



# Mémoire de Magister

Présenté à l'université de Guelma  
Faculté des sciences et de l'ingénierie

Département : Génie électrique  
Option : Signaux et images

Par : DRICI DJALEL

---

## Méthodes Markoviennes et l'analyse d'ondelettes pour le traitement d'images satellitaires

---

Sous la direction du Docteur : **BOUKROUCHE Abdelhani**

### JURY

**Président :**

**TEBBIKH Hichem**      Professeur      Université de Guelma

**Examineurs :**

**DEHICI Abdolkader**      Maître de conférence      Université de Guelma

**LAKEL Rabah**      Maître de conférence      Université de Annaba

## Remerciements

**C**e travail a été réalisé au sein de l'équipe du Laboratoire d'Automatique et d'informatique de Guelma (LAIG).

Je tiens à remercier sincèrement tous les membres de l'équipe du laboratoire pour la l'ambiance studieuse et leurs multiples encouragements.

Je tiens à remercier mon encadreur Monsieur Abdelhani BOUKROUCHE Maître de conférence à l'université de Guelma qui a suivi avec intérêt le déroulement de mes travaux, Je suis reconnaissant de sa patience et de ses conseils avisés.

Mes sincères remerciement à Monsieur Hichem TABBIKH professeur à université de Guelma d'avoir accepter de présider le jury, je tiens à lui exprimer mon extrême gratitude.

Je remercie Monsieur DEHICI Abdelkader, Maître de conférence à l'Université de Guelma pour l'intérêt qu'il a bien voulu porter à ce travail en acceptant de faire partie du jury.

Je remercie Monsieur, LAKEL Rabah, Maître de conférence à l'Université de Annaba d'avoir accepter d'expertiser ce travail et de faire partie de mon jury de soutenance.

Mes vifs remerciements vont également à mes enseignants, Madame KACHI djemaa, Messieurs BEDDA mouldi, et YOUNSI Abdelaziz, SIRIDI Abdelhamid pour leurs encouragements et conseils.

Je remercier tous ceux qui m'ont aider de près ou de loin pour l'élaboration de ce modeste mémoire.

## ملخص

أثناء نقل الإشارة و تخزينها قد تتعرض إلى بعض التشوهات، سواء ناتجة عن وسيلة التقاطها أو المحيط الذي أخذت فيه، و محاولة استرجاعها ليس بالأمر السهل، خاصة إذا كنا نعرف أسباب التدهور، لان اغلبها موصوفة بالترددات العالية التي تجعل هذه الأخيرة من ضمن المسائل العكسية المطروحة بشكل سيئ.

في هذه الأطروحة، تطرقنا إلى دراسة المسائل العكسية، وكيفية حلها وجعلها مطروحة بشكل جيد، من اجل استرجاع الصور ومحاولة الوصول إلى صيغة اقرب مما يمكن لحقيقتها، وذلك من خلال دراستنا إلى حقل ماركوف الذي يقوم بمعالجة الصور وبشكل دقيق، وكذلك التحليل المتعدد الموجات الذي بدوره يساعد على تسريع عملية الاسترجاع و تحسين الصورة أيضا.

لقد عرضنا هنا مبادئ وطرق هذه التقنيات، وكيفية استغلال ايجابياتها في معالجة الصور الملتقطة بالقمر الصناعي ومن ثم الاستفادة منها في شتى المجالات.

## **Résumé**

La modélisation de l'image et de ses discontinuités par champ markovien a pour objectif dans la restauration d'image de préserver les contours.

Dans cette mémoire l'image est d'abord décomposée à différents niveaux de résolutions par transformée en ondelettes. Un champ markovien est défini pour chaque résolution et l'image est restaurée itérativement en fonction des résolutions croissantes. Cet algorithme est analysé comparativement à un algorithme monorésolution et montre son efficacité en terme de réduction de temps de calcul et qualité de l'image restaurée.

## **MOTS CLÉS**

Champ de Markov, processus de lignes, restauration d'image, analyse multirésolution, relaxation déterministe.

