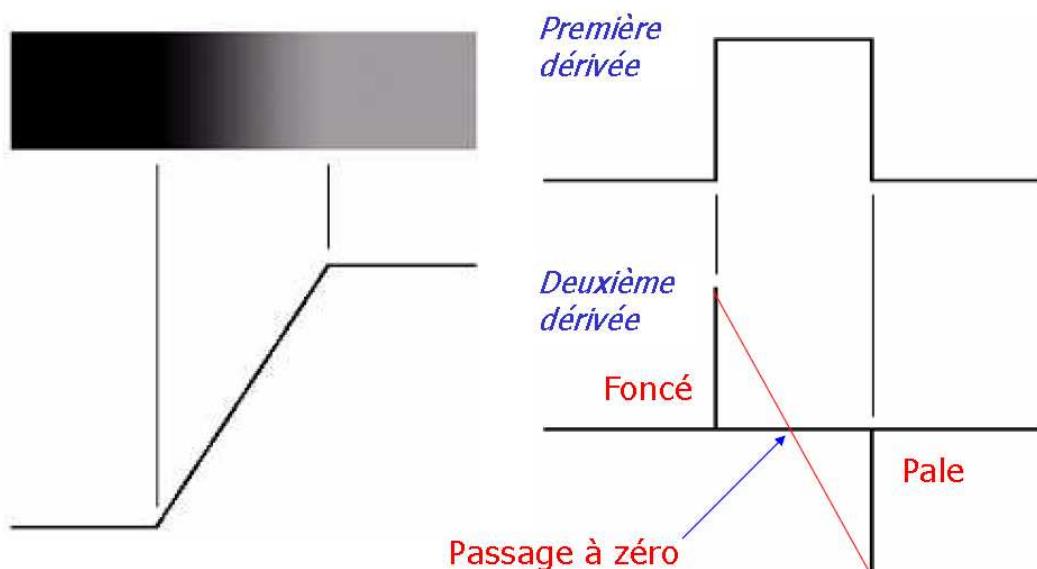


FIGURE 4.49 – Illustration de la dérivée seconde pour la détection de contours.

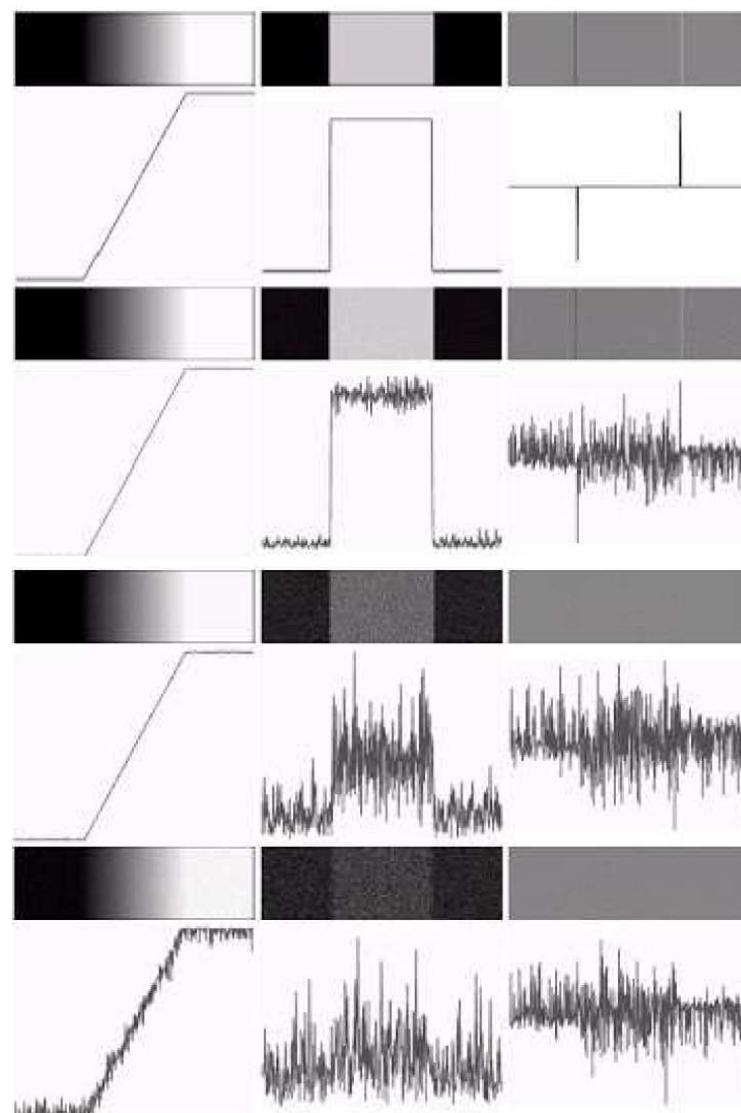


FIGURE 4.50 – Effet du bruit sur la détection des bords par les dérivées première et seconde.

fonction de transfert  $H_x(u, v) = j2\pi u$ , c'est-à-dire filtrer  $f(x, y)$  avec le filtre de réponse impulsionale  $h_x(x, y)$  (transformée de FOURIER inverse de  $H_x$ ). De même, on peut définir la réponse impulsionale  $h_y(x, y)$  correspondant au filtre dérivée dans la direction  $\vec{e}_y$ . Le module de la fonction de transfert étant égal à  $2\pi|u|$ , on voit clairement qu'il y a un effet accentuateur des hautes fréquences, d'où le caractère passe-haut de l'opérateur dérivée. C'est pareil pour  $h_y(x, y)$ .

En adoptant une notation vectorielle de la dérivée, on définit le gradient  $\nabla f$  de l'image  $f$  par

$$\nabla f = \frac{\partial f}{\partial x} \vec{e}_x + \frac{\partial f}{\partial y} \vec{e}_y \quad (4.36)$$

Le gradient d'une fonction  $f(x, y)$  est donc une fonction vectorielle. Mais plutôt que de recourir aux composantes en  $x$  et  $y$ , on peut caractériser un gradient par son amplitude et sa direction.

### Amplitude du gradient

L'amplitude du gradient est donnée par

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \quad (4.37)$$

On approxime parfois l'amplitude du gradient par l'expression

$$|\nabla f| \simeq \left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial f}{\partial y} \right| \quad (4.38)$$

pour éviter le calcul de la racine carrée et les élévations au carré afin d'accélérer le calcul mais cette formule n'est qu'une approximation. La figure 4.51 illustre le calcul de  $|\nabla f|$  par l'utilisation de cette approximation.

### Direction du gradient

La direction du gradient est donnée par

$$\tan^{-1} \left( \frac{\left( \frac{\partial f}{\partial y} \right)}{\left( \frac{\partial f}{\partial x} \right)} \right) \quad (4.39)$$

### Dérivée directionnelle

Étant donnée la définition du gradient d'une fonction, il est maintenant possible de définir sa dérivée directionnelle par

$$\frac{\partial f}{\partial \vec{\theta}} = \vec{\theta} \cdot \nabla f \quad (4.40)$$

Il s'agit du produit scalaire du gradient de  $f$  et d'un vecteur caractérisant la direction dans laquelle on désire calculer la dérivée. Si la direction  $\vec{\theta}$  fait un angle  $\theta$  avec l'axe  $x$ , c'est-à-dire  $\vec{\theta} = \cos \theta \vec{e}_x + \sin \theta \vec{e}_y$ , la dérivée directionnelle peut encore s'écrire

$$\frac{\partial f}{\partial \vec{\theta}} = \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} \quad (4.41)$$

Pour  $\theta = 0$ , nous retrouvons la dérivée directionnelle dans la direction  $\vec{e}_x$ . La dérivée directionnelle est utile pour mettre en évidence les contours d'une image qui ont une orientation particulière.

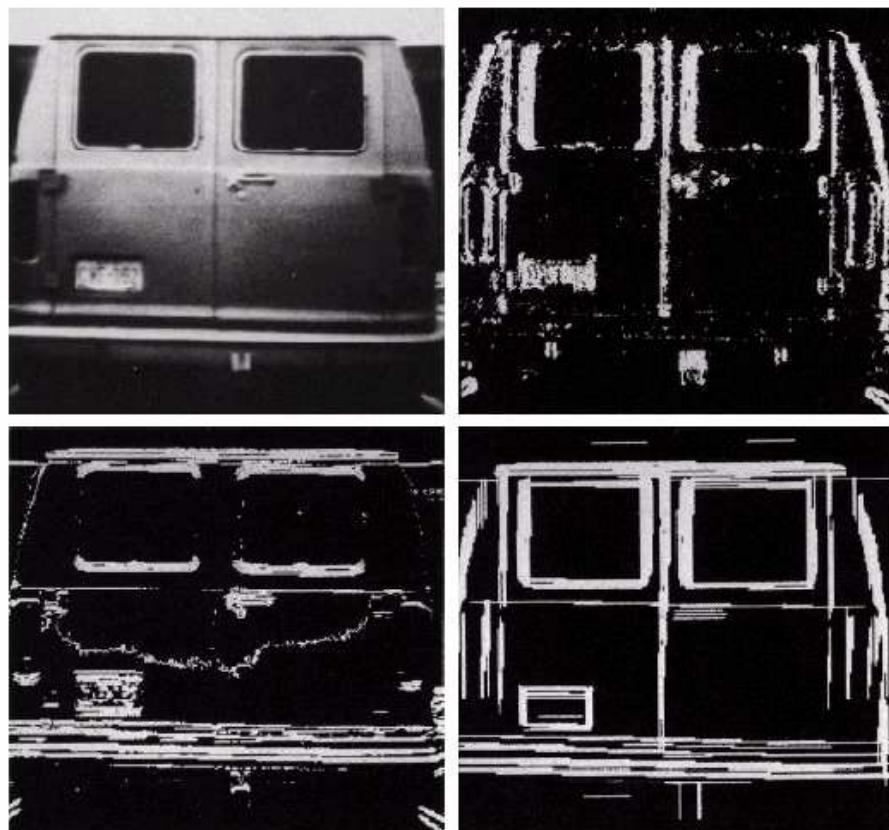


FIGURE 4.51 – Calcul de l'amplitude du gradient d'une image. En haut à gauche : image originale  $f$ . En haut à droite :  $|\frac{\partial f}{\partial x}|$ . En bas à gauche :  $|\frac{\partial f}{\partial y}|$ . En bas à droite :  $|\frac{\partial f}{\partial x}| + |\frac{\partial f}{\partial y}|$ .

#### 4.5.1.2 Opérateur de dérivée seconde et Laplacien

Comme pour la dérivée première, on peut définir les réponses impulsionales  $h_{xx}(x, y)$  et  $h_{yy}(x, y)$  correspondant aux filtres dérivée seconde dans les directions  $\vec{e}_x$  et  $\vec{e}_y$ . On définit alors le laplacien d'une fonction par

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (4.42)$$

On peut calculer facilement la transformée de FOURIER de cet opérateur

$$\nabla^2 f \rightleftharpoons -4\pi^2(u^2 + v^2) F(u, v) \quad (4.43)$$

Ici encore, on remarque l'effet accentuateur des hautes fréquences.

#### 4.5.1.3 Calcul pratique des dérivées et masques de convolution

En pratique, les images sur lesquelles nous devons calculer la dérivée sont échantillonnées et sont fournies sous la forme d'une matrice de pixels. Il est alors nécessaire d'échantillonner l'opérateur de dérivée première et seconde en les approximant au mieux.

##### Dérivée première

Considérons une fonction  $f(x)$  à une dimension. Une première approximation de sa dérivée première est donnée par

$$f'(x) \simeq \frac{f(x+h) - f(x-h)}{2h} \quad (4.44)$$

où  $h$  est la pas d'échantillonnage. Cette approximation correspond à l'utilisation du masque de convolution centré suivant

$$\frac{1}{2h} \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \quad (4.45)$$

Pour son application, le facteur  $\frac{1}{2h}$  est en général omis. Ce petit masque s'applique directement sur la grille de pixels de l'image à traiter, en plaçant le 0 sur le pixel en cours de traitement. Une approximation de la dérivée première dans la direction  $y$  est donnée par

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad (4.46)$$

Dans la foulée, on peut imaginer l'utilisation du masque suivant

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (4.47)$$

pour procéder à une dérivation directionnelle dans la direction diagonale  $135^\circ$ . La figure 4.52 montre le résultatat de l'application de divers opérateurs de dérivée première. L'image originale est définie sur 256 niveaux de gris ; les images gradients ont été décalées de 128. En effet, l'opérateur de dérivation modifie la dynamique de l'image originale. Si celle-ci présente une dynamique de  $[0, 255]$ , l'image gradient présente une dynamique de  $[-255, 255]$ .

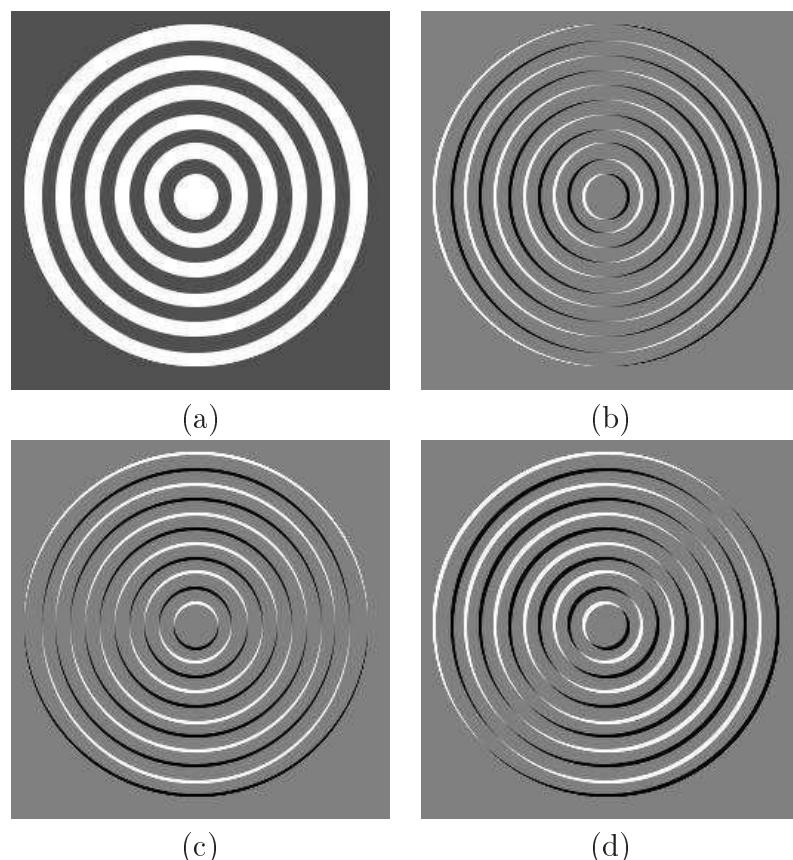


FIGURE 4.52 – Illustrations de l’application de masques de dérivées premières : (a) Image originale. (b) Application d’un masque horizontal. (c) Application d’un masque vertical. (d) Application d’un masque à  $135^\circ$ .

### Filtres gradients de PREWITT

Les formes de base effectuent un gradient ligne par ligne ou colonne par colonne. Les directions horizontales et verticales sont dès lors privilégiées. Il existe des formes qui permettent de fournir un gradient aux caractéristiques plus isotropes et qui permettent en plus de limiter l'effet du bruit en réalisant un filtrage passe-bas dans la direction perpendiculaire à celle dans laquelle on calcule la dérivée. Pour y parvenir, ces formes s'étendent dans les deux directions. Les filtres gradients de PREWITT appartiennent à cette classe d'opérateur gradient. Ils valent

$$\frac{1}{3} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{et} \quad \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad (4.48)$$

### Filtres gradients de SOBEL

Les filtres de SOBEL constituent une alternative aux filtres de PREWITT. Ils s'expriment comme suit

$$\frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{et} \quad \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (4.49)$$

### Dérivée seconde et filtres Laplacien

Considérons à nouveau une fonction  $f(x)$  à une dimension. Pour la dérivée seconde, nous avons l'approximation suivante

$$f''(x) \simeq \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \quad (4.50)$$

Cette approximation conduit tout naturellement aux deux masques de convolution suivant

$$\begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad (4.51)$$

On peut obtenir un filtre bidimensionnel en combinant les deux expressions précédentes ou lui préférer un masque du type

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (4.52)$$

ou encore

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (4.53)$$

La figure 4.53 illustre l'application de ces différents masques.

### 4.5.2 Opérateurs non-linéaires basés sur la morphologie mathématique

Si  $B$  est un élément structurant de petite taille, la différence ensembliste  $X - (X \ominus B)$  fournit les contours intérieurs de l'image  $X$ , alors que la différence  $(X \oplus B) - X$  produit les contours extérieurs de l'image. On peut réaliser des opérations similaires sur des images en niveaux de gris.

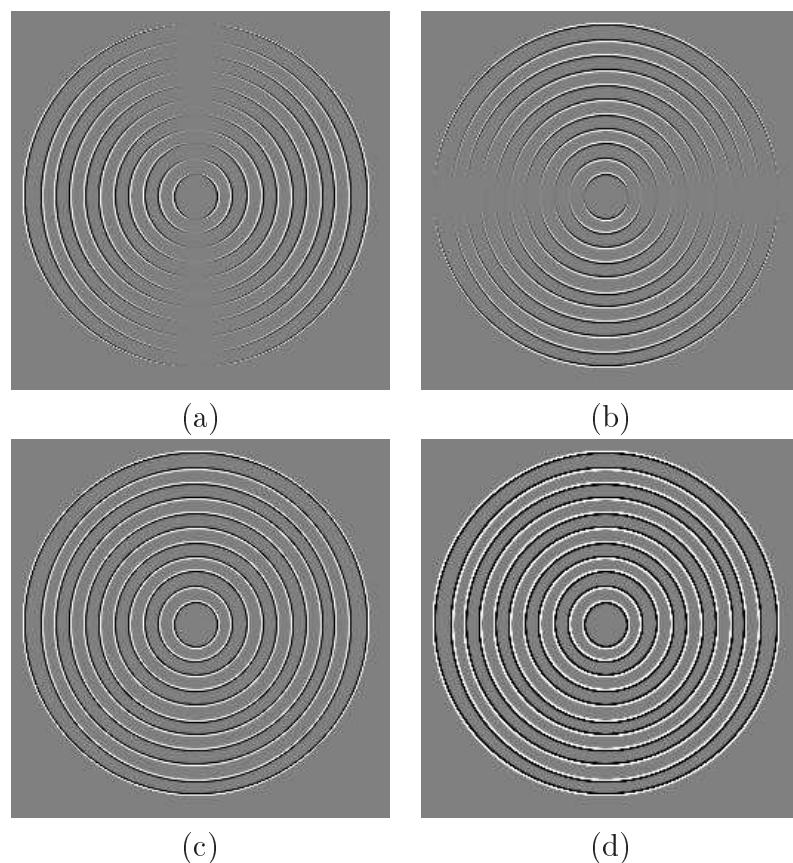


FIGURE 4.53 – Illustrations de l’application de masques de dérivées secondees (décalage de 128) : (a) Laplacien horizontal. (b) Laplacien vertical. (c) Laplacien obtenu avec le masque (4.52). (d) Laplacien obtenu avec le masque (4.53).