## **Abstract**

The restoration methods involving Markov random fields with line process for the discontinuities are known to give good results in edge preserving image restoration . The Markov model is developed on the image decomposed at several resolutions using a wavelet transform.

Simulated Annealing algorithm or ICM (Iterated Conditional Modes) performs the minimization of the non convex criterion . Several multiresolution algorithms are developed and results are compared with those in monoresolution . These multiresolution algorithms provide better results with less computation time .

## **KEY WORDS**

Markov random field, line processus, image restoration, multiresolution analysis, deterministic relaxation.

## LISTE DES FIGURES

### Chapitre I

Figure 1.1 : Modèle du processus d'observation (acquisition d'image bruité).	09
Figure 1.2 : Voisinages (a) d'ordre 1 et (b) d'ordre 2 et cliques correspondantes.	12
Figure 1.3 : Probabilité d'acceptation pour la dynamique de Metropolis.	32
Figure 1.4 : Probabilité d'acceptation sigmoïde pour la dynamique de Barker (loi du Thermostat). Remarque : plus la force est grande, et plus elle "tire" vers le nouvel état proposé.	32
Figure. 1.5 : Paysage énergétique et importance de l'initialisation dans le cas de la relaxation déterministe (Pour simplifier la visualisation on a représenté un espace des configurations unidimensionnel).	46

### Chapitre II

Figure 2.1 : a) transformée de Fourier d'un signal $s(t)$ , b) bon échantillonnage, c) repliement du spectre.	52
Figure 2.2 : la fonction $\phi(t) = X_{[-\tau, \tau]}(t)$ $\phi(t) = X_{[-\tau, \tau]}(t)$ .	53
Figure 2.3 : Fenêtrage temps-fréquence de la STFT ( $t^* = w^* = 0$ ).	56
Figure 2.4 : Fenêtrage temps-fréquence pour la transformée en ondelettes.	63
Figure 2.5 : représente la nature hiérarchique de l'analyse multirésolution .	66
Figure 2.6 : Décomposition en cascade.	70
Figure 2.7 : Reconstruction unidimensionnelle par ondelettes.	71
Figure 2.8 : Fonction d'échelle et ondelette de Meyer.	73
Figure 2.9 : Fonctions d'échelles et ondelettes de Daubechies pour $N = 2, 4, 8$ .	74
Figure 2.10 : Fonction d'échelle et ondelette Symlets pour $N = 8$ .	75
Figure 2.11 : Fonction d'échelle et ondelette Coiflets pour $N = 5$ .	76
Figure 2.12 : Décomposition et reconstruction en deux dimensions.	78
Figure 2.13 : Dispositions spatiales pour deux niveaux de décomposition.	79
Figure 2.14 : Analyse multirésolution appliquée à l'image de Souk Ahras à l'aide de l'algorithme de Malat.	82
Figure 2.15 : Analyse multirésolution appliquée à l'image de Lena à l'aide de l'algorithme de Mallat.	83

### Chapitre III

Figure 3.1 : Machinerie d'un satellite.	86
---	----

Figure 3.2 : Capteur à transfert de charge.	88
Figure 3.3 : Le principe du balayage en râteau ou puch-broom.	89
Figure 3.4 : Exemple d'image satellitaire.	90
Figure 3.5 : Instrument d'acquisition.	92
Figure 3.6 : Exemple de La méthode ICM.	95
Figure 3.7 : Profil a la ligne 65.	96
Figure 3.8 : Exemple de La méthode recuit simulé.	98
Figure 3.9 : Profil a la ligne 65	99
Figure 3.10 : Évolution du SNR énergie (rapport des énergies) en fonction de l'indice de résolution de l'image.	100
Figure 3.11 : Schéma d'une itération de l'algorithme de restauration multirésolution.	103
Figure 3.12 : La décomposition d'image par l'analyse multirésolution.	104
Figure 3.13 : La restauration d'image par la méthode de recuit simulé et l'analyse multirésolution.	105
Figure 3.14 : Schéma de la dernière itération de l'algorithme multirésolution avec traitement de la sous-image de coefficients en ondelettes $C_1^Y$ .	107
Figure 3.15 : Exemple d'application sur l'image de Souk Ahras.	108
Figure 3.16. Profil l'image de Souk Ahras a la ligne 100	109
Figure 3.17 : Exemple d'application sur l'image d'hôtel Emarate.	110
Figure 3.18 : Profil l'image d'hôtel Emarate a la ligne 175.	111
Figure 3.19 : Exemple d'application sur l'image de cote méditerranée.	112
Figure 3.20 : Profil l'image méditerranée a la ligne 475.	113

# TABLE DE MATIERES

	<b>Page</b>
<b>Introduction générale</b>	01
<b><i>Chapitre I Le Champs de Markov en analyse d'image</i></b>	
<b><i>Partie I</i></b>	
1.1. Introduction	03
1.2. Régularisation Bayesienne	04
1.2.1. Introduction et notations	04
1.2.2. Règle de Bayes	05
1.2.2.1 L'estimateur du MAP	06
1.2.2.2 L'estimateur du MPM	06
1.2.2.3 L'estimateur du champ moyen	07
1.2.3. Lien statistique entre observation et étiquettes	07
1.2.4. Probabilité a priori : modèle Markovien spatial	10
1.3. Champs de Markov	11
1.3.1. Définitions	11
1.3.2. Théorème d'équivalence	13
1.3.3. Le problème de l'estimation des paramètres	14
1.4. Estimateur MAP et minimisation de l'énergie	15
1.4.1 Construction de la fonction de coût	15
1.4.2. Remarque sur l'estimateur MPM et MF	17
1.5. Débruitage d'images binaires et modèle d'Ising	18
1.5.1. Définition de la fonction de coût ou énergie	18
1.5.2. Choix des paramètres $\beta_S$ et $\beta_o$	20
1.6. Conclusion	22
<b><i>Partie II</i></b>	
1.7. Recuit Simulé	23
1.7.1. Introduction	23
1.7.2. Les algorithmes d'optimisation : généralités	24
1.8. Algorithmes stochastiques	25
1.8.1. Introduction	25
1.8.2. Les procédures d'échantillonnage dynamiques	26

1.8.2.1	Chaînes de Markov. Théorème d'échantillonnage dynamique	26
1.8.2.2	Echantillonneur de Metropolis (ou dynamique de Metropolis)	30
1.8.2.3	L'échantillonneur de Gibbs	36
1.8.2.4	Régime transitoire	39
1.8.2.5	Parallélisation des procédures d'échantillonnage	40
1.9.	Conclusion	42

### **Partie III**

1.10.	La procédure de refroidissement ou Recuit Simulé	43
1.10.1	Recuit continu	43
1.10.2.	Algorithmes avec descente en température par paliers	44
1.10.3.	Critère d'arrêt : définition heuristique du "recuit sérieux"	44
1.11.	Algorithmes déterministes	46
1.11.1.	Introduction	46
1.12.	Conclusion	48

### **Chapitre II La Transformée en Ondelettes**

2.1.	Introduction	49
2.1.1.	Analyse de Fourier	50
2.1.1.1.	Transformée de Fourier	50
2.1.1.2.	Echantillonnage	51
2.1.1.3.	Inconvénients de l'analyse de Fourier	52
2.1.2.	Analyse temps fréquence	53
2.1.2.1.	Le fenêtrage	53
2.1.2.2.	Transformée de Fourier à fenêtre glissante	55
2.1.2.2.1.	Propriétés de la STFT	56
2.1.2.3.	STFT discrète	57
2.1.2.4.	Bilan sur la STFT	58
2.1.3.	Les Ondelettes	59
2.1.3.1	Introduction	59
2.1.3.2	Transformée en ondelettes continue	59
2.1.3.3	Fenêtre temps-fréquence	61
2.1.3.4	Quelques propriétés	63
2.1.3.5	Transformée en ondelettes discrète	64
2.2.	Analyse multirésolution et algorithmes	65