## Gradient d'érosion

L'opérateur

$$GE(f) = f - (f \ominus B) \quad (4.54)$$

appelé gradient d'érosion constitue un moyen commode d'accentuer les transitions d'une image en niveaux de gris  $f$ .

## Gradient de dilatation

Un opérateur qui remplit la même fonction que le précédent est le gradient de dilatation

$$GD(f) = (f \oplus B) - f \quad (4.55)$$

Par combinaison de ces deux opérateurs, il est possible de synthétiser une kyrielle de nouveaux opérateurs, tous destinés à extraire les contours de l'image. L'objectif d'une utilisation des deux opérateurs est la symétrisation du traitement d'une image et de son arrière-fond. Les exemples sont nombreux (voir figures 4.54 et 4.55).

## Gradient de BEUCHER

Il est défini par

$$GE(f) + GD(f) = (f \oplus B) - (f \ominus B) \quad (4.56)$$

et est illustré à la figure 4.54.

## Chapeau haut-de-forme (“top-hat” en anglais)

Cet opérateur est égal à l'image moins l'ouvert :

$$f - (f \circ B) \quad (4.57)$$

## Opérateurs de détection de contours marqués

Ils sont définis par

$$\min(GE(f), GD(f)) \quad (4.58)$$

et

$$\max(GE(f), GD(f)) \quad (4.59)$$

## Laplacien non-linéaire

Enfin, le laplacien non-linéaire est défini par

$$GD(f) - GE(f) \quad (4.60)$$

Tous ces opérateurs sont illustrés à la figure 4.55.

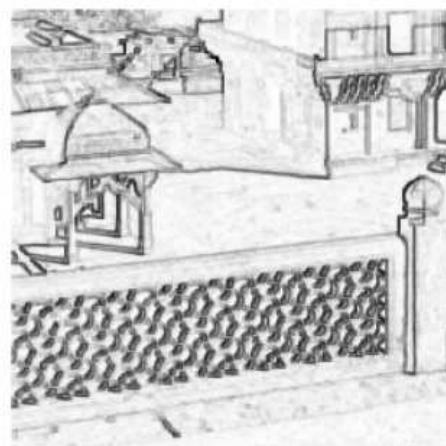
(a) Image originale  $f$ (b)  $f \oplus B$ (c)  $f \ominus B$ (d)  $(f \oplus B) - (f \ominus B)$  (vidéo inverse)

FIGURE 4.54 – Gradient de BEUCHER.

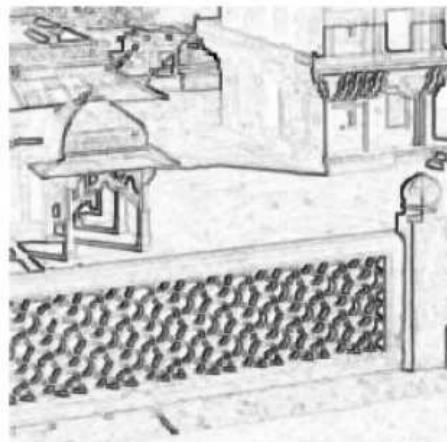
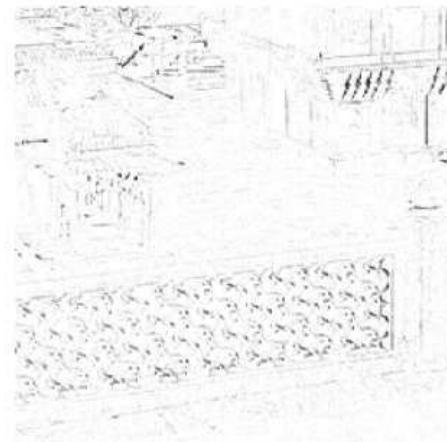
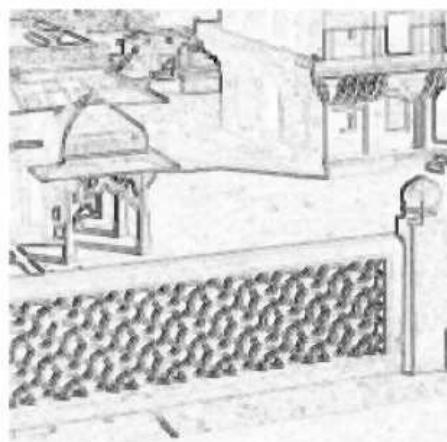
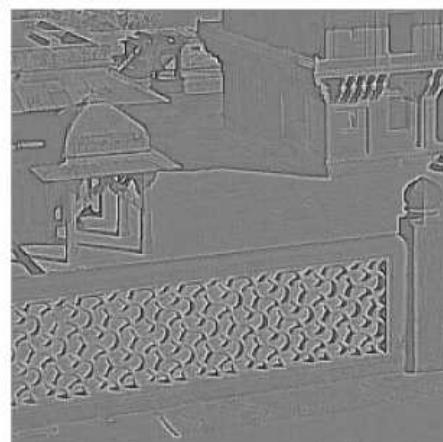
(a)  $(f \oplus B) - (f \ominus B)$ (b)  $f - (f \circ B)$ (c)  $\max(GE(f), GD(f))$ (d)  $GD(f) - GE(f)$ 

FIGURE 4.55 – Différents détecteurs de contours non-linéaires (vidéo inverse).



FIGURE 4.56 – Une image originale et le résultat d'une segmentation.

## 4.6 Traitement spécifique : Segmentation et seuillage

La segmentation est une des branches du traitement d'images qui s'occupe de l'analyse d'image et de scène. Elle vise donc des applications automatisées de vision par ordinateur et la robotique. Dans ce domaine, l'entrée est toujours une image mais la sortie est une description de l'image. La plupart des descriptions nécessitent une détection préalable des formes présentes dans l'image. C'est cette étape qui est appelée segmentation. Les techniques de segmentation sont nombreuses et présentent chacune des avantages et des inconvénients. Le plus souvent, il faudra trouver la méthode adaptée à l'image traitée. Dans cette introduction, nous nous limiterons à la segmentation par seuillage.

### 4.6.1 Définition

Le seuillage vise à sous-diviser l'image en constituants distincts appelés objets ou encore régions selon l'application visée. Par exemple, la figure 4.56 illustre une image d'une fleur et le résultat d'une segmentation ayant extrait trois régions : le fond, les pétales et le pistille.

La segmentation est donc normalement basée sur

- les discontinuités : arrêtes, changements abruptes d'intensité, ...
- les similitudes (zones homogènes) : couleurs, textures, intensités, ...

et plusieurs approches sont possibles :

- l'*approche régions* qui consiste à rechercher les zones dans l'image sur un critère d'homogénéité (comme à la figure 4.56),
- l'*approche contours* qui consiste à rechercher les discontinuités entre régions. Dans ce cas, les opérateurs de détection de contours que nous avons étudiés plus haut sont utilisés. Cependant, les détecteurs ne donnent pas des contours fermés et sont sensibles au bruit. Dans ce contexte, une étape de seuillage du gradient (ou du laplacien) est nécessaire au préalable. Ensuite, il faut fermer les contours et conserver les contours significatifs. Ce qui n'est pas tâche aisée...
- l'*approche duale* qui vise à combiner les deux approches précédentes. En effet, il existe une dualité entre régions et contours. Une région est délimitée par un contour et un contour sépare deux régions.

Une méthode simple et très populaire pour la segmentation de régions dans les images est le seuillage qui peut être binaire ou à plusieurs niveaux (le seuillage de la figure 4.56 est à 3 niveaux).

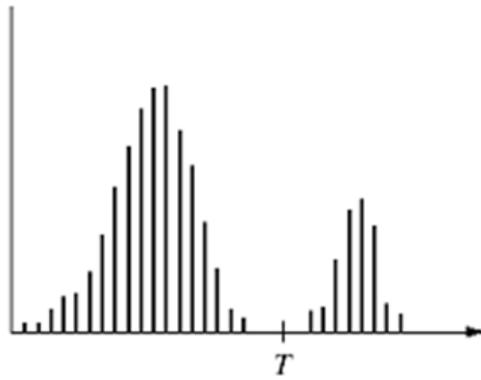


FIGURE 4.57 – Choix du seuil dans le cas d'un histogramme simple.

#### 4.6.2 Seuillage simple

La méthode consiste à choisir un seuil  $T$  qui va séparer les pixels de l'image en deux zones, une comprenant les pixels d'intensité supérieure à  $T$  et ceux d'intensité inférieure à  $T$ . Pour une image en niveau de gris  $f(m, n)$ , l'image résultant de la segmentation est alors une *image binaire*  $g$  donnée par

$$g(m, n) = \begin{cases} 1 & \text{si } f(m, n) > T \\ 0 & \text{si } f(m, n) \leq T \end{cases} \quad (4.61)$$

Le principal souci est le choix de la valeur du seuil  $T$ . Lorsque l'histogramme est clairement *bimodal*, comme celui de la figure 4.57, le fond est bien séparé des objets présents dans l'image et la segmentation donne de bons résultats. La figure 4.58 illustre un résultat de seuillage simple valable<sup>4</sup>. Il est clair que pour obtenir un résultat valable, il doit y avoir un contraste suffisant entre les objets à extraire et le fond. Si tel n'est pas le cas, il faut prévoir un traitement préalable de rehaussement ou de restauration de l'image<sup>5</sup>.

Par contre, lorsque l'image originale présente des problèmes d'éclairage non uniforme, problème courant en pratique, le seuillage simple ne donne pas de bons résultats. La figure 4.59 illustre la situation. Dans ce cas, il est nécessaire de recourir à des techniques plus complexes combinant seuillage, opérations morphologiques ou autres... Nous n'entrerons cependant pas dans le détail de ces méthodes.

#### 4.6.3 Seuillage multiple

Le seuillage multiple permet d'extraire d'une image plusieurs régions. Elle est basée sur le choix de plusieurs seuils  $T_1, T_2, \dots$  permettant ainsi de scinder l'image en sous-groupes de pixels. L'image résultant de la segmentation n'est plus binaire mais *multimodale*. Par exemple, pour un seuillage à deux niveaux, l'expression de l'image résultante est

$$g(m, n) = \begin{cases} 2 & \text{si } f(m, n) > T_2 \\ 1 & \text{si } T_2 \geq f(m, n) > T_1 \\ 0 & \text{si } f(m, n) \leq T_1 \end{cases} \quad (4.62)$$

où les valeurs 0, 1, 2 sont arbitraires et peuvent donc être choisies en fonction de l'utilisation que l'on veut faire du résultat. Un résultat de seuillage multiple est présenté à la figure 4.56.

4. Source : [http://www.ext.upmc.fr/urfirfirst/image\\_numerique/segmentation.html](http://www.ext.upmc.fr/urfirfirst/image_numerique/segmentation.html)

5. Dans les applications industrielles, il est préférable de travailler en environnement contrôlé, c'est-à-dire qu'il est nécessaire de mettre en place un dispositif efficace d'éclairage afin d'obtenir des résultats de segmentation optimaux.

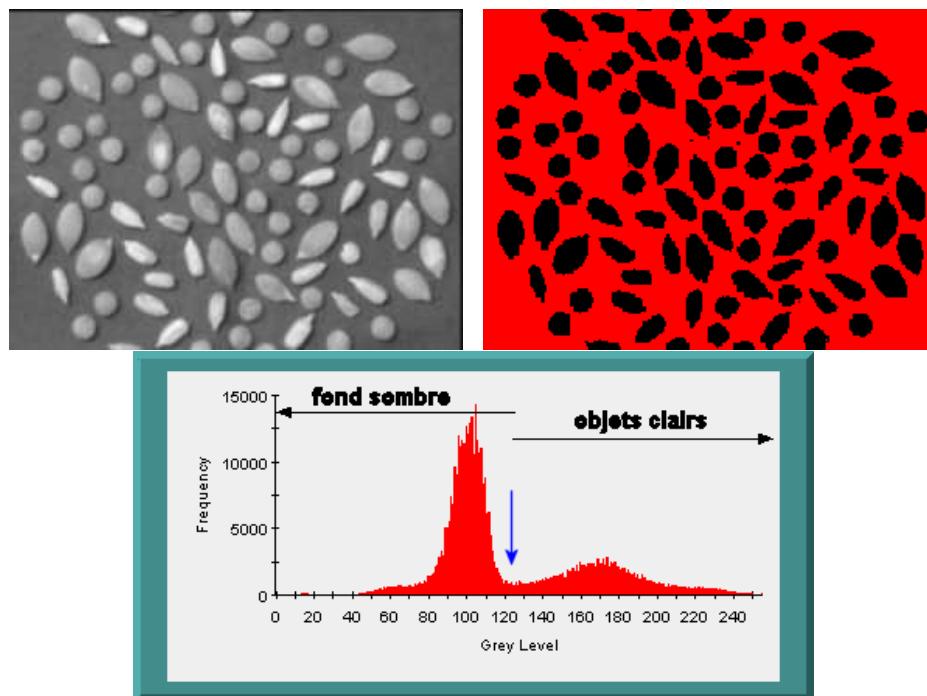


FIGURE 4.58 – Seuillage simple d'une image en 256 niveaux de gris ( $T = 120$ ).

Les avantages des méthodes de seuillage sont assez évidents. Elles sont simples à implémenter et permettent le temps réel étant donné la simplicité des opérations. Par contre, les inconvénients sont également présents : apparition de faux éléments, choix entre seuillage simple ou multiple, et surtout quels choix réaliser pour le (les) seuil(s) ? Voici quelques moyens simples permettant de choisir une valeur de seuil dans le cas du seuillage simple :

- valeur obtenue par essais-erreurs (on aurait pu s'en douter... mais ce n'est pas vraiment à cela que l'on s'attendait !)
- valeur moyenne des tons de gris
- valeur médiane entre le ton maximum et le ton minimum
- ...

Plusieurs algorithmes ont néanmoins vu le jour afin de réaliser un seuillage automatique en recherchant une valeur du seuil qui balance au mieux les deux sections de l'histogramme. Nous en présentons un ci-après.

#### 4.6.4 Seuillage automatique

Voici un exemple d'algorithme simple de seuillage simple automatique :

1. Choisir un seuil  $T$  initial (moyenne, médiane, ...)
2. On obtient alors 2 groupes de pixels : le groupe  $G_1$  contenant les pixels d'intensité supérieure à  $T$  et le groupe  $G_2$  contenant les pixels d'intensité inférieure à  $T$ .
3. Calculer les moyennes des niveaux de gris sur les deux groupes de pixels  $G_1$  et  $G_2$  ; ceci fournit les deux moyennes  $\mu_1$  et  $\mu_2$ .
4. Mettre à jour la valeur du seuil

$$T \rightarrow \frac{\mu_1 + \mu_2}{2}$$

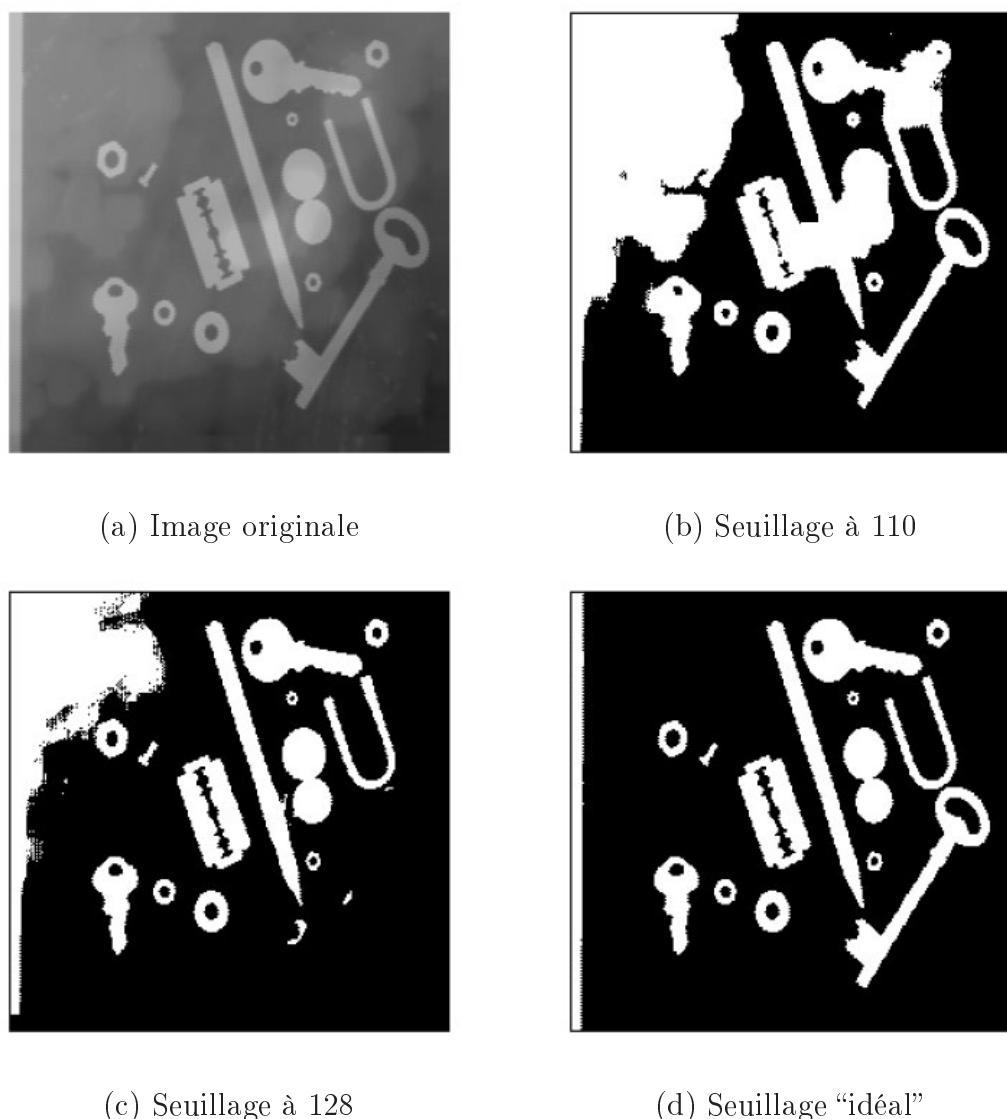


FIGURE 4.59 – Seuillage simple dans le cas de problème d'éclairage non uniforme.

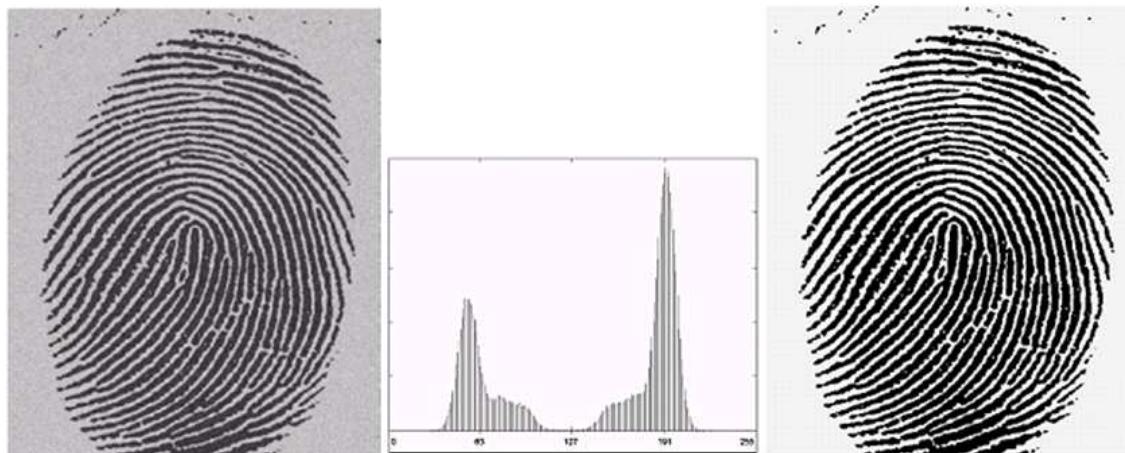


FIGURE 4.60 – Résultat de l’algorithme de seuillage automatique, seuil trouvé :  $T = 125$ .

5. Recommencer à l’étape 1 jusqu’à ce que la valeur du seuil  $T$  converge vers une valeur constante.

Le résultat de l’application de cet algorithme est illustré à la figure 4.60. Évidemment, pour qu’un tel algorithme fournisse de bons résultats, il faut à nouveau un contraste important entre le fond et les objets à détecter.

## 4.7 Bibliographie

1. “Traitement numérique des images”, M. VAN DROOGENBROECK, Université de Liège, Faculté des Sciences appliquées.  
URL : <http://www2.ulg.ac.be/telecom/teaching/notes/totali/elen016/index.html>
2. “Traitement d’images”, A. BOUCHER, Institut de la Francophonie pour l’informatique.  
URL : [http://www.ifi.auf.org/personnel/Alain.Boucher/cours/traitement\\_images/index.html](http://www.ifi.auf.org/personnel/Alain.Boucher/cours/traitement_images/index.html)



# Quatrième partie

## Analyse de FOURIER



# Chapitre 5

## Séries de FOURIER

Ce chapitre a pour but de définir les séries de FOURIER. Celles-ci constituent un outil essentiel dans le domaine de l'analyse et du traitement du signal.

### 5.1 Introduction

L'étude de la musique et en particulier des signaux audio est une manière concrète d'aborder la théorie mathématique des séries de FOURIER. Un son est une onde sonore, “plus ou moins” périodique qui se propage dans l'air, le “plus ou moins” venant du fait qu'un son “harmonieux” peut évoluer au cours du temps. Un son peut être représenté à l'aide d'une fonction mathématique, dont un exemple est donné à la figure 5.1.

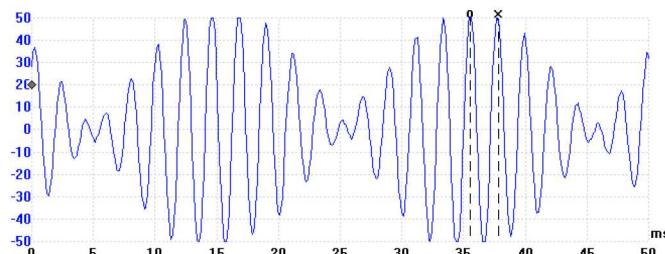


FIGURE 5.1 – Un son, caractérisé par une certaine répétition.

Le fait qu'un son soit plus ou moins périodique le différencie de ce qu'on appelle communément un bruit. Un exemple de bruit est donné à la figure 5.2.

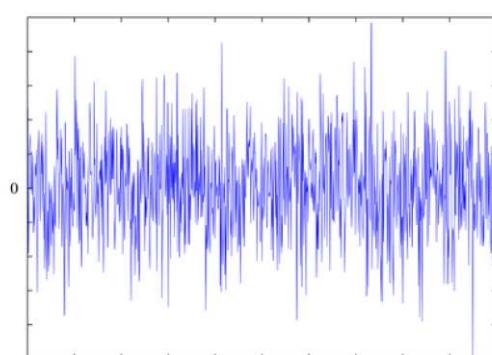


FIGURE 5.2 – Du bruit, aucun aspect répétitif ni structuré.

Apparemment, créer un son consiste donc à créer des déformations périodiques de l'air. Cependant, cela n'est pas aussi simple que cela. En effet, considérons le son ré à 290 Hz. Ce son correspond donc à une oscillation de la pression de l'air à une cadence de 290 déformations par seconde. Cependant, il existe des milliers de ré à 290 Hz différents, caractérisés chacun par leur spécificité, on parle encore du timbre de ce son. On peut facilement distinguer un ré 290 Hz issu d'un piano de celui généré par une guitare. L'illustration graphique des similitudes et différences d'un même son ré 290 Hz générés par deux instruments différents est fournie aux figures 5.3 et 5.4.

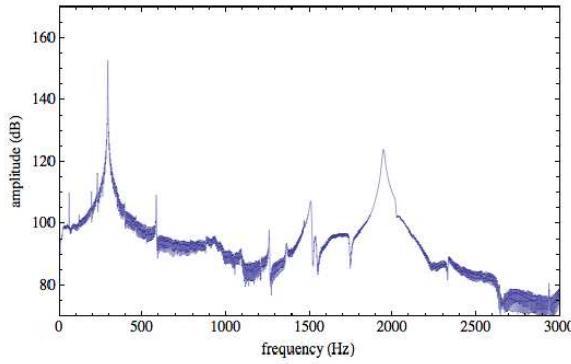


FIGURE 5.3 – Une note jouée sur un piano.

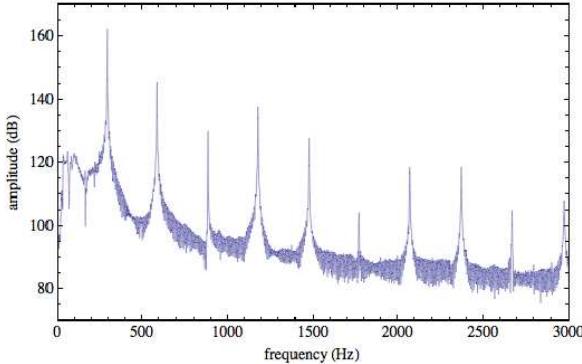


FIGURE 5.4 – Même note jouée sur une guitare.

On peut remarquer une corrélation importante entre les deux graphiques correspondant à la même note pour des instruments différents. On y voit clairement apparaître un “pic”, commun aux deux graphiques, à la fréquence de la note (290 Hz). Ce pic porte le nom d’harmonique fondamentale du son ou de la note, et 290 Hz est appelée “fréquence fondamentale” de la note. Le reste correspond au timbre du son, ce qui différencie un instrument d’un autre. En particulier, sur la figure 5.4, on voit la présence d’autres pics situés aux multiples de la fréquence 290 Hz. Ces pics s’appellent les harmoniques (secondaires) du son ou de la note. Selon l’instrument, l’amplitude des harmoniques secondaires peut être nulle ou non.

Une note de musique peut donc être vue comme la superposition d’une onde (harmonique) fondamentale à la fréquence de la note et d’une multitude d’ondes (harmoniques) secondaires dont les fréquences sont des multiples entiers de la fréquence fondamentale, l’amplitude de ces harmoniques dépendant de l’instrument considéré et constituant le timbre de la note. Nous arrivons donc à la conclusion que nous pouvons créer toute une variété de sons en additionnant une fonction sinusoïdale (onde fondamentale) et un ensemble de fonctions sinusoïdales de

fréquence multiple de la fréquence fondamentale. Mais inversément, peut-on décrire tout son comme une somme d'harmoniques ? La réponse à cette question est donnée par l'étude des séries de FOURIER.

Mais avant de se lancer dans leur étude, il convient de définir quelques notions comme celle de périodicité d'une fonction ou encore celle de fonction continue par morceaux.

### 5.1.1 Fonction continue par morceaux

Avant de donner la définition précise de cette notion, la figure 5.5 montre un exemple de fonction continue par morceaux sur l'intervalle  $]a_0, a_4[$ . Il convient donc tout d'abord de rappeler la notion de continuité.

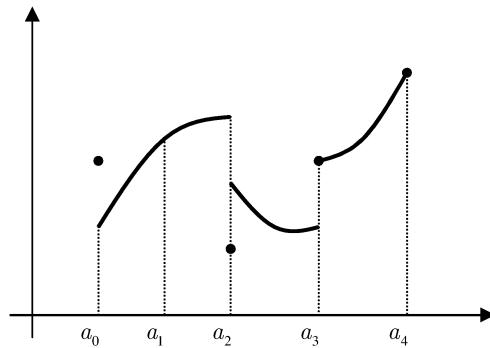


FIGURE 5.5 – Fonction continue par morceaux.

**Rappel (Fonction continue sur un ouvert).** Une fonction  $g$  d'une variable  $t$ , notée  $g(t)$ , est continue sur un intervalle ouvert  $]a, b[$ , ce que l'on note  $g \in C_0(]a, b[)$ , si

- $g(t)$  est définie pour tout  $t \in ]a, b[$ , et
- pour tout  $t_0 \in ]a, b[$ , nous avons

$$\lim_{t \rightarrow t_0} g(t) = g(t_0)$$

Dans cette définition, l'intervalle considéré est ouvert. Qu'en est-il sur le fermé ? Donc...

**Rappel (Fonction continue sur un fermé).** Une fonction  $g(t)$  est continue sur un intervalle fermé  $[a, b]$ , ce que l'on note  $g \in C_0([a, b])$ , si

- $g$  est continue sur l'ouvert  $]a, b[$ , et
- $g$  est définie en  $a$  et  $b$ , et
- la limite à droite en  $a$  et la limite à gauche en  $b$  existent et sont telles que

$$g(a^+) \triangleq \lim_{t \rightarrow a^+} g(t) = g(a) \text{ et } g(b^-) \triangleq \lim_{t \rightarrow b^-} g(t) = g(b)$$

Nous pouvons à présent définir la notion de fonction continue par morceaux.

**Définition (Fonction continue par morceaux).** Une fonction  $g(t)$  est continue par morceaux sur l'intervalle  $[a, b]$  lorsqu'elle ne présente qu'un nombre fini de discontinuités sur cet intervalle, c'est-à-dire si on peut trouver un nombre fini de points  $a_1, \dots, a_p$  ( $a = a_0 < a_1 < \dots < a_p < a_{p+1} = b$ ), tels que pour tout  $i = 0, \dots, p$ ,

- $g(t)$  est continue sur  $]a_i, a_{i+1}[$ , et
- les limites

$$\lim_{t \rightarrow a_i^+} g(t) \text{ et } \lim_{t \rightarrow a_{i+1}^-} g(t)$$

sont des réels.

En particulier, la fonction donnée en exemple à la figure 5.5 est continue sur  $]a_0, a_2[, ]a_2, a_3[$  et  $]a_3, a_4]$  et présente des discontinuités en  $a_0, a_2$  et  $a_3$ .

### 5.1.2 Fonctions périodiques

Intuitivement, une fonction est périodique si elle se répète à intervalles réguliers. Précisons.

**Définition (Fonction périodique).** Une fonction  $g(t)$  est périodique si elle satisfait la relation suivante

$$g(t) = g(t + T) \quad (5.1)$$

pour tout  $t$  où  $g$  est définie et où  $T$  est une constante positive appelée une période de  $g$ .

Remarquons que si  $T$  est une période de  $g$ , alors  $2T, 3T, 4T, \dots$  sont également des périodes de  $g$ . La plus petite valeur de  $T$  satisfaisant la relation (5.1) est appelée **période fondamentale** de la fonction  $g$ . On définit alors la **fréquence fondamentale**  $f_0$  de la fonction  $g$  comme l'inverse de cette période fondamentale

$$f_0 = \frac{1}{T} \quad (5.2)$$

Avant de passer aux choses sérieuses, considérons la propriété suivante. Soit une fonction  $g(t)$  périodique de période  $T$  et continue par morceaux sur  $[t_0, t_0 + T]$ . Alors, pour tout  $t_0$ , nous avons

$$\int_{t_0}^{t_0+T} g(t) dt = \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) dt \quad (5.3)$$

En effet,

$$\int_{t_0}^{t_0+T} g(t) dt = \int_{t_0}^{-\frac{T}{2}} g(t) dt + \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) dt + \int_{+\frac{T}{2}}^{t_0+T} g(t) dt$$

Si l'on effectue le changement de variable  $t' = t - T$  dans la dernière intégrale, nous obtenons

$$\int_{t_0}^{t_0+T} g(t) dt = \int_{t_0}^{-\frac{T}{2}} g(t) dt + \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) dt + \int_{-\frac{T}{2}}^{t_0} g(t' + T) dt'$$

Etant donné que  $g(t' + T) = g(t')$ , la somme de la première et de la dernière intégrale est nulle, d'où le résultat annoncé.

Cette propriété va nous permettre, dans la suite, pour toute fonction périodique de période  $T$ , de nous limiter à son étude sur une seule de ses périodes, en particulier sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$ .

## 5.2 Polynômes de FOURIER (ou trigonométriques)

Soit une fonction  $g(t)$  périodique de période fondamentale  $T$  et continue par morceaux sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$ . Afin de faire le lien avec l'introduction, nous allons à présent tenter d'approximer cette fonction par la somme d'une onde fondamentale de fréquence  $f_0 = 1/T$  et d'un ensemble d'ondes "secondaires" de fréquence  $2f_0, 3f_0, \dots$ . Nous introduisons donc le polynôme suivant.

### 5.2.1 Définition

Les fonctions

$$\cos(2\pi n f_0 t) \text{ et } \sin(2\pi n f_0 t)$$

où  $n = 1, 2, 3, \dots$  sont périodiques et de périodes égales à  $\frac{1}{n f_0}$ . Elles sont donc également de période  $\frac{1}{f_0} = T$ . Toute combinaison linéaire de ces fonctions est donc également une fonction périodique de période  $T$ . Nous définissons donc le **polynôme trigonométrique** suivant

$$p(t) = a_0 + \sum_{n=1}^N (a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t))$$

(5.4)

où  $N$  est un entier supérieur ou égal à 1 et où  $a_0, a_n$  et  $b_n$  sont des constantes réelles. La constante  $a_0$  correspond à la valeur moyenne de  $p(t)$ <sup>1</sup>. Dans cette expression, on voit bien apparaître l'harmonique fondamentale (ou onde fondamentale) :

$$H_1(t) = a_1 \cos(2\pi f_0 t) + b_1 \sin(2\pi f_0 t) \quad (5.5)$$

mais également les harmoniques secondaires

$$H_n(t) = a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t) \quad (5.6)$$

dont les fréquences respectives sont bien des multiples entiers de la fréquence fondamentale  $f_0$ .

La question qui va donc se poser dans la suite est de savoir dans quelle mesure le polynôme trigonométrique (5.4) va pouvoir approximer la fonction  $g(t)$ . Il sera alors indispensable de déterminer les constantes  $a_0, a_n$  et  $b_n$  pour une fonction  $g(t)$  donnée mais également le nombre d'harmoniques  $N$  qu'il faudra retenir. Avant de répondre à ces questions, nous devons encore aborder les considérations mathématiques suivantes.

### 5.2.2 Intégrales intéressantes

Nous donnons ici la valeur de quelques intégrales dont nous allons avoir besoin dans la section suivante :

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} \cos(2\pi m f_0 t) \cos(2\pi n f_0 t) dt = \begin{cases} 0 & \text{si } m \neq n \\ \frac{T}{2} & \text{si } m = n \end{cases} \quad (5.7)$$

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi m f_0 t) \sin(2\pi n f_0 t) dt = \begin{cases} 0 & \text{si } m \neq n \\ \frac{T}{2} & \text{si } m = n \end{cases} \quad (5.8)$$

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi m f_0 t) \cos(2\pi n f_0 t) dt = 0 \quad (5.9)$$

1. Dans le cas des signaux, on parle encore de composante continue ou DC du signal  $p(t)$ .

où  $m$  et  $n$  sont des naturels non nuls. Le calcul de ces intégrales est laissé au bon soin du lecteur. Nous pouvons encore donner les deux intégrales suivantes :

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} \cos(2\pi m f_0 t) dt = \begin{cases} 0 & \text{si } m \neq 0 \\ T & \text{si } m = 0 \end{cases} \quad (5.10)$$

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi m f_0 t) dt = 0 \quad (5.11)$$

où  $m$  est un entier.

### 5.2.3 Calcul des coefficients $a_0$ , $a_n$ et $b_n$

Etant donné la définition (5.4) du polynôme trigonométrique, il est à présent possible d'exprimer les coefficients  $a_0$ ,  $a_n$  et  $b_n$  pour un polynôme  $p(t)$  donné. Les formules qui suivent nous permettront dans la suite de calculer ces mêmes coefficients pour la fonction  $g(t)$  que nous souhaitons approximer.

Commençons par le coefficient  $a_0$ . Pour cela, il suffit d'intégrer la relation (5.4) entre  $-\frac{T}{2}$  et  $+\frac{T}{2}$ . On obtient ainsi

$$\begin{aligned} \int_{-\frac{T}{2}}^{+\frac{T}{2}} p(t) dt &= \int_{-\frac{T}{2}}^{+\frac{T}{2}} a_0 dt + \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sum_{n=1}^N (a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)) dt \\ &= a_0 T + \sum_{n=1}^N a_n \int_{-\frac{T}{2}}^{+\frac{T}{2}} \cos(2\pi n f_0 t) dt + \sum_{n=1}^N b_n \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi n f_0 t) dt \end{aligned}$$

Etant données les deux relations (5.10) et (5.11), il vient finalement

$$a_0 = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} p(t) dt \quad (5.12)$$

Cette dernière expression montre bien que  $a_0$  représente la valeur moyenne de  $p(t)$  sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$ .