2.2.1.	Présentation de l'analyse multirésolution	65
2.2.1.1	Définition d'une analyse multirésolution	66
2.2.1.2	Relations de reconstruction et de décomposition	67
2.2.1.3	Algorithmes	68
2.2.1.3.1	Algorithme de décomposition	69
2.2.1.3.2	Algorithme de reconstruction	70
2.2.2.	Bilan	71
2.3.	Exemples d'ondelettes	72
2.3.1.	Ondelettes orthogonales	72
2.4.	Bases d'ondelettes a deux dimensions	76
2.4.1.	Principe	76
2.4.2.	Transformée en ondelettes rapide à 2D	77
2.4.3.	Les critères de qualités des ondelettes utilisées en traitement	79
2.4.4.	Choix d'ondelette	81
2.4.5.	Exemple	81
2.5.	Conclusion	
<b><i>Chapitre III Application à la restauration d'images</i></b>		
3.1.	Généralités	85
3.1.1.	Application de l'imagerie satellitaire	85
3.1.1.1.	Les spatiocartes	85
3.1.1.2.	Les modèles numériques de terrain (MNT)	85
3.1.1.3.	Systèmes d'information géographique (SIG)	85
3.1.1.4.	Suivi des subsidences	86
3.2.	Acquisition des images satellites	86
3.2.1.	La charge utile	87
3.2.2.	La plateforme	88
3.3.	Acquisition des images	88
3.3.1.	Les système d'acquisition	88
3.4.	L'obtention de l'image satellite numérique	90
3.5.	Différents modes d'acquisition	91
3.5.1.	Le mode multibande ou multispectral	91
3.5.2.	Le mode panchromatique	91
3.5.3.	Les différents instruments	91

3.6. Transfert des images du satellite vers la Terre	92
3.7. Traitement et interprétation des images	93
3.8. Application de la méthode ICM	94
3.9. Application de la méthode recuit simulé	97
3.10 Analyse du rapport signal a bruit (SNR) en fonction de la résolution	100
3.11. Restauration multirésolution	102
3.11 .1. L'algorithme multirésolutio	102
3.11.2. Interprétation des Résultats	105
3.12. Restauration des coefficients d'ondelettes	106
3.13. Résultats	108
3.14. Interprétation des résultats	113
Conclusion et perspective	114
<b>Bibliographie</b>	115

# **Introduction générale**

# **Champs de Markov en analyse d'images**

# Partie 1

## Introduction générale

D'une manière générale, les images numériques (images satellites, médicales...) sont dégradées lors de leur acquisition.

Les traitements de restauration sont souvent indispensables pour pouvoir améliorer la qualité de ces images numériques, et pouvoir les exploiter ultérieurement. Nous nous intéressons dans ce travail au problème de la suppression de bruit sur une image dégradée par un bruit blanc additif. Ce problème fait partie du cas général des 'problèmes mal posés'. Le but est de lisser l'image, tout en préservant ses contours.

La modélisation de l'image par champ de Markov

L'utilisation des champs de Markov en traitement d'image est attractif puisque ceux-ci fournissent un cadre probabiliste très générale pour l'analyse d'image et permettent d'avoir une interprétation à la fois locale et globale des dépendances des variables aléatoires, ils sont devenu très populaire depuis les travaux de Geman et Geman. Ces travaux ont permis de développer des algorithmes de simulation des champs de Markov en les interprétant comme des champs de Gibbs.

Le modèle utilisé est celui de la régularisation bayésienne par champs de Markov, ces approches permettent de modéliser notre connaissance sur les processus étudiés, les problèmes mathématiques posés (minimisation de fonctions non convexes, nombre de variables important), impliquent un temps de calcul important pour la recherche de la solution optimale (méthode de relaxation du type recuit simulé). Les méthodes déterministes permettent dans certains cas, d'approcher la solution optimale et sont plus rapides (MFA, ICM, GNC). Cependant, leur temps de calcul reste important pour des images 256 x 256 pixels par exemple.

On a présenté une méthode de restauration utilisant une analyse multirésolution de l'image et une modélisation markovienne. L'hypothèse de Markov est alors faite pour toutes les images aux différents niveaux de résolution, au même titre qu'elle est faite usuellement à la résolution la plus fine (image numérique d'origine).