Pour calculer  $a_n$ , nous multiplions l'expression (5.4) par  $\cos(2\pi k f_0 t)$  ( $k = 1, \dots, N$ ) avant de l'intégrer entre  $-\frac{T}{2}$  et  $+\frac{T}{2}$ . Cela donne :

$$\begin{aligned} \int_{-\frac{T}{2}}^{+\frac{T}{2}} p(t) \cos(2\pi k f_0 t) dt &= \int_{-\frac{T}{2}}^{+\frac{T}{2}} a_0 \cos(2\pi k f_0 t) dt \\ &\quad + \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sum_{n=1}^N (a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)) \cos(2\pi k f_0 t) dt \\ &= a_0 \int_{-\frac{T}{2}}^{+\frac{T}{2}} \cos(2\pi k f_0 t) dt \\ &\quad + \sum_{n=1}^N a_n \int_{-\frac{T}{2}}^{+\frac{T}{2}} \cos(2\pi n f_0 t) \cos(2\pi k f_0 t) dt \\ &\quad + \sum_{n=1}^N b_n \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi n f_0 t) \cos(2\pi k f_0 t) dt \end{aligned}$$

En regard des intégrales fournies à la section 5.2.2, il vient que seule l'intégrale faisant intervenir les deux cosinus (pour  $n = k$ ) est non nulle. Il vient donc

$$a_k = \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} p(t) \cos(2\pi k f_0 t) dt \quad (5.13)$$

pour  $k = 1, \dots, N$ .

Le calcul des coefficients  $b_n$  est relativement similaire à celui des coefficients  $a_n$ . Il suffit de multiplier la relation (5.4) par  $\sin(2\pi k f_0 t)$  ( $k = 1, \dots, N$ ) avant de l'intégrer entre  $-\frac{T}{2}$  et  $+\frac{T}{2}$ . Il vient

$$\begin{aligned} \int_{-\frac{T}{2}}^{+\frac{T}{2}} p(t) \sin(2\pi k f_0 t) dt &= \int_{-\frac{T}{2}}^{+\frac{T}{2}} a_0 \sin(2\pi k f_0 t) dt \\ &\quad + \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sum_{n=1}^N (a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)) \sin(2\pi k f_0 t) dt \\ &= a_0 \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi k f_0 t) dt \\ &\quad + \sum_{n=1}^N a_n \int_{-\frac{T}{2}}^{+\frac{T}{2}} \cos(2\pi n f_0 t) \sin(2\pi k f_0 t) dt \\ &\quad + \sum_{n=1}^N b_n \int_{-\frac{T}{2}}^{+\frac{T}{2}} \sin(2\pi n f_0 t) \sin(2\pi k f_0 t) dt \end{aligned}$$

De la même manière que précédemment, il vient que seule l'intégrale faisant intervenir les deux sinus (pour  $n = k$ ) est non nulle. Nous obtenons donc

$$b_k = \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} p(t) \sin(2\pi k f_0 t) dt \quad (5.14)$$

pour  $k = 1, \dots, N$ .

Nous disposons à présent de tous les outils nécessaires pour mettre en place le lien entre les polynômes trigonométriques (5.4) et une fonction périodique  $g(t)$  continue par morceaux.

## 5.3 Théorème de FOURIER

Avant d'énoncer le théorème attendu, commençons par étudier l'exemple suivant.

### 5.3.1 Exemple : Fonction carrée

Considérons la fonction périodique  $g(t)$  de période  $T$ , continue par morceaux sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$  et définie par

$$g(t) = \begin{cases} -1 & \text{si } -\frac{T}{2} \leq t < 0 \\ +1 & \text{si } 0 \leq t < \frac{T}{2} \end{cases} \quad (5.15)$$

Celle-ci présente des discontinuités en  $0$ ,  $-\frac{T}{2}$  et  $+\frac{T}{2}$ .

Bien qu'elle ne soit pas un polynôme trigonométrique, calculons les coefficients  $a_0$ ,  $a_k$  et  $b_k$  ( $k = 1, \dots, N$ ) pour la fonction  $g(t)$ . Le coefficient  $a_0$  est donné par

$$\begin{aligned} a_0 &= \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) dt \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^0 (-1) dt + \frac{1}{T} \int_0^{+\frac{T}{2}} 1 dt \\ &= \frac{1}{T} (-t) \Big|_{-\frac{T}{2}}^0 + \frac{1}{T} t \Big|_0^{+\frac{T}{2}} \\ &= -\frac{1}{2} + \frac{1}{2} \\ &= 0 \end{aligned}$$

Venons-en au calcul de  $a_k$  pour  $k = 1, \dots, N$ . En partant de (5.13), nous avons

$$\begin{aligned} a_k &= \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \cos(2\pi k f_0 t) dt \\ &= \frac{2}{T} \int_{-\frac{T}{2}}^0 (-\cos(2\pi k f_0 t)) dt + \frac{2}{T} \int_0^{+\frac{T}{2}} \cos(2\pi k f_0 t) dt \end{aligned}$$

Dans la première intégrale, nous effectuons le changement de variable  $t' = -t$ . Il vient

$$\begin{aligned} a_k &= \frac{2}{T} \int_{\frac{T}{2}}^0 \cos(2\pi k f_0 t') dt' + \frac{2}{T} \int_0^{+\frac{T}{2}} \cos(2\pi k f_0 t) dt \\ &= 0 \end{aligned}$$

Calculons à présent  $b_k$  pour  $k = 1, \dots, N$ . L'application de (5.14) à  $g(t)$  fournit

$$\begin{aligned} b_k &= \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \sin(2\pi k f_0 t) dt \\ &= \frac{2}{T} \int_{-\frac{T}{2}}^0 (-\sin(2\pi k f_0 t)) dt + \frac{2}{T} \int_0^{+\frac{T}{2}} \sin(2\pi k f_0 t) dt \end{aligned}$$

Le changement de variable  $t' = -t$  dans la première intégrale conduit à

$$\begin{aligned} b_k &= \frac{2}{T} \int_{\frac{T}{2}}^0 \sin(-2\pi k f_0 t') dt' + \frac{2}{T} \int_0^{+\frac{T}{2}} \sin(2\pi k f_0 t) dt \\ &= \frac{4}{T} \int_0^{+\frac{T}{2}} \sin(2\pi k f_0 t) dt \\ &= \frac{4}{T} \left[ \frac{-\cos(2\pi k f_0 t)}{2\pi k f_0} \right]_0^{\frac{T}{2}} \\ &= \frac{2}{\pi k} \left( -\cos\left(2\pi k f_0 \frac{T}{2}\right) + 1 \right) \\ &= \frac{2(1 - \cos(k\pi))}{\pi k} \end{aligned}$$

Finalement, nous obtenons, si l'on tient compte de la parité de  $k$ ,

$$b_k = \begin{cases} 0 & \text{si } k \text{ est pair} \\ \frac{4}{\pi k} & \text{si } k \text{ est impair} \end{cases}$$

A partir des coefficients calculés, nous construisons à présent le polynôme trigonométrique suivant

$$p(t) = \sum_{n=1}^N \frac{2(1 - \cos(n\pi))}{\pi n} \sin(2\pi n f_0 t)$$

où on a tenu compte que  $a_0 = 0$  et que  $a_k = 0$  pour  $k = 1, \dots, N$ . Etudions à présent dans quelle mesure le polynôme  $p(t)$  ainsi construit se rapproche de  $g(t)$ . Pour cela, la table 5.1 montre  $g(t)$  et  $p(t)$  pour différentes valeurs de  $N$ .

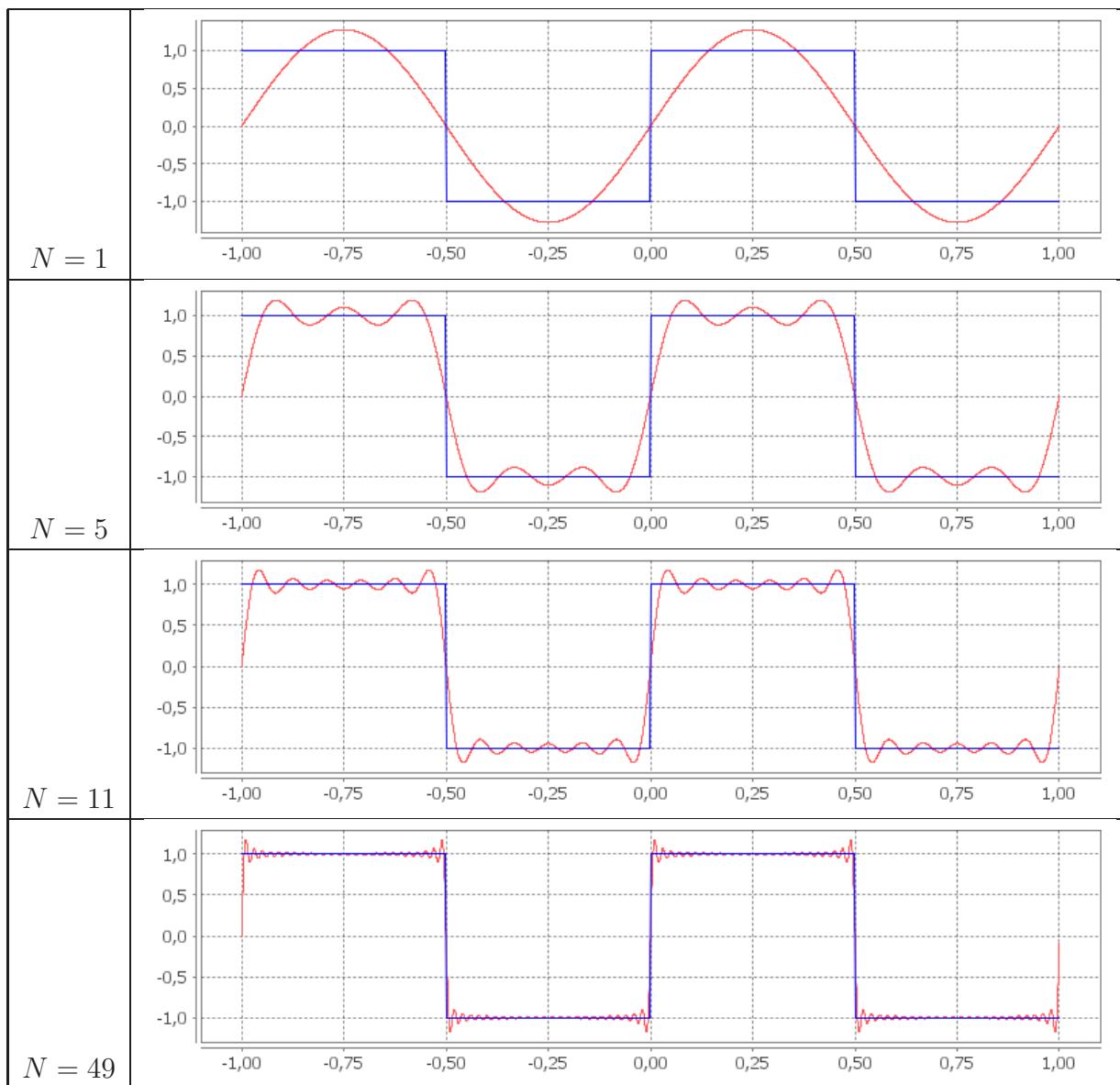


TABLE 5.1 – Polynôme trigonométrique construit à partir des coefficients de la fonction carrée (pour  $T = 1$ ).

A la table 5.1, pour  $N = 1$ , on observe clairement l'onde fondamentale. On peut également faire les observations suivantes :

- Plus le nombre d'harmoniques  $N$  augmente, plus la différence entre la fonction  $g(t)$  et le polynôme trigonométrique  $p(t)$  semble diminuer.

- Aux points de discontinuités, une différence importante subsiste. Cela est dû au fait que l'on tente d'approcher une fonction discontinue par une somme de fonctions continues.

Néanmoins, l'intuition nous fait penser que si l'on fait tendre  $N$  vers l'infini, la différence entre  $p(t)$  et  $g(t)$  tend vers 0. C'est la même intuition que FOURIER eut en son temps. En faisant tendre  $N$  vers l'infini, la somme de  $N$  termes, apparaissant dans (5.4), devient une série mathématique qui porte le nom de série de FOURIER.

### 5.3.2 Définition et énoncé du théorème

**Définition (Série de FOURIER).** Soit une fonction périodique  $g(t)$  de période  $T = \frac{1}{f_0}$  et continue par morceaux sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$ . La série de FOURIER de  $g(t)$ , notée  $SF_g(t)$ , est définie par

$$SF_g(t) = a_0 + \sum_{n=1}^{+\infty} (a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)) \quad (5.16)$$

où les coefficients  $a_0$ ,  $a_n$  et  $b_n$  sont donnés par

$$a_0 = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) dt \quad (5.17)$$

$$a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \cos(2\pi n f_0 t) dt \quad (5.18)$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \sin(2\pi n f_0 t) dt \quad (5.19)$$

Nous pouvons à présent énoncer le théorème de FOURIER donnant les conditions pour que la série (5.16) converge.

**Théorème (Théorème de FOURIER).** Si la fonction périodique  $g(t)$  de période  $T$  est continûment dérivable par morceaux sur  $[-\frac{T}{2}, +\frac{T}{2}]$ , alors

1. la série de FOURIER  $SF_g(t)$  converge en tout réel ;
2. en tout  $t \in ]-\frac{T}{2}, +\frac{T}{2}[$ , la série de FOURIER converge vers la moyenne de la limite à gauche et de la limite à droite en  $t$  de  $g$ , c'est-à-dire

$$SF_g(t) = \frac{1}{2} (g(t^-) + g(t^+))$$

en particulier, si  $f$  est continue en  $t$ , la série de FOURIER converge vers  $g(t)$ , c'est-à-dire

$$SF_g(t) = g(t)$$

3. Aux extrémités  $-\frac{T}{2}$  et  $+\frac{T}{2}$ , la série de FOURIER converge vers la moyenne de la limite à droite en  $-\frac{T}{2}$  et la limite à gauche en  $+\frac{T}{2}$ , c'est-à-dire

$$SF_g\left(-\frac{T}{2}\right) = SF_g\left(+\frac{T}{2}\right) = \frac{1}{2} \left(g\left(-\frac{T}{2}^+\right) + g\left(+\frac{T}{2}^-\right)\right)$$

Nous pouvons à présent revenir sur l'exemple de la fonction carrée (5.15) introduite plus haut. La fonction carrée  $g(t)$  respecte les hypothèses du théorème de FOURIER et nous pouvons

donc écrire

$$g(t) = \sum_{n=1}^{+\infty} \frac{2(1 - \cos(n\pi))}{\pi n} \sin(2\pi n f_0 t)$$

ou encore, en tenant compte de la parité de  $n$ ,

$$g(t) = \sum_{n=0}^{+\infty} \frac{4}{\pi(2n+1)} \sin(2\pi(2n+1)f_0 t)$$

Aux points de discontinuités, la série de FOURIER converge vers la moyenne de +1 et -1, c'est-à-dire 0. Ceci peut être observé à la table (5.1).

## 5.4 Mise en forme des séries de FOURIER

La manière (5.16) d'écrire la série de FOURIER avec une somme de sinus et de cosinus n'est pas son unique représentation. Comme nous allons le voir, elle peut également être représentée par une somme d'exponentielles, ce qui en simplifiera l'écriture.

### 5.4.1 Forme complexe ou exponentielle

Commençons par nous rappeler que les fonctions cosinus et sinus ont une représentation exponentielle complexe. Ainsi,

$$\cos(2\pi n f_0 t) = \frac{e^{j2\pi n f_0 t} + e^{-j2\pi n f_0 t}}{2} \quad \text{et} \quad \sin(2\pi n f_0 t) = \frac{e^{j2\pi n f_0 t} - e^{-j2\pi n f_0 t}}{2j}$$

Les harmoniques  $H_n(t)$  (voir (5.5) et (5.6)) peuvent dès lors s'écrire

$$\begin{aligned} H_n(t) &= a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t) \\ &= a_n \frac{e^{j2\pi n f_0 t} + e^{-j2\pi n f_0 t}}{2} + b_n \frac{e^{j2\pi n f_0 t} - e^{-j2\pi n f_0 t}}{2j} \\ &= \frac{a_n}{2} (e^{j2\pi n f_0 t} + e^{-j2\pi n f_0 t}) - \frac{j b_n}{2} (e^{j2\pi n f_0 t} - e^{-j2\pi n f_0 t}) \\ &= \frac{a_n - j b_n}{2} e^{j2\pi n f_0 t} + \frac{a_n + j b_n}{2} e^{-j2\pi n f_0 t} \end{aligned}$$

On pose alors

$$c_n = \frac{1}{2} (a_n - j b_n) \tag{5.20}$$

pour  $n$  entier différent de 0. Remarquons que, si on remplace  $n$  par  $-n$  dans cette dernière expression, nous obtenons

$$c_{-n} = \frac{1}{2} (a_{-n} - j b_{-n}) = \frac{1}{2} (a_n + j b_n) = \overline{c_n}$$

étant donné que  $a_{-n} = a_n$  et que  $b_{-n} = -b_n$  (voir (5.18) et (5.19)). Nous pouvons donc réécrire les harmoniques sous la forme

$$H_n(t) = c_n e^{j2\pi n f_0 t} + c_{-n} e^{-j2\pi n f_0 t} \tag{5.21}$$

Si on pose en plus

$$c_0 = a_0 \tag{5.22}$$

la série de FOURIER (5.16) peut à présent s'écrire

$$\begin{aligned} SF_g(t) &= c_0 + \sum_{n=1}^{+\infty} (c_n e^{j2\pi n f_0 t} + c_{-n} e^{-j2\pi n f_0 t}) \\ &= c_0 + \sum_{n=1}^{+\infty} c_n e^{j2\pi n f_0 t} + \sum_{n=1}^{+\infty} c_{-n} e^{-j2\pi n f_0 t} \\ &= c_0 + \sum_{n=1}^{+\infty} c_n e^{j2\pi n f_0 t} + \sum_{n'=-\infty}^{-1} c_{n'} e^{j2\pi n' f_0 t} \end{aligned}$$

où, dans la dernière somme, nous avons posé  $n = -n'$ . Finalement, en rassemblant les trois termes, nous obtenons

$$SF_g(t) = \sum_{n=-\infty}^{+\infty} c_n e^{j2\pi n f_0 t}$$

(5.23)

Cette dernière formule constitue la forme complexe ou exponentielle de la série de FOURIER de  $g(t)$ . Nous allons à présent voir que les coefficient  $c_n$  peuvent également avoir une représentation complexe.

### Calcul des coefficients $c_n$

En repartant de l'expression (5.20) de  $c_n$  et des définitions (5.18) et (5.19) de  $a_n$  et  $b_n$ , nous pouvons écrire

$$\begin{aligned} c_n &= \frac{1}{2} (a_n - jb_n) \\ &= \frac{1}{2} \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \cos(2\pi n f_0 t) dt - \frac{j}{2} \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \sin(2\pi n f_0 t) dt \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) (\cos(2\pi n f_0 t) - j \sin(2\pi n f_0 t)) dt \end{aligned}$$

et finalement, en utilisant la relation d'EULER, nous obtenons

$$c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) e^{-j2\pi n f_0 t} dt$$

(5.24)

Remarquons que cette dernière expression est valable tout entier  $n$ , y compris  $n = 0$ . En effet, en remplaçant  $n$  par 0 dans cette dernière expression, nous retrouvons bien  $c_0 = a_0$ .

### Remarque

Si une série converge, son terme général tend vers 0 lorsque l'indice  $n$  tend vers l'infini. Dès lors, pour  $n$  qui tend vers l'infini, nous avons  $\|c_n e^{j2\pi n f_0 t}\| \rightarrow 0$  et donc  $\|c_n\| \rightarrow 0$ . Nous pouvons donc écrire finalement

$$\lim_{n \rightarrow \pm\infty} \|c_n\| = 0$$

mais encore

$$\lim_{n \rightarrow \pm\infty} a_n = \lim_{n \rightarrow \pm\infty} b_n = 0$$

### 5.4.2 Forme trigonométrique

Une autre représentation intéressante de la série de FOURIER est de voir la fonction  $g(t)$  comme une somme de sinusoïdes déphasées dont l'amplitude et la phase dépendent de  $g(t)$ . Pour cela, commençons par mettre les coefficients complexes  $c_n$  sous leur forme polaire :

$$c_n = r_n e^{j\theta_n}$$

où nous avons les relations suivantes entre  $(a_n, b_n)$  et  $(r_n, \theta_n)$  :

$$\begin{cases} r_n &= \frac{1}{2} \sqrt{a_n^2 + b_n^2} \\ \tan \theta_n &= -\frac{b_n}{a_n} \end{cases}$$

Les harmoniques  $H_n(t)$  provenant de (5.21), pour  $n = 1, \dots$  peuvent alors s'écrire

$$\begin{aligned} H_n(t) &= c_n e^{j2\pi n f_0 t} + c_{-n} e^{-j2\pi n f_0 t} \\ &= c_n e^{j2\pi n f_0 t} + \overline{c_n} e^{-j2\pi n f_0 t} \\ &= r_n e^{j\theta_n} e^{j2\pi n f_0 t} + r_n e^{-j\theta_n} e^{-j2\pi n f_0 t} \\ &= r_n e^{j(2\pi n f_0 t + \theta_n)} + r_n e^{-j(2\pi n f_0 t + \theta_n)} \\ &= 2r_n \cos(2\pi n f_0 t + \theta_n) \end{aligned}$$

La série de FOURIER peut finalement s'écrire sous la forme

$$SF_g(t) = a_0 + 2 \sum_{n=-\infty}^{+\infty} r_n \cos(2\pi n f_0 t + \theta_n)$$

(5.25)

qui constitue la représentation trigonométrique de la série de FOURIER de la fonction  $g(t)$ .

### 5.4.3 Cas particuliers : fonctions paires, impaires et demi-onde

Dans le cas de certaines fonctions, nous allons voir que l'expression des coefficients  $a_n$  et  $b_n$  se simplifie, voir même s'annule complètement.

#### Fonction paire : $g(-t) = g(t)$

Dans ce cas, en partant de (5.18), les coefficients  $a_n$  peuvent s'écrire

$$\begin{aligned} a_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \cos(2\pi n f_0 t) dt \\ &= \frac{4}{T} \int_0^{+\frac{T}{2}} g(t) \cos(2\pi n f_0 t) dt \end{aligned}$$

étant donné que l'intégrant est une fonction paire, obtenue par le produit de deux fonctions paires.

Les coefficients  $b_n$  (voir (5.19)) sont, quant à eux, égaux à

$$\begin{aligned} b_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \sin(2\pi n f_0 t) dt \\ &= 0 \end{aligned}$$

vu que l'intégrant est une fonction impaire, obtenue par le produit d'une fonction paire et d'une fonction impaire.

On remarque dès lors que les coefficients  $c_n$  sont purement réels et valent  $c_n = \frac{1}{2} a_n$ . Du coup, la série de FOURIER d'une fonction  $g(t)$  paire ne comporte que des termes en cosinus.

**Fonction impaire :**  $g(-t) = -g(t)$

Dans ce cas, en partant de (5.18), les coefficients  $a_n$  peuvent s'écrire

$$\begin{aligned} a_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \cos(2\pi n f_0 t) dt \\ &= 0 \end{aligned}$$

étant donné que l'intégrant est une fonction impaire, obtenue par le produit d'une fonction impaire et d'une fonction paire.

Les coefficients  $b_n$  (voir (5.19)) sont, quant à eux, égaux à

$$\begin{aligned} b_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g(t) \sin(2\pi n f_0 t) dt \\ &= \frac{4}{T} \int_0^{+\frac{T}{2}} g(t) \sin(2\pi n f_0 t) dt \end{aligned}$$

vu que l'intégrant est une fonction paire, obtenue par le produit de deux fonctions impaires.

On remarque dès lors que les coefficients  $c_n$  sont purement imaginaires et valent  $c_n = \frac{i}{2} b_n$ . Du coup, la série de FOURIER d'une fonction  $g(t)$  impaire ne comporte que des termes en sinus.

**Fonction demi-onde :**  $g(t - \frac{T}{2}) = -g(t)$

Pour ces fonctions, si  $n$  est pair, les fonctions  $\cos(2\pi n f_0 t)$  et  $\sin(2\pi n f_0 t)$  sont de période  $\frac{T}{2}$  et nous avons

$$\int_{-\frac{T}{2}}^0 g(t) \cos(2\pi n f_0 t) dt = - \int_0^{+\frac{T}{2}} g(t) \cos(2\pi n f_0 t) dt$$

Dès lors,  $a_n = 0$  et de même  $b_n = 0$ . Par conséquent, tous les coefficients de FOURIER d'indice pair, y compris  $c_0 = a_0$ , sont nuls. La fonction carrée (voir (5.15)) est un exemple de fonction demi-onde.

## 5.5 Formule de PARSEVAL

Dans le domaine de l'analyse des signaux, il est courant de différencier les signaux de puissance des signaux d'énergie. On pourrait montrer que les signaux périodiques ont une énergie infinie et une puissance moyenne finie, ils font partie de la catégorie des signaux de puissance. La formule de PARSEVAL va nous permettre de calculer la puissance moyenne d'une fonction périodique connaissant les coefficients de sa série de FOURIER.

Commençons par donner la définition de la puissance moyenne  $P$  d'une fonction périodique  $g(t)$  de période  $T$  :

$$P = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g^2(t) dt \tag{5.26}$$

Nous avons le résultat important suivant.

**Théorème (Formule de PARSEVAL).** Si la fonction périodique  $g(t)$  de période  $T$  est continûment dérivable par morceaux sur  $[-\frac{T}{2}, +\frac{T}{2}]$ , alors

$$\frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g^2(t) dt = \sum_{n=-\infty}^{+\infty} \|c_n\|^2 = a_0^2 + \frac{1}{2} \sum_{n=1}^{+\infty} (a_n^2 + b_n^2) \quad (5.27)$$

La preuve de cette formule est relativement aisée. En effet,

$$\begin{aligned} \int_{-\frac{T}{2}}^{+\frac{T}{2}} g^2(t) dt &= \int_{-\frac{T}{2}}^{+\frac{T}{2}} \left( \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} c_n c_m e^{j2\pi n f_0 t} e^{j2\pi m f_0 t} \right) dt \\ &= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} c_n c_m \int_{-\frac{T}{2}}^{+\frac{T}{2}} e^{j2\pi(n+m)f_0 t} dt \end{aligned}$$

Mais étant données les intégrales (5.10) et (5.11), nous savons grâce à la formule d'EULER que

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} e^{j2\pi(n+m)f_0 t} dt = \begin{cases} 0 & \text{si } n+m \neq 0 \\ T & \text{si } n+m = 0 \end{cases}$$

Dès lors,

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} g^2(t) dt = \sum_{n=-\infty}^{+\infty} c_n c_{-n} T = T \sum_{n=-\infty}^{+\infty} \|c_n\|^2$$

vu que  $c_{-n} = \overline{c_n}$ .

## 5.6 Exemple : Droite de FOURIER

Un premier exemple a déjà été donné plus haut lors de l'étude de la fonction carrée. Pour clôturer ce chapitre, nous allons étudier un second exemple, celui de la fonction "dents de scie". Celle-ci, sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$ , se comporte comme la droite d'équation  $g(t) = t$ . On en parle donc souvent en utilisant l'appellation de "droite de FOURIER".

Soit la fonction périodique  $g(t)$  de période  $T$ , définie sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$  par

$$g(t) = \frac{2}{T} t \quad (5.28)$$

Cette fonction respecte les hypothèses du théorème de FOURIER, sa série de FOURIER existe donc. Calculons ses coefficients. Etant donné que la fonction  $g(t)$  est impaire, les coefficients  $a_0$  et  $a_n$  sont nuls. Les coefficients  $b_n$  sont alors obtenus par

$$\begin{aligned} b_n &= \frac{4}{T} \int_0^{+\frac{T}{2}} \frac{2}{T} t \sin(2\pi n f_0 t) dt \\ &= \frac{8}{T^2} \int_0^{+\frac{T}{2}} t \sin(2\pi n f_0 t) dt \end{aligned}$$

En utilisant la technique de l'intégration par parties, il vient

$$\begin{aligned} b_n &= \frac{8}{T^2} \left\{ \left[ -\frac{t}{2\pi n f_0} \cos(2\pi n f_0 t) \right]_0^{\frac{T}{2}} + \frac{1}{2\pi n f_0} \int_0^{+\frac{T}{2}} \cos(2\pi n f_0 t) dt \right\} \\ &= \frac{8}{T^2} \frac{1}{2\pi n f_0} \left[ -\frac{T}{2} \cos\left(2\pi n f_0 \frac{T}{2}\right) + \frac{1}{2\pi n f_0} \sin\left(2\pi n f_0 \frac{T}{2}\right) \right] \\ &= -\frac{2}{n\pi} \cos(n\pi) \end{aligned}$$

Finalement, en tenant compte de la parité de  $n$ , nous obtenons

$$b_n = \begin{cases} -\frac{2}{n\pi} & \text{si } n \text{ est pair} \\ +\frac{2}{n\pi} & \text{si } n \text{ est impair} \end{cases} = -\frac{2(-1)^n}{n\pi}$$

La série de FOURIER de  $g(t)$  peut enfin s'écrire

$$SF_g(t) = -\frac{2}{\pi} \sum_{n=1}^{+\infty} \frac{(-1)^n}{n} \sin(2\pi n f_0 t) \quad (5.29)$$

Les polynômes de FOURIER pour différentes valeurs de  $N$  sont présentés à la table 5.2. En choisissant  $T = 2$ , nous pouvons finalement déduire que sur l'intervalle  $] -1, +1 [$ , nous avons

$$t = \frac{2}{\pi} \left( \sin(\pi t) - \frac{1}{2} \sin(2\pi t) + \frac{1}{3} \sin(3\pi t) - \dots \right) \quad (5.30)$$

### 5.6.1 Phénomène de GIBBS

Nous avons vu lors de l'analyse de la fonction carrée, et maintenant de la fonction dents de scie, que plus on additionne des termes trigonométriques (en augmentant  $N$ ), plus la fonction obtenue tend vers la fonction carrée (ou dents de scie). Mais nous remarquons aussi que sur les points anguleux de la fonction (aux discontinuités), bien qu'on puisse ajouter un grand nombre de termes, un défaut d'approximation persistera.

Etant donné que tout terme de la série de FOURIER est une fonction continue, il est normal que la série ne puisse approcher uniformément la fonction aux points de discontinuité. Des études ont déjà montré que l'amplitude d'oscillation autour des discontinuités est toujours plus importante. Ce phénomène porte le nom de phénomène de GIBBS. Quantitativement, il en ressort que très régulièrement l'oscillation y est de 18% plus importante que sur le reste de la fonction. C'est le cas de la fonction carrée mais aussi de la fonction dents de scie.

## 5.7 Exercices

- Déterminer le développement en série de FOURIER de la fonction périodique  $g(t)$  de période  $T$  et définie sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$  par

$$g(t) = \begin{cases} 0 & \text{si } -\frac{T}{2} \leq t < 0 \\ \frac{2}{T}t & \text{si } 0 \leq t < \frac{T}{2} \end{cases}$$

- Déterminer le développement en série de FOURIER de la fonction périodique  $g(t)$  de période  $2\pi$  et définie sur l'intervalle  $[-\pi, +\pi[$  par  $g(t) = t^2$ .
- Déterminer le développement en série de FOURIER de la fonction triangle, c'est-à-dire de la fonction périodique  $g(t)$  de période  $T$  et définie sur l'intervalle  $[-\frac{T}{2}, +\frac{T}{2}]$  par

$$g(t) = \begin{cases} -\frac{4t}{T} - 1 & \text{si } -\frac{T}{2} \leq t < 0 \\ \frac{4t}{T} - 1 & \text{si } 0 \leq t < \frac{T}{2} \end{cases}$$

## 5.8 Références

1. Mathématiques appliquées. A. PETRY. 2013. Haute Ecole de la Province de Liège.
2. Séries de FOURIER et applications. A. ETESSE, F. JACOBS, J.-F. ZHANG. 2011.

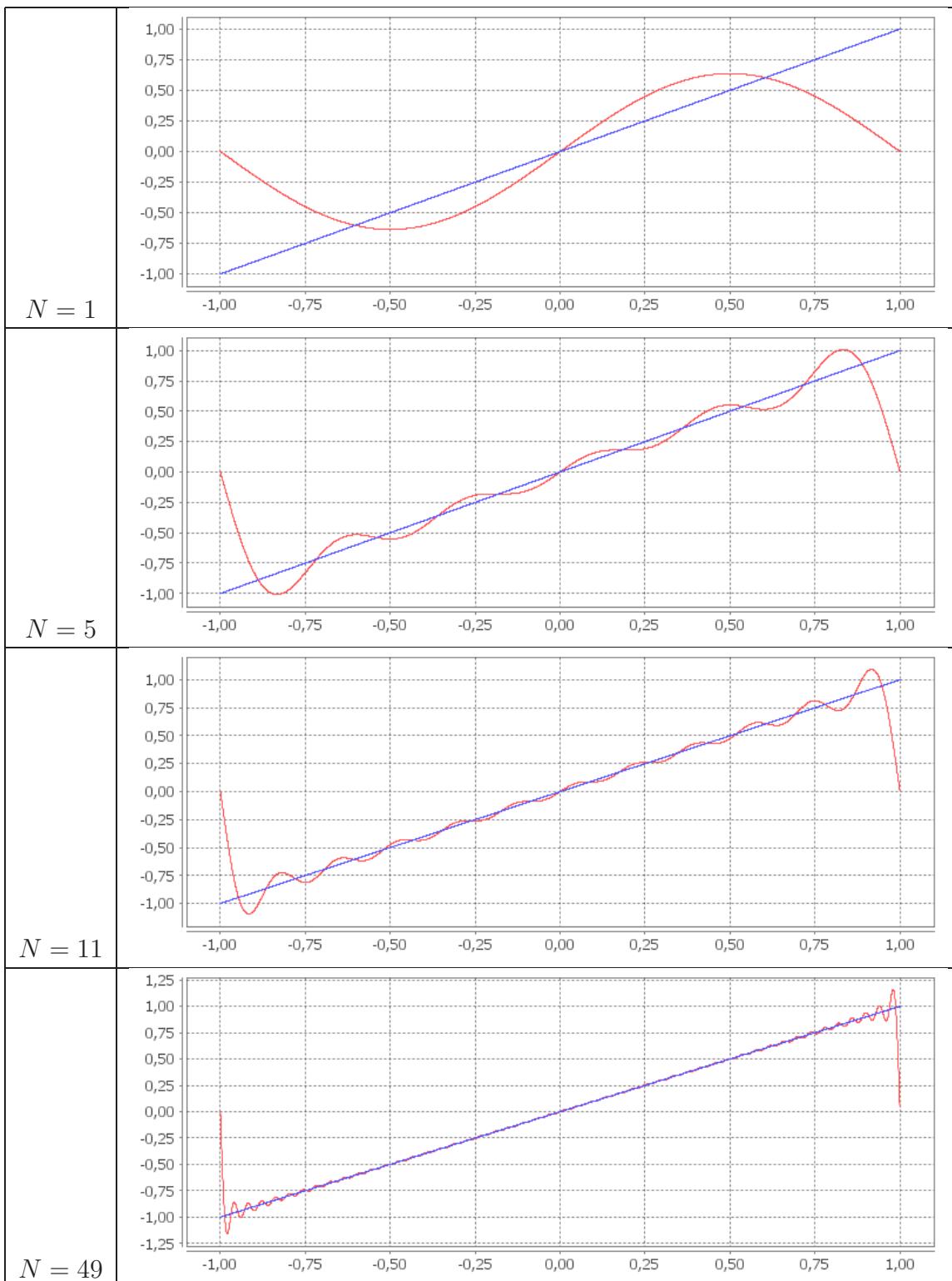


TABLE 5.2 – Polynôme trigonométrique construit à partir des coefficients de la fonction “dents de scie” (pour  $T = 2$ ).

# Chapitre 6

## Transformée de FOURIER 1D

### 6.1 Transformée de FOURIER 1D

#### 6.1.1 Définitions

Soit  $g(t)$  un signal déterministe.

**Définition [Transformée de FOURIER].** La transformée de FOURIER du signal  $g(t)$  est définie par

$$G(f) = \int_{-\infty}^{+\infty} g(t) e^{-j2\pi t f} dt \quad (6.1)$$

À partir de  $G(f)$ , il est possible de retrouver exactement le signal  $g(t)$  au moyen de

**Définition [Transformée de FOURIER inverse].** La transformée de FOURIER inverse de  $G(f)$  est définie par

$$g(t) = \int_{-\infty}^{+\infty} G(f) e^{j2\pi t f} df \quad (6.2)$$

Pour que la transformée de FOURIER d'un signal  $g(t)$  existe, il faut que l'intégrale (7.1) converge et fournit un résultat fini quelque soit la fréquence  $f$ . Mathématiquement, il est possible d'introduire un ensemble de contraintes sur la fonction  $g(t)$ . Dans le cadre de ce cours,  $g(t)$  est un signal et on pourrait montrer que pour les signaux physiques, c'est-à-dire les signaux d'énergie, la transformée de FOURIER existe toujours.

Les signaux de puissance, ayant une énergie infinie, ne possède pas de transformée de FOURIER au sens classique des mathématiques. En particulier, il est impossible de calculer la transformée de FOURIER des signaux périodiques, comme une sinusoïde. En effet, dans ce cas, l'intégrale (7.1) donne un résultat infini quelque soit  $f$ ... Cela est particulièrement gênant pour l'étude de nombreux systèmes de traitement ou de transmission du signal qui font abondamment usage de ce genre de signal. Néanmoins, la théorie des distributions introduit un signal particulier, l'impulsion de DIRAC, qui permet de résoudre ce problème. Nous y reviendrons très bientôt.