On présente un algorithme multirésolution performant pour la restauration d'image bruitée, qui est le point de départ de recherches sur d'autres modélisations multirésolution mieux adaptées.

Nous exploitons deux propriétés de la multirésolution :

- D'une part les techniques multirésolutions permettent d'accélérer significativement la convergence des algorithmes de traitement d'image ;
- D'autre part elles permettent une meilleure adaptation des traitements en fonction de la résolution et donc d'améliorer significativement la qualité de la restauration.

La méthode de relaxation utilisée pour la minimisation du critère non convexe est la méthode aléatoire, appelée recuit simulé.

## Organisation du document

Nous avons organisé ce document de la façon suivante :

Dans le chapitre 1, la première partie nous avons présenté une approche probabiliste générale permettant de régulariser les problèmes mal posés rencontrés en vision bas-niveau. Ceci est possible grâce à l'utilisation d'estimateurs bayesiens. Dans la deuxième partie nous présentons la technique dite du recuit simulé, méthode d'optimisation stochastique permettant en principe de venir à bout de n'importe quel problème d'optimisation difficile alors que la troisième partie porte la procédure de refroidissement (une loi adéquate de décroissance de la température) celles-ci vont permettre d'approcher la configuration d'étiquettes minimisant une fonction de coût donnée

Dans le chapitre 2, nous présenterons tout d'abord un panorama de l'analyse temps-fréquence et leurs propriétés fondamentales. En suite nous rappellerons les théories de la transformée en ondelettes et de l'analyse multirésolution.

Le chapitre 3 porte sur l'étude théorique et expérimentale des algorithmes de restauration d'images bruitées à base de ICM (Iterated Conditional Modes) et de champs de Markov avec l'analyse multirésolution.

## 1.1. Introduction

L'utilisation des champs de Markov en traitement d'image est attractif puisque ceux-ci fournissent un cadre probabiliste très général pour l'analyse d'image et permettent d'avoir une interprétation à la fois locale et globale des dépendances des variables aléatoires, ils sont devenu très populaire depuis les travaux de Geman et Geman [10]. Ces travaux ont permis de développer des algorithmes de simulation des champs de Markov en les interprétant comme des champs de Gibbs.

Le modèle utilisé est celui la régularisation Bayesienne par champs de Markov, Ces approches permettent de modéliser notre connaissance sur Les processus étudiés, Nous allons introduire dans un premier temps quelques définitions et propriétés générales sur les Régularisations Bayesienne puis, nous nous intéresserons aux propriétés des modèles Markovien puis, les algorithmes d'optimisation sur le champs de Markov.

## 1.2. Régularisation Bayésienne

### 1.2.1. Introduction et notations

L'extraction de primitives de l'image en vision bas niveau est en général un problème mal posé au sens d'Hadamard [4], i.e., pour une donnée unique du problème, il existe plusieurs solutions possibles compatibles avec l'observation. L'introduction d'un certain nombre de contraintes permet de restreindre l'ensemble des solutions. Cette technique de tri a priori porte le nom de régularisation. Par rapport à la régularisation standard, qui met définitivement de côté un sous-ensemble de solutions, la régularisation Bayésienne est une approche probabiliste le problème est modélisé de telle sorte qu'une distribution de probabilités pondérant chacune des solutions potentielles est construite sur l'ensemble des solutions possibles.

Les notations suivantes seront appliquées tout au long de ce travail. Leur interprétation concrète dépendra du problème traité :

Le support de l'image (grille de pixels ou sites) :

$S = \{0 \dots n - 1\} \times \{0 \dots m - 1\}$  ensemble de sites de l'image de taille  $n \times m$ .  $N = mxn$  est donc le nombre de pixels ou sites de l'image.  $q, r$  ou  $s \in S$  sera la notation usuelle Pour un site particulier de l'image. Enfin,  $V(s)$  dénotera un ensemble de sites autour de  $s \in S$  appelé voisinage de  $s$ .