Dans la suite, nous dirons que  $g(t)$  et  $G(f)$  forment une paire de transformées de FOURIER représentée par

$$g(t) \rightleftharpoons G(f)$$

En général,  $G(f)$  est une fonction à valeurs complexes, ce qui n'est pas spécialement pour ravir le lecteur... Pour rappel, un nombre complexe peut s'exprimer en utilisant la notation

module-argument. Nous pouvons donc écrire

$$G(f) = \|G(f)\| e^{j\theta(f)} \quad (6.3)$$

où

- $\|G(f)\|$  est appelé *module* de  $G(f)$ , ou encore *spectre* de  $g(t)$ , et
- $\theta(f)$  est appelée phase de  $G(f)$ .

Dans le cas important où  $g(t)$  est un **signal à valeurs réelles**, nous avons

$$G^*(f) = G(-f)$$

où \* représente le complexe conjugué. Il vient

$$\boxed{\|G(-f)\| = \|G(f)\|}$$

$$\boxed{\theta(-f) = -\theta(f)}$$

Dès lors, nous pouvons déduire deux propriétés importantes d'un signal à valeurs réelles :

- Le *spectre* du signal est une *fonction paire* de la fréquence, c'est-à-dire que le graphe de  $\|G(f)\|$  est symétrique par rapport à l'axe vertical.
- La *phase* du signal est une *fonction impaire* de la fréquence, c'est-à-dire que le graphe de  $\theta(f)$  est symétrique par rapport à l'origine des axes.

Ces deux propriétés sont très importantes dans le cas du filtrage linéaire des signaux. En effet, si un traitement dans le domaine fréquentiel modifie le spectre et/ou la phase du signal de telle sorte qu'une, au moins, de ces deux propriétés ne soit plus vérifiée, le signal obtenu après transformée de FOURIER inverse ne sera plus réel...

### 6.1.2 Propriétés

#### 1. Linéarité

Soient  $g_1(t) \rightleftharpoons G_1(f)$  et  $g_2(t) \rightleftharpoons G_2(f)$ . Alors, pour toutes constantes  $c_1$  et  $c_2$ , nous avons

$$c_1 g_1(t) + c_2 g_2(t) \rightleftharpoons c_1 G_1(f) + c_2 G_2(f) \quad (6.4)$$

#### 2. Dilatation temporelle

Soit  $g(t) \rightleftharpoons G(f)$ . Nous avons

$$g(at) \rightleftharpoons \frac{1}{|a|} G\left(\frac{f}{a}\right) \quad (6.5)$$

La fonction  $g(at)$  représente une version de  $g(t)$  compressée dans le temps par un facteur  $a$  tandis que la fonction  $G(f/a)$  représente une version de  $G(f)$  dilatée en fréquence par le même facteur  $a$ . Dès lors, il vient qu'une compression dans le domaine temporel équivaut à une dilatation dans le domaine fréquentiel et vice versa.

#### 3. Dualité

Si  $g(t) \rightleftharpoons G(f)$ , alors

$$G(t) \rightleftharpoons g(-f) \quad (6.6)$$

## 4. Translation temporelle

Si  $g(t) \rightleftharpoons G(f)$ , alors

$$g(t - t_0) \rightleftharpoons G(f) e^{-j2\pi f t_0} \quad (6.7)$$

Il en résulte que le fait de translater la fonction  $g(t)$  de  $t_0$  ne modifie pas le module de la transformée de FOURIER, par contre sa phase est modifiée d'un facteur linéaire  $-j2\pi f t_0$ .

## 5. Translation fréquentielle

Si  $g(t) \rightleftharpoons G(f)$ , alors

$$g(t) e^{j2\pi f_0 t} \rightleftharpoons G(f - f_0) \quad (6.8)$$

La multiplication de la fonction  $g(t)$  par le facteur  $e^{j2\pi f_0 t}$  est équivalente à une translation de la transformée de FOURIER  $G(f)$  dans le domaine fréquentiel. On appelle encore cette propriété théorème de *modulation*.

## 6. Dérivée temporelle

Si  $g(t) \rightleftharpoons G(f)$ , alors

$$\frac{d}{dt} g(t) \rightleftharpoons j2\pi f G(f) \quad (6.9)$$

## 7. Intégration temporelle

Si  $g(t) \rightleftharpoons G(f)$  et que  $G(0) = 0$ , ce qui signifie que le signal  $g(t)$  n'a pas de composante continue ou encore que le signal  $g(t)$  oscille en moyenne autour de la valeur 0, alors

$$\int_{-\infty}^t g(x) dx = \int g(t) dt \rightleftharpoons \frac{1}{j2\pi f} G(f) \quad (6.10)$$

## 8. Multiplication dans le domaine temporel

Si  $g_1(t) \rightleftharpoons G_1(f)$  et  $g_2(t) \rightleftharpoons G_2(f)$ , alors

$$g_1(t) g_2(t) \rightleftharpoons \int_{-\infty}^{+\infty} G_1(\lambda) G_2(f - \lambda) d\lambda = (G_1 \otimes G_2)(f) \quad (6.11)$$

L'intégrale apparaissant dans cette expression est connue sous le nom d'*intégrale de convolution* dans le domaine fréquentiel. L'opérateur  $\otimes$  est appelé opérateur de convolution et la nouvelle fonction  $(G_1 \otimes G_2)$  ainsi créée est appelé *produit de convolution* de  $G_1$  et  $G_2$ . Dès lors, une multiplication dans le domaine temporel est équivalente à une convolution dans le domaine fréquentiel.

## 9. Convolution dans le domaine temporel

Si  $g_1(t) \rightleftharpoons G_1(f)$  et  $g_2(t) \rightleftharpoons G_2(f)$ , alors

$$(g_1 \otimes g_2)(t) = \int_{-\infty}^{+\infty} g_1(\tau) g_2(t - \tau) d\tau \rightleftharpoons G_1(f) G_2(f) \quad (6.12)$$

Dès lors, le produit de convolution dans le domaine temporel équivaut à un simple produit dans le domaine fréquentiel. Cette propriété est essentielle dans l'étude des systèmes linéaires, elle porte le nom de *théorème de convolution*.

### 10. Égalité de PARSEVAL

Si  $g(t) \rightleftharpoons G(f)$ , alors

$$\int_{-\infty}^{+\infty} g^2(t) dt = \int_{-\infty}^{+\infty} \|G(f)\|^2 df \quad (6.13)$$

Cette propriété porte également le nom de *théorème d'énergie de RAYLEIGH*. La quantité  $\|G(f)\|^2$  est appelée *densité spectrale d'énergie* du signal  $g(t)$ .

### 6.1.3 Exemples

#### Exemple 1

Considérons une impulsion rectangulaire  $g(t)$  de durée  $T$ , d'amplitude  $A$  et centrée à l'origine. Afin de définir ce signal, nous introduisons la fonction

$$\text{rect}(t) = \begin{cases} 1 & \text{si } -\frac{1}{2} < t < \frac{1}{2} \\ 0 & \text{sinon} \end{cases}$$

appelée fonction rectangle. Dès lors, nous pouvons écrire

$$g(t) = A \text{rect}\left(\frac{t}{T}\right)$$

La transformée de FOURIER du signal  $g(t)$  est donnée par

$$\begin{aligned} G(f) &= \int_{-T/2}^{T/2} A e^{-j2\pi f t} dt \\ &= AT \left( \frac{\sin(\pi f T)}{\pi f T} \right) \end{aligned}$$

Afin de simplifier la notation précédente, nous introduisons la fonction sinus cardinal, *sinc*, définie par

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \quad (6.14)$$

Cette fonction est représentée à la figure 6.1.

Finalement, il vient

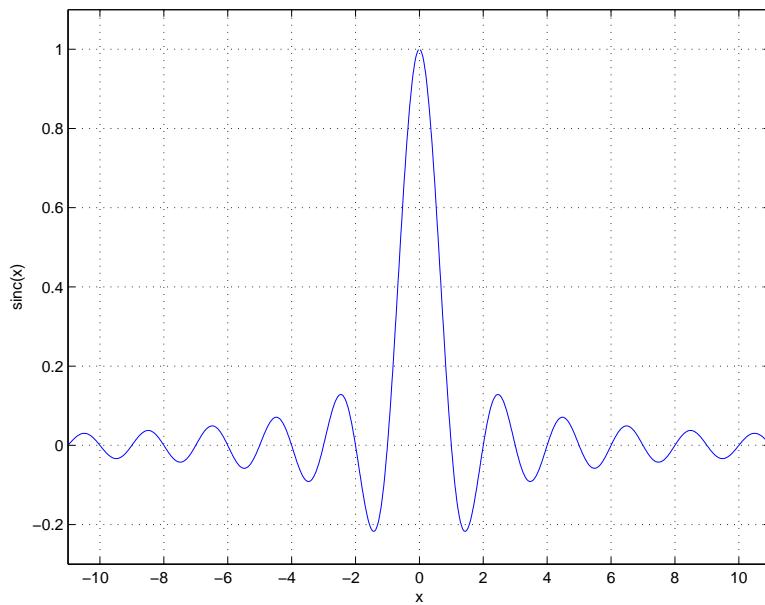
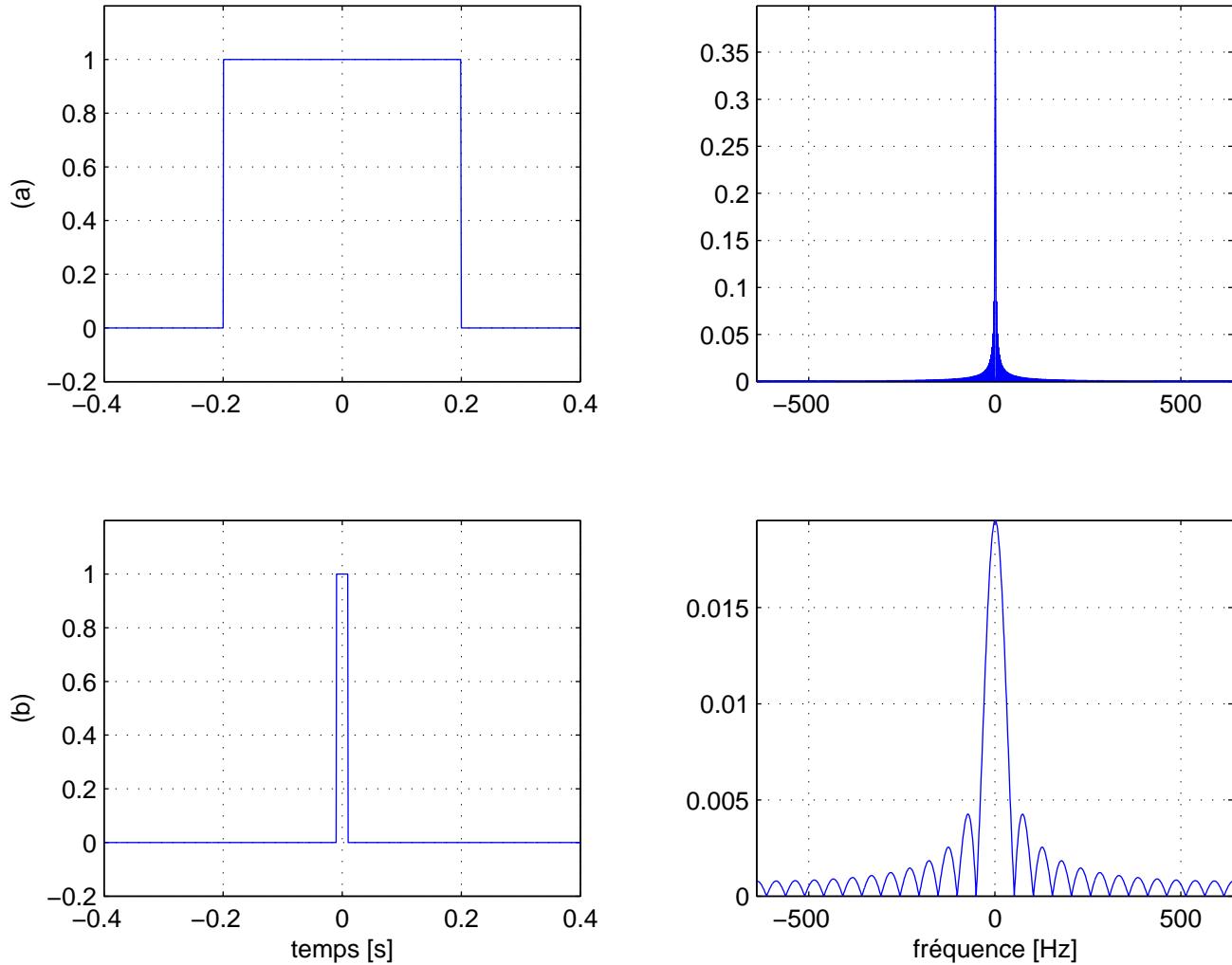
$$G(f) = AT \text{sinc}(fT)$$

Nous avons donc la paire de transformée de FOURIER suivante :

$$A \text{rect}\left(\frac{t}{T}\right) \rightleftharpoons AT \text{sinc}(fT) \quad (6.15)$$

Ce signal, ainsi que son spectre, sont représentés à la figure 6.2 pour  $T = 0,4 [s]$  et  $T = 0,02 [s]$  ( $A = 1$  dans les deux cas).

Dans les deux exemples de la figure 6.2, la propriété de dilatation de la transformée de FOURIER est bien illustrée. Pour passer de (a) à (b), la durée de l'impulsion a été réduite d'un facteur 20, on observe donc une contraction dans le domaine temporel. Par contre, dans le domaine fréquentiel, on observe bien une dilatation selon l'axe des fréquences. Nous constatons donc ici un phénomène important : plus un signal est bref, plus il s'étend sur l'axe des fréquences. On dit qu'il présente une grande *bande passante*.

FIGURE 6.1 – La fonction  $\text{sinc}(x)$ .FIGURE 6.2 – L'impulsion rectangulaire et son spectre : (a)  $A = 1$  et  $T = 0,4 [s]$ . (b)  $A = 1$  et  $T = 0,02 [s]$ .

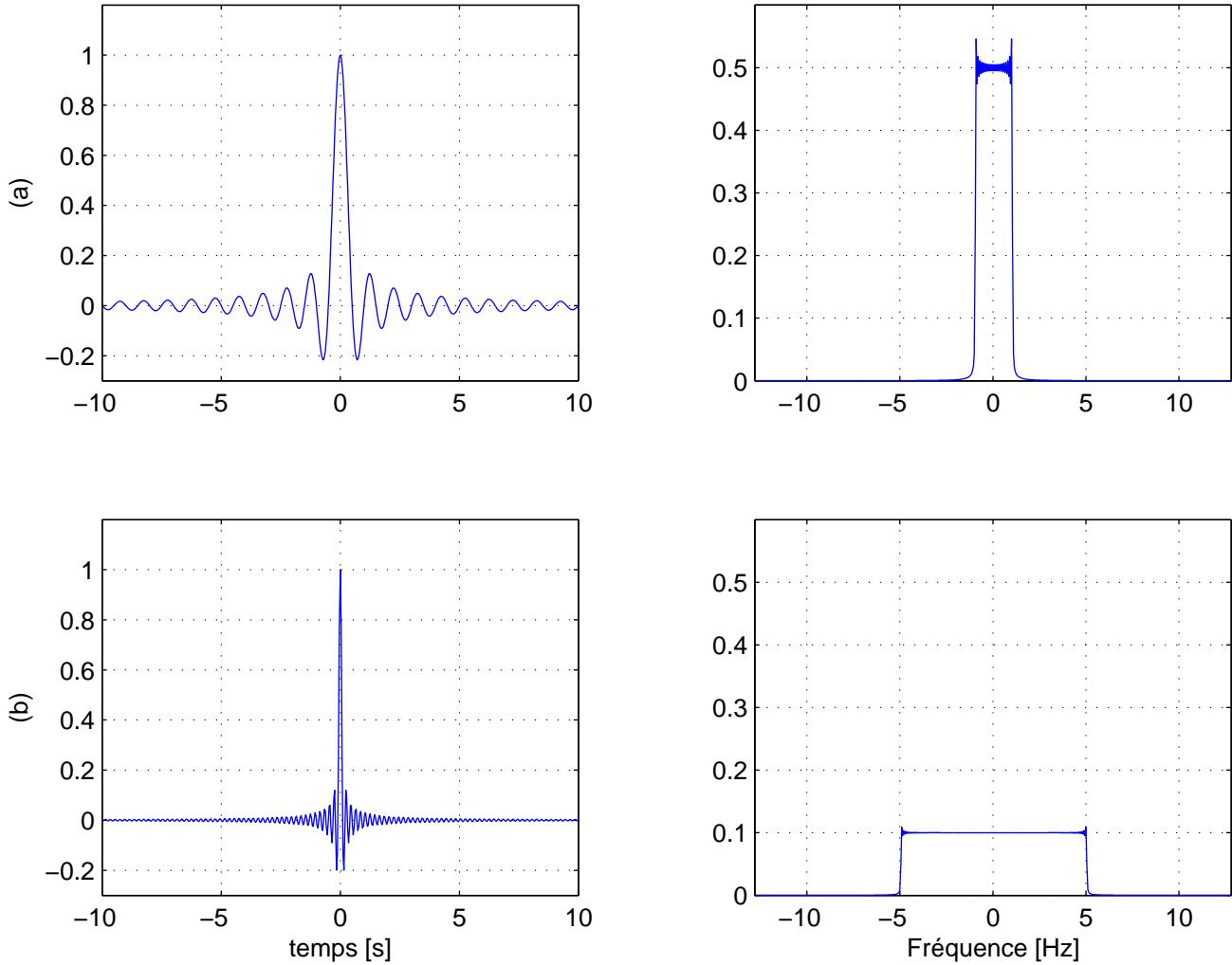


FIGURE 6.3 – Le signal  $g(t) = A \text{sinc}(2Wt)$  et son spectre : (a)  $W = 1 \text{ [Hz]}$  et  $A = 1$ . (b)  $W = 5 \text{ [Hz]}$  et  $A = 1$ .

### Exemple 2

Considérons le signal suivant

$$g(t) = A \text{sinc}(2Wt)$$

En appliquant la propriété de dualité de la transformée de FOURIER à la relation (6.15), et étant donné que la fonction rectangle est une fonction paire, il vient

$$AT \text{sinc}(tT) \rightleftharpoons A \text{rect}\left(\frac{f}{T}\right)$$

En posant  $T = 2W$ , nous obtenons la paire de transformée de FOURIER suivante

$$A \text{sinc}(2Wt) \rightleftharpoons \frac{A}{2W} \text{rect}\left(\frac{f}{2W}\right) \quad (6.16)$$

La figure 6.3 illustre le signal  $g(t)$ , ainsi que son spectre, pour  $W = 1 \text{ [Hz]}$  et  $W = 5 \text{ [Hz]}$  ( $A = 1$  dans les deux cas).

On observe que le spectre du signal  $g(t)$  est nul pour des fréquences supérieures à  $+W$  et inférieures à  $-W$ . Tout signal présentant cette particularité est dit à *bande limitée* et  $W$

est appelée *bande passante* du signal. Dans le cas particulier du signal  $A \text{sinc}(2Wt)$ , le spectre est constant sur l'intervalle  $[-W, +W]$ . Il n'en est pas ainsi pour tous les signaux à bande limitée.

Pour conclure ces exemples, nous dirons donc qu'un signal est à *bande limitée* si son spectre s'étend sur une *bande finie* de fréquences et vaut 0 en dehors de cette zone.

## 6.2 La fonction Delta de DIRAC

La fonction Delta de DIRAC, encore appelée *impulsion de DIRAC*, définie et décrite ci-dessous n'est pas une fonction au sens classique des mathématiques. Elle est issue de la théorie des distributions qui sort du cadre de ce cours. Néanmoins, elle va nous être d'un grand secours dans la suite de notre apprentissage des signaux.

### 6.2.1 Définition

La fonction Delta de DIRAC, notée  $\delta(t)$ , est définie par

$$\delta(t) = 0 \text{ pour tout } t \neq 0$$

et

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1$$

Donc, reprenons. La fonction Delta de DIRAC est une fonction qui est nulle partout sauf en 0 où elle n'est pas définie. Néanmoins, son intégrale vaut 1. Mouais, admettons...

Il est possible de donner une autre définition de la fonction  $\delta(t)$  qui incorpore les deux relations précédentes :

$$\boxed{\int_{-\infty}^{+\infty} g(t) \delta(t - t_0) dt = g(t_0)} \quad (6.17)$$

où  $g(t)$  est une fonction continue. Donc, si on prend un signal  $g(t)$ , qu'on le multiplie par une impulsion de DIRAC centrée sur l'instant  $t_0$  et que l'on fait l'intégrale de ce produit, on obtient la valeur du signal  $g(t)$  à l'instant  $t_0$ . Quel intérêt ? Jusque là, aucun... un peu de patience. Nous demandons juste au lecteur d'admettre pour le moment cette définition un peu surprenante.

### 6.2.2 Transformée de FOURIER

Par définition, la transformée de FOURIER de  $\delta(t)$  est donnée par

$$\int_{-\infty}^{+\infty} \delta(t) e^{-j2\pi tf} dt$$

Étant donné que la fonction  $e^{-j2\pi tf}$  évaluée en  $t = 0$  vaut 1, il vient

$$\boxed{\delta(t) \rightleftharpoons 1} \quad (6.18)$$

Donc, le spectre de la fonction Delta de DIRAC s'étend uniformément sur tout l'intervalle de fréquence  $]-\infty, +\infty[$ . Ce signal présente donc une *bande passante infinie*.

Valant zéro presque partout et n'étant pas définie à l'origine, il n'est pas facile de représenter graphiquement la fonction  $\delta(t)$ . Par convention, on la représentera comme à la figure 6.4a. Sa

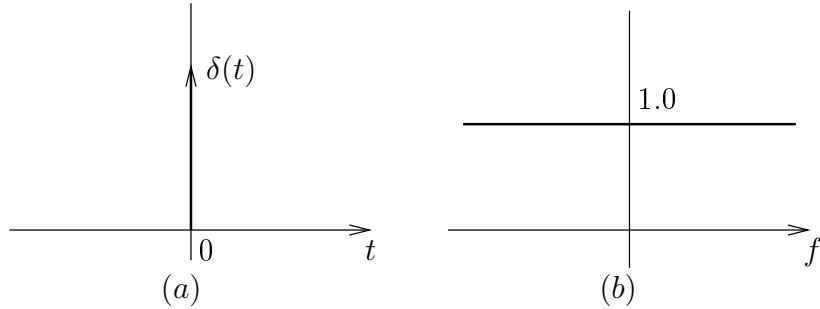


FIGURE 6.4 – (a) La fonction Delta de DIRAC. (b) La transformée de FOURIER de  $\delta(t)$ .

transformée de FOURIER est représentée à la figure 6.4b. On appelle également la fonction Delta de DIRAC l'*impulsion de DIRAC* étant donné son caractère (extrêmement !) limité dans le temps. On parlera également de *raie de DIRAC*.

### 6.2.3 Applications directes

#### Signal continu ou signal DC

Considérons le signal continu suivant

$$g(t) = A$$

Ce signal n'a pas de transformée de FOURIER au sens classique des mathématiques car l'intégrale (7.1) n'existe pas. Néanmoins, en appliquant la propriété de dualité (7.7) de la transformée de FOURIER à la relation (7.14) et étant donné que la fonction  $\delta$  est paire, il vient

$$1 \rightleftharpoons \delta(f) \quad (6.19)$$

et donc

$$A \rightleftharpoons A \delta(f) \quad (6.20)$$

La transformée de FOURIER d'un signal continu comporte donc une seule impulsion de DIRAC située en  $f = 0$ . On dira que ce signal présente une seule *harmonique* en  $f = 0$ . La figure 6.5a illustre ce signal et son spectre pour  $A = 1$ .

#### Signal complexe exponentiel

En appliquant la propriété de translation fréquentielle (7.9) de la transformée de FOURIER à la relation (7.15), nous obtenons

$$e^{j2\pi f_0 t} \rightleftharpoons \delta(f - f_0) \quad (6.21)$$

Le spectre d'un signal complexe exponentiel de fréquence  $f_0$  se limite à une raie située en  $f = f_0$ .

#### Signal sinusoïdal

Considérons le signal

$$g(t) = A \cos(2\pi f_0 t)$$

pour lequel nous connaissons bien (enfin j'espère !) la relation suivante

$$\cos(2\pi f_0 t) = \frac{e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}}{2}$$

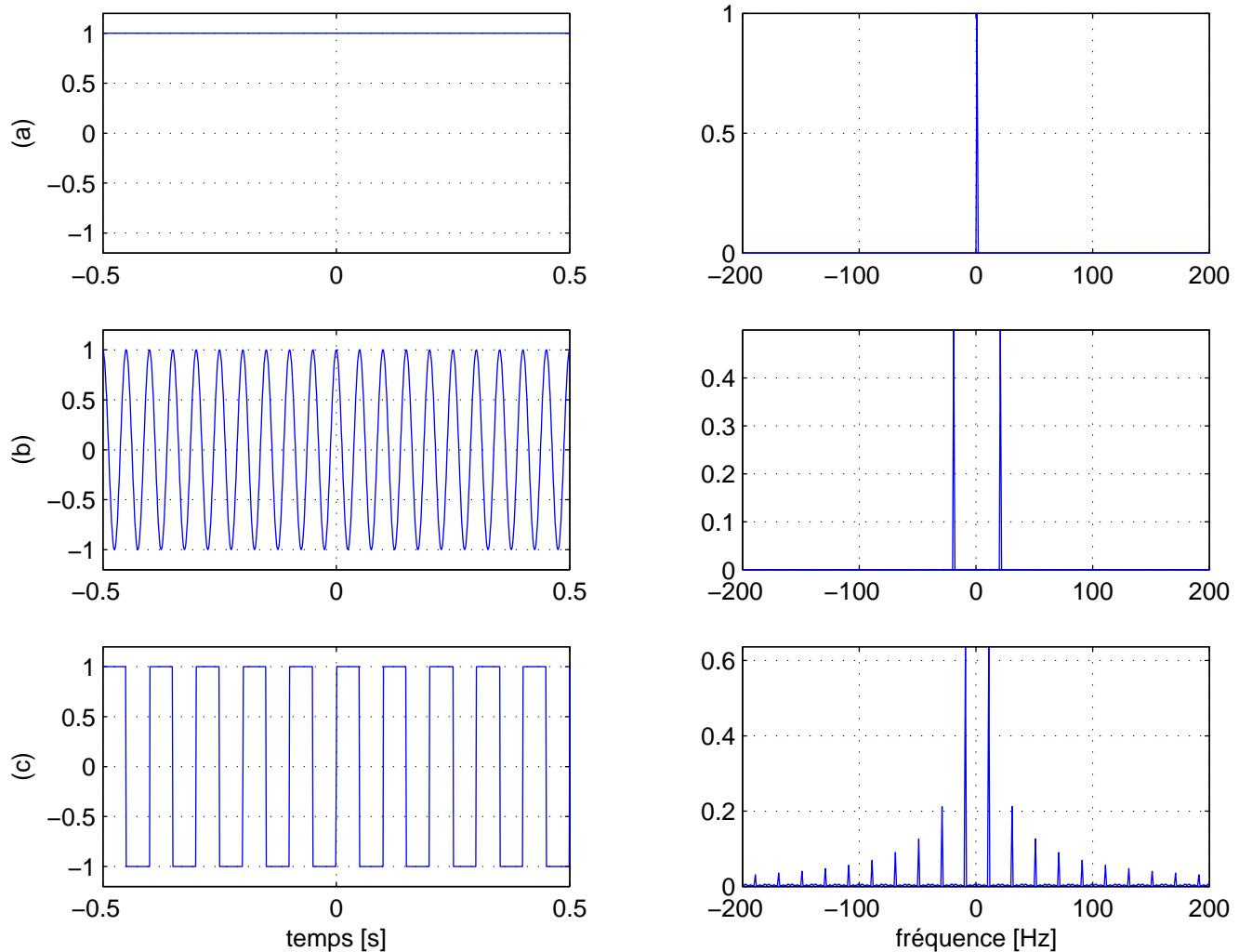


FIGURE 6.5 – (a) Signal continu ( $A = 1$ ). (b) Signal cosinusoidal ( $A = 1$  et  $f_0 = 20$  [Hz]). (c) Signal périodique rectangulaire de fréquence fondamentale  $f_0 = 10$  [Hz].

En utilisant la propriété de linéarité (7.5) de la transformée de FOURIER, il vient

$$\cos(2\pi f_0 t) \Leftrightarrow \frac{\delta(f - f_0) + \delta(f + f_0)}{2} \quad (6.22)$$

Le spectre d'un signal cosinusoidal comporte donc 2 raies pondérées par le facteur 1/2 situées en  $f = \pm f_0$ . Dès lors,

$$G(f) = \frac{A}{2} (\delta(f - f_0) + \delta(f + f_0))$$

Le signal  $g(t)$ , ainsi que son spectre sont représentés à la figure 6.5b pour  $A = 1$  et  $f_0 = 20 [Hz]$ .

De la même manière, on pourrait montrer que

$$\sin(2\pi f_0 t) \Leftrightarrow \frac{\delta(f - f_0) - \delta(f + f_0)}{2j} \quad (6.23)$$

On peut remarquer que les fonctions sinus et cosinus ont le même spectre mais pas la même phase. En effet, graphiquement, ils ont exactement le même comportement ; leur différence réside dans le déphasage de 90°.

### Signaux périodiques

La plupart des signaux périodiques ont une énergie infinie et donc n'ont pas de transformée de FOURIER au sens classique des mathématiques. Néanmoins, l'impulsion de DIRAC et les applications que nous venons d'en tirer vont permettre de nous en sortir.

Tout signal périodique  $g(t)$  fini (c'est-à-dire dont les valeurs restent bornées) de période fondamentale  $T_0$ , et donc de fréquence fondamentale  $f_0 = 1/T_0$ , peut être décomposé en une somme infinie de "sinusoïdes" en usant de la série de FOURIER :

$$g(t) = \sum_{n=-\infty}^{+\infty} c_n e^{j2\pi(\frac{n}{T_0})t} = \sum_{n=-\infty}^{+\infty} c_n e^{j2\pi(nf_0)t} \quad (6.24)$$

où les coefficients  $c_n$  de FOURIER sont donnés par

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{+T_0/2} g(t) e^{-j2\pi(\frac{n}{T_0})t} dt \quad (6.25)$$

En appliquant la propriété de linéarité (7.5) à la relation (6.24) et étant donné la relation (7.17), il vient

$$G(f) = \sum_{n=-\infty}^{+\infty} c_n \delta(f - nf_0) \quad (6.26)$$

Le spectre d'un signal périodique est donc constitué d'une infinité de raies situées aux multiples de la fréquence fondamentale du signal. Toutes ces raies constituent les *harmoniques* du signal. On dit que le spectre d'un signal périodique est *discret*, car le spectre est nul entre les différentes raies. La raie située à l'origine

$$c_0 \delta(f)$$

est appelée *composante continue* du signal. Si  $c_0 = 0$ , nous dirons que le signal  $g(t)$  n'a pas de composante continue. Lorsque l'on parle de la *composante fondamentale* du signal, il s'agit de la composante

$$c_{-1}\delta(f + f_0) + c_1\delta(f - f_0)$$

qui représente les raies situées à la fréquence fondamentale  $f_0$  du signal. Un exemple de signal périodique et son spectre sont représentés à la figure 6.5c.

## 6.3 Retour à la transformée de FOURIER

Nous pouvons à présent donner une meilleure interprétation physique de la transformée de FOURIER. À la section précédente, nous avons observé que le signal (co)sinusoïdal a un comportement particulier vis-à-vis de la transformée de FOURIER. En effet, son spectre est constitué d'une seule raie située à la fréquence de la (co)sinusoïde. Les signaux périodiques, quant à eux, présentent un spectre composé de raies ou d'harmoniques situées aux multiples de la fréquence fondamentale du signal. Chaque harmonique représente donc une composante du signal, composante correspondant à une fréquence. Pour les signaux non périodiques, on pourrait dire que les harmoniques sont tellement proches que le spectre devient une fonction continue de la fréquence. Dès lors, on peut considérer que la grandeur  $G(f)$  constitue la composante fréquentielle du signal  $g(t)$  à la fréquence  $f$ . La transformée de FOURIER peut donc être vue comme une généralisation de la série de FOURIER qui n'est applicable qu'aux signaux périodiques.

Plusieurs notions sont à retenir. La *bande passante* d'un signal est une caractéristique importante de celui-ci ; il s'agit de la bande de fréquence dans laquelle le spectre du signal est non nul. Son spectre est donc nul en dehors de sa bande passante.

Une autre notion importante est celle des hautes et des basses fréquences d'un signal. Les hautes fréquences d'un signal correspondent à ses composantes  $G(f)$  pour  $f$  élevé ; il s'agit donc des variations rapides de  $g(t)$ . Par contre, les basses fréquences du signal correspondent à ses composantes  $G(f)$  pour  $f$  considéré autour de 0. La valeur  $G(0)$  constitue la composante continue du signal, c'est-à-dire sa valeur moyenne. Les basses fréquences correspondent aux variations lentes du signal. Notons qu'il n'y a pas de distinction nette entre les basses et les hautes fréquences. Cela dépend de l'application et de la gamme de fréquence dans laquelle on travaille.

## 6.4 Quelques signaux fondamentaux

### 6.4.1 Définitions

1. Fonction rectangle

$$\text{rect}(t) = \begin{cases} 1 & \text{si } -\frac{1}{2} < t < \frac{1}{2} \\ 0 & \text{sinon} \end{cases}$$

2. Fonction échelon

$$u(t) = \begin{cases} 1 & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$$

3. Fonction signe

$$\text{sign}(t) = \begin{cases} 1 & \text{si } t > 0 \\ -1 & \text{si } t < 0 \end{cases}$$

4. Fonction sinc

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$$

### 6.4.2 Paires de transformées de FOURIER

$$\begin{aligned}
rect\left(\frac{t}{T}\right) &\Leftrightarrow T \operatorname{sinc}(fT) \\
\operatorname{sinc}(2Wt) &\Leftrightarrow \frac{1}{2W} rect\left(\frac{f}{2W}\right) \\
e^{-at}u(t), a > 0 &\Leftrightarrow \frac{1}{a + j2\pi f} \\
e^{-a|t|}, a > 0 &\Leftrightarrow \frac{2a}{a^2 + (2\pi f)^2} \\
e^{-\pi t^2} &\Leftrightarrow e^{-\pi f^2} \\
\delta(t) &\Leftrightarrow 1 \\
1 &\Leftrightarrow \delta(f) \\
\delta(t - t_0) &\Leftrightarrow e^{-j2\pi f_0 t} \\
e^{j2\pi f_0 t} &\Leftrightarrow \delta(f - f_0) \\
\cos(2\pi f_0 t) &\Leftrightarrow \frac{1}{2} [\delta(f - f_0) + \delta(f + f_0)] \\
\sin(2\pi f_0 t) &\Leftrightarrow \frac{1}{2j} [\delta(f - f_0) - \delta(f + f_0)] \\
sign(t) &\Leftrightarrow \frac{1}{j\pi f} \\
\frac{1}{\pi t} &\Leftrightarrow -j sign(f) \\
u(t) &\Leftrightarrow \frac{1}{2}\delta(f) + \frac{1}{j2\pi f} \\
\sum_{k=-\infty}^{+\infty} \delta(t - kT_0) &\Leftrightarrow \frac{1}{T_0} \sum_{n=-\infty}^{+\infty} \delta\left(f - \frac{n}{T_0}\right)
\end{aligned}$$

# Chapitre 7

## Transformée de FOURIER 2D

### 7.1 Transformée de FOURIER 2D

#### 7.1.1 Définition

Soit  $f(x, y)$  un signal 2D (nous dirons image dans la suite) déterministe.

**Définition [Transformée de FOURIER 2D].** La transformée de FOURIER de  $f(x, y)$  est définie par

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi(xu+yv)} dx dy \quad (7.1)$$

À partir de  $F(u, v)$ , il est possible de retrouver exactement l'image  $f(x, y)$  au moyen de

**Définition [Transformée de FOURIER 2D inverse].** La transformée de FOURIER inverse de  $F(u, v)$  est définie par

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{j2\pi(xu+yv)} du dv \quad (7.2)$$

Pour que la transformée de FOURIER d'une image  $f(x, y)$  existe, il faut que l'intégrale (7.1) converge et fournit un résultat fini quelque soit la fréquence  $(u, v)$ . Mathématiquement, il est possible d'introduire un ensemble de contraintes sur la fonction  $f(x, y)$ .

Les signaux de puissance, ayant une énergie infinie, ne possède pas de transformée de FOURIER au sens classique des mathématiques. En particulier, il est impossible de calculer la transformée de FOURIER des signaux périodiques, comme une sinusoïde. En effet, dans ce cas, l'intégrale (7.1) donne un résultat infini quelque soit  $(u, v)$ ... Cela est particulièrement gênant pour l'étude de nombreux systèmes de traitement du signal qui font abondamment usage de ce genre de signal. Néanmoins, la théorie des distributions introduit un signal particulier, l'impulsion de DIRAC  $\delta(x, y)$ , qui permet de résoudre ce problème. Nous y reviendrons très bientôt.

Dans la suite, nous dirons que  $f(x, y)$  et  $F(u, v)$  forment une paire de transformées de FOURIER représentée par

$$f(x, y) \rightleftharpoons F(u, v)$$

En général,  $F(u, v)$  est une fonction à valeurs complexes, ce qui n'est pas spécialement pour ravir le lecteur... Pour rappel, un nombre complexe peut s'exprimer en utilisant la notation module-argument. Nous pouvons donc écrire

$$F(u, v) = \|F(u, v)\| e^{j\theta(u, v)} \quad (7.3)$$

où

- $\|F(u, v)\|$  est appelé *module* de  $F(u, v)$ , ou encore *spectre* de  $f(x, y)$ , et
- $\theta(u, v)$  est appelée *phase* de  $f(x, y)$ .

Dans le cas important où  $f(x, y)$  est une image à valeurs réelles, nous avons

$$F^*(u, v) = F(-u, -v)$$

où \* représente le complexe conjugué. Il vient

$$\boxed{\|F(-u, -v)\| = \|F(u, v)\|}$$

$$\boxed{\theta(-u, -v) = -\theta(u, v)}$$

Dès lors, nous pouvons déduire deux propriétés importantes d'une image à valeurs réelles :

- Le *spectre*  $\|F(u, v)\|$  de l'image est symétrique par rapport à l'origine  $(0, 0)$  du système d'axes  $u - v$ .
- La *phase*  $\theta(u, v)$  de l'image est anti-symétrique par rapport à l'origine  $(0, 0)$  du système d'axes  $u - v$ .

Ces deux propriétés sont très importantes dans le cas du filtrage linéaire des images. En effet, si un traitement dans le domaine fréquentiel modifie le spectre et/ou la phase de l'image de telle sorte qu'une, au moins, de ces deux propriétés ne soit plus vérifiée, l'image obtenue après transformée de FOURIER inverse ne sera plus réelle...

### 7.1.2 Propriétés

#### 1. Séparabilité

En permutant l'ordre d'intégration dans (7.1), nous avons

$$F(u, v) = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi xu} dx \right] e^{-j2\pi yu} dy \quad (7.4)$$

La transformée de FOURIER d'une image  $f(x, y)$  peut se réaliser en deux étapes : (i) transformée de FOURIER 1D de la fonction  $f(x, y)$  pour tout  $y$  fixé, transformant la variable  $x$  en la variable  $u$  et (ii) transformée de FOURIER 1D de la fonction obtenue pour tout  $u$  fixé, transformant la variable  $y$  en la variable  $v$ .

#### 2. Linéarité

Soient  $f_1(x, y) \rightleftharpoons F_1(u, v)$  et  $f_2(x, y) \rightleftharpoons F_2(u, v)$ . Alors, pour toutes constantes  $c_1$  et  $c_2$ , nous avons

$$c_1 f_1(x, y) + c_2 f_2(x, y) \rightleftharpoons c_1 F_1(u, v) + c_2 F_2(u, v) \quad (7.5)$$

#### 3. Dilatation spatiale ou homothétie

Soit  $f(x, y) \rightleftharpoons F(u, v)$ . Nous avons

$$f(ax, by) \rightleftharpoons \frac{1}{|ab|} F\left(\frac{u}{a}, \frac{v}{b}\right) \quad (7.6)$$

La fonction  $f(ax, by)$  représente une version de  $f(x, y)$  compressée dans l'espace par un facteur  $a$  dans la direction  $x$  et par un facteur  $b$  dans la direction  $y$ . Une compression

dans le domaine spatial équivaut à une dilatation dans le domaine fréquentiel et vice versa.

#### 4. Dualité

Si  $f(x, y) \rightleftharpoons F(u, v)$ , alors

$$F(x, y) \rightleftharpoons f(-u, -v) \quad (7.7)$$

#### 5. Translation spatiale

Si  $f(x, y) \rightleftharpoons F(u, v)$ , alors

$$f(x - x_0, y - y_0) \rightleftharpoons F(u, v) e^{-j2\pi(ux_0 + vy_0)} \quad (7.8)$$

Il en résulte que le fait de translater la fonction  $f(x, y)$  de  $(x_0, y_0)$  ne modifie pas le module de la transformée de FOURIER, par contre sa phase est modifiée d'un facteur  $-j2\pi(ux_0 + vy_0)$ .

#### 6. Translation fréquentielle

Si  $f(x, y) \rightleftharpoons F(u, v)$ , alors

$$f(x, y) e^{j2\pi(u_0x + v_0y)} \rightleftharpoons F(u - u_0, v - v_0) \quad (7.9)$$

La multiplication de la fonction  $f(x, y)$  par le facteur  $e^{j2\pi(u_0x + v_0y)}$  est équivalente à une translation de la transformée de FOURIER  $F(u, v)$  dans le domaine fréquentiel.

#### 7. Aire

Si  $f(x, y) \rightleftharpoons F(u, v)$ , alors

$$F(0, 0) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy \quad (7.10)$$

Le coefficient  $F(0, 0)$ , appelé parfois *composante DC*, est la somme des valeurs de pixel de l'image. Sa dynamique est donc très importante par rapport aux autres coefficients. Ceci pose des problèmes pratiques qui amène à traiter séparément ce coefficient.

#### 8. Multiplication dans le domaine spatial

Si  $f(x, y) \rightleftharpoons F(u, v)$  et  $g(x, y) \rightleftharpoons G(u, v)$ , alors

$$f(x, y) g(x, y) \rightleftharpoons \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(\alpha, \beta) G(u - \alpha, v - \beta) d\alpha d\beta = (F \otimes G)(u, v) \quad (7.11)$$

L'intégrale apparaissant dans cette expression est connue sous le nom d'*intégrale de convolution* dans le domaine fréquentiel. L'opérateur  $\otimes$  est appelé opérateur de convolution et la nouvelle fonction  $(F \otimes G)$  ainsi créée est appelée *produit de convolution* de  $F$  et  $G$ . Dès lors, une multiplication dans le domaine spatial est équivalente à une convolution dans le domaine fréquentiel.

#### 9. Convolution dans le domaine spatial

Si  $f(x, y) \rightleftharpoons F(u, v)$  et  $g(x, y) \rightleftharpoons G(u, v)$ , alors

$$(f \otimes g)(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\alpha, \beta) g(x - \alpha, y - \beta) d\alpha d\beta \rightleftharpoons F(u, v) G(u, v) \quad (7.12)$$

Dès lors, le produit de convolution dans le domaine spatial équivaut à un simple produit dans le domaine fréquentiel.

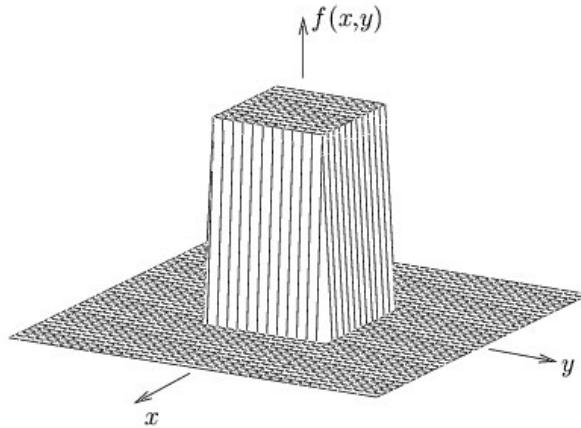


FIGURE 7.1 – Illustration de la fonction Rectangle.

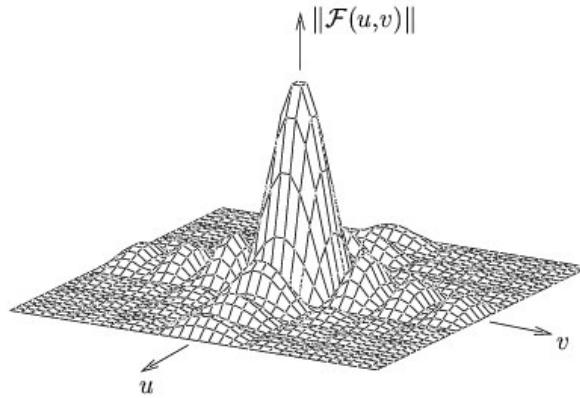


FIGURE 7.2 – Module de la transformée de FOURIER de la fonction Rectangle.

### 7.1.3 Exemples

#### Exemple 1

Considérons l'image  $f(x,y)$  définie par

$$f(x,y) = A \operatorname{Rect}_{a,b}(x,y)$$

où

$$\operatorname{Rect}_{a,b}(x,y) = \begin{cases} 1 & \text{si } |x| < \frac{a}{2}, |y| < \frac{b}{2} \\ 0 & \text{sinon} \end{cases}$$

Cette fonction est appelée *fonction Rectangle*. L'image  $f(x,y)$  vaut donc  $A$  à l'intérieur du rectangle de longueur  $a$  et de largeur  $b$  dont les côtés sont parallèles aux axes  $x$  et  $y$ , et zéro partout ailleurs. La transformée de FOURIER de l'image  $f(x,y)$  est donnée par

$$\begin{aligned} F(u,v) &= \int_{-a/2}^{+a/2} dx \int_{-b/2}^{+b/2} dy A e^{-j2\pi(xu+yv)} \\ &= Aab \operatorname{sinc}(au) \operatorname{sinc}(bv) \end{aligned}$$

Une représentation de l'image  $f(x,y)$  est donnée à la figure 7.1 tandis que le module de sa transformée de FOURIER est représenté à la figure 7.2.

#### Exemple 2

Considérons l'image  $f(x,y)$  définie par

$$f(x,y) = A \operatorname{Disque}_R(x,y)$$

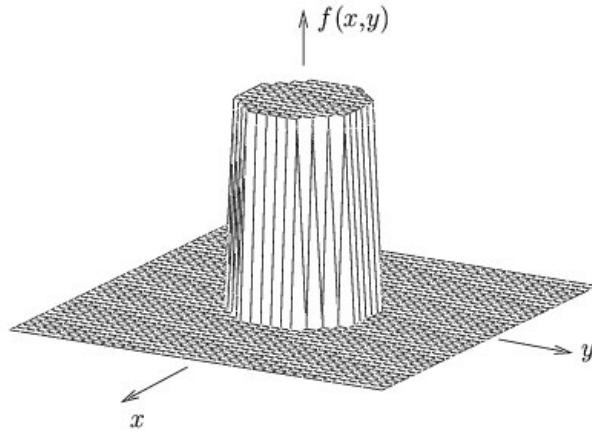


FIGURE 7.3 – Illustration de la fonction Disque.

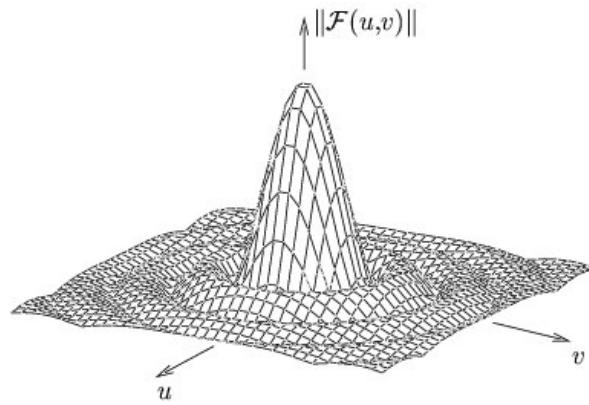


FIGURE 7.4 – Module de la transformée de FOURIER de la fonction Disque.

où

$$\text{Disque}_R(x, y) = \begin{cases} 1 & \text{si } \sqrt{x^2 + y^2} < R \\ 0 & \text{sinon} \end{cases}$$

est appelée *fonction disque*. L'image  $f(x, y)$  vaut donc  $A$  sur le disque de rayon  $R$  et zéro partout ailleurs. La transformée de FOURIER de l'image  $f(x, y)$  est donnée par

$$\begin{aligned} F(u, v) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi(xu+yv)} dx dy \\ &= \int_0^R r dr \int_0^{2\pi} A e^{-j2\pi r(u \cos \theta + v \sin \theta)} d\theta \end{aligned}$$

où nous avons effectué un changement de variables des coordonnées cartésiennes vers les coordonnées polaires. En se basant sur les propriétés des fonctions de BESSEL, nous obtenons le résultat suivant

$$F(u, v) = A R \frac{J_1(2\pi R \sqrt{u^2 + v^2})}{\sqrt{u^2 + v^2}}$$

où  $J_1(r)$  est la fonction de BESSEL d'ordre 1. Il est à noter que la transformée de Fourier de l'image  $f(x, y)$  est purement réelle et à symétrie radiale, c'est-à-dire qu'elle ne dépend que de la distance  $\sqrt{u^2 + v^2}$  à l'origine. L'image  $f(x, y)$  est représentée à la figure 7.3 tandis que le module de sa transformée de FOURIER est représenté à la figure 7.4.

### Exemple 3 : dual de la fonction Rectangle

Pour rappel, nous avons la paire de transformée de FOURIER suivante

$$\text{Rect}_{a,b}(x, y) \Leftrightarrow ab \text{sinc}(au) \text{sinc}(bv)$$

En utilisant la propriété de dualité (relation 7.7) de la transformée de FOURIER, nous obtenons

$$\text{sinc}(ax) \text{sinc}(by) \Leftrightarrow \frac{1}{ab} \text{Rect}_{a,b}(u, v)$$

La nouvelle image ainsi obtenue est dite à bande limitée car son spectre est limité à une région finie du plan  $u - v$ . En l'occurrence, cette région est, dans ce cas-ci, un rectangle de longueur  $a$  et de largeur  $b$  centré à l'origine. D'après la définition de la fonction rectangle, le spectre est nul est dehors de ce rectangle.

### Exemple 4 : dual de la fonction Disque

Pour rappel, nous avons la paire de transformée de Fourier suivante

$$\text{Disque}_R(x, y) \Leftrightarrow R \frac{J_1(2\pi R\sqrt{u^2 + v^2})}{\sqrt{u^2 + v^2}}$$

En utilisant la propriété de dualité (relation 7.7) de la transformée de Fourier, nous obtenons

$$\frac{J_1(2\pi f_0 \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}} \Leftrightarrow \frac{1}{f_0} \text{Disque}_{f_0}(u, v)$$

La nouvelle image ainsi obtenue est à bande limitée. Dans ce cas, le spectre est limité au disque de rayon  $f_0$  centré à l'origine.  $f_0$  est appelée fréquence radiale de coupure.

## 7.2 La fonction Delta de DIRAC

La fonction Delta de DIRAC, encore appelée *impulsion de DIRAC*, définie et décrite ci-dessous n'est pas une fonction au sens classique des mathématiques. Elle est issue de la théorie des distributions qui sort du cadre de ce cours. Néanmoins, elle va nous être d'un grand secours dans la suite.

### 7.2.1 Définition

La fonction Delta de DIRAC, notée  $\delta(x, y)$ , est définie par

$$\delta(x, y) = 0 \text{ pour tout } (x, y) \neq (0, 0)$$

et

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x, y) dx dy = 1$$

La fonction Delta de DIRAC est une fonction qui est nulle partout sauf à l'origine  $(0, 0)$  où elle n'est pas définie. Néanmoins, son intégrale vaut 1.

Il est possible de donner une autre définition de la fonction  $\delta(x, y)$  qui incorpore les deux relations précédentes :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(x - x_0, y - y_0) dx dy = f(x_0, y_0) \quad (7.13)$$

où  $f(x, y)$  est une fonction continue. Donc, si on prend une image  $f(x, y)$ , qu'on le multiplie par une impulsion de DIRAC centrée sur l'instant sur le pixel  $(x_0, y_0)$  et que l'on fait l'intégrale de ce produit, on obtient la valeur de l'image  $f(x, y)$  au pixel  $(x_0, y_0)$ .

### 7.2.2 Transformée de FOURIER

Par définition, la transformée de FOURIER de  $\delta(x, y)$  est donnée par

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x, y) e^{-j2\pi(xu+yv)} dx dy$$

Étant donné que la fonction  $e^{-j2\pi(xu+yv)}$  évaluée à l'origine  $(0, 0)$  vaut 1, il vient

$$\boxed{\delta(x, y) = 1} \quad (7.14)$$

Donc, le spectre de la fonction Delta de DIRAC s'étend uniformément sur tout l'intervalle de fréquence  $]-\infty, +\infty[ \times ]-\infty, +\infty[$ . Cette image présente donc une *bande passante infinie*.

### 7.2.3 Applications directes

#### Image continue ou constante

Considérons l'image continue suivante

$$f(x, y) = A$$

Cette image n'a pas de transformée de FOURIER au sens classique des mathématiques car l'intégrale (7.1) n'existe pas. Néanmoins, en appliquant la propriété de dualité (7.7) de la transformée de FOURIER à la relation (7.14) et étant donné que la fonction  $\delta$  est symétrique par rapport à l'origine, il vient

$$\boxed{1 = \delta(u, v)} \quad (7.15)$$

et donc

$$A = A \delta(u, v) \quad (7.16)$$

La transformée de FOURIER d'une image continue ou constante comporte donc une seule raie de DIRAC située à l'origine  $(0, 0)$  du plan  $u - v$ .

#### Image complexe exponentielle

En appliquant la propriété de translation fréquentielle (7.9) de la transformée de FOURIER à la relation (7.15), nous obtenons

$$e^{j2\pi(u_0x+v_0y)} = \delta(u - u_0, v - v_0) \quad (7.17)$$

Le spectre d'une image complexe exponentielle de fréquence  $(u_0, v_0)$  se limite donc à une raie située en  $(u_0, v_0)$  du plan  $u - v$ .

### Image sinusoïdale

Considérons l'image

$$f(x, y) = A \cos(2\pi(u_0x + v_0y))$$

pour lequel nous avons la relation suivante

$$\cos(2\pi(u_0x + v_0y)) = \frac{e^{j2\pi(u_0x + v_0y)} + e^{-j2\pi(u_0x + v_0y)}}{2}$$

En utilisant la propriété de linéarité (7.5) de la transformée de FOURIER, il vient

$$\cos(2\pi(u_0x + v_0y)) \Leftrightarrow \frac{\delta(u - u_0, v - v_0) + \delta(u + u_0, v + v_0)}{2} \quad (7.18)$$

Le spectre d'une image cosinusoïdale comporte donc 2 raies pondérées par le facteur 1/2 situées en  $(u_0, v_0)$  et  $(-u_0, -v_0)$ . Dès lors,

$$F(u, v) = \frac{A}{2} (\delta(u - u_0, v - v_0) + \delta(u + u_0, v + v_0))$$

De la même manière, on pourrait montrer que

$$\sin(2\pi(u_0x + v_0y)) \Leftrightarrow \frac{\delta(u - u_0, v - v_0) - \delta(u + u_0, v + v_0)}{2j} \quad (7.19)$$

On peut remarquer que les fonctions sinus et cosinus ont le même spectre mais pas la même phase.