L'image (image "réelle" observée par le capteur) :

$I = \{I(s) \in G_I, s \in S\}$  est le champ image (considéré comme un champ aléatoire).

$G_I$  est un ensemble continu ou discret, fini ou infini de niveaux de gris de l'image réelle. Le champ (déterminé)  $i = \{i(s) \in G_I, s \in S\}$  représente une image particulière (Réalisation du champ image). On notera  $\Omega_I$  l'ensemble des réalisations possibles du champ image :  $\Omega_I = G_I^{n \times m}$

L'observation (résultat de l'observation de l'image "réelle") :

$O = \{O(s) \in G_O, s \in S\}$  est le champ d'observation associé à l'image.  $G_O$  est un ensemble continu ou discret, fini ou infini des valeurs possibles pour l'observation. Le champ  $O = \{O(s) \in G_O, s \in S\}$  est une réalisation particulière du champ  $O$ . On notera  $\Omega_o = G_o^{n \times m}$  l'ensemble des réalisations possibles du champ d'observation.

Le champ de primitives ou étiquettes de l'image :

$E = \{E(s) \in G_E, s \in S\}$  est le champ d'étiquettes ou primitives associé à l'image.

$e = \{e(s) \in G_E, s \in S\}$  est une réalisation particulière du champ d'étiquettes.  $e(s)$  est l'étiquette associée au site  $s$  à valeurs dans  $G_E = \{a_1, a_2, \dots, a_p\}$ , ensemble discret et fini des valeurs possibles pour les étiquettes. L'ensemble des réalisations possibles du champ d'étiquettes sera noté  $\Omega_E = G_E^{n \times m}$

On utilisera indifféremment les notations  $E(s)$  ou  $E_S$  pour caractériser le champ d'étiquettes en l'occurrence au site  $s$  de la grille  $S$ .

### 1.2.2. Règle de Bayes

La règle de Bayes est à la base du formalisme. Soit le champ d'observations  $O$  issu des données observées et  $E$  le champ d'étiquettes que l'on cherche à estimer. La distribution de probabilités sur l'ensemble  $\Omega_E$  de toutes les réalisations possibles de  $E$  connaissant  $O=o$  s'appelle distribution a posteriori du champ à estimer et s'obtient par la formule de Bayes :

$$\Pr(E = e / O = o) = \frac{\Pr(E = e) \Pr(O = o / E = e)}{\Pr(O = o)} \quad (1.1)$$

L'application de la règle de Bayes implique d'une part (a) la possibilité de décrire statistiquement le processus d'observation par la distribution de probabilités  $\Pr(O/E)$ , et d'autre part (b) la connaissance de la distribution de probabilités a priori  $\Pr(E = e)$  du champ d'étiquettes. Si ces deux conditions sont remplies, on peut construire à partir de la distribution a posteriori plusieurs estimateurs du champ d'étiquettes recherché ; un estimateur  $\hat{e}$  est une application associant à toute réalisation du champ d'observation  $o$ , une configuration d'étiquettes estimée  $\hat{e}(o)$  :

$$\begin{aligned} \hat{e} : \Omega_O &\rightarrow \Omega_E \\ o &\rightarrow \hat{e}(o) \end{aligned} \quad (1.2)$$

La qualité d'un estimateur est définie par rapport à une distance  $C$  sur l'ensemble  $\Omega_E$ , aussi appelé fonction de coût. On appelle estimateur bayésien optimal relativement à la fonction de coût  $C$ , un estimateur  $\hat{e}$  pour lequel l'espérance mathématique de la fonction de coût (ou risque) est minimale (la moyenne prise sur toutes les

réalisations conjointes possibles de  $E$  et de  $O$  (le risque ainsi défini est une mesure du coût moyen "d'erreur" de l'estimateur) :

$$\begin{aligned} \text{Risque}(\hat{e}) &= \langle C(\hat{e}(O), E) \rangle_{\Omega_E \times \Omega_O} = \int_{o \in \Omega_O} \int_{c \in \Omega_E} C(\hat{e}(o), e) \cdot p_{EO}(o, e) \mu(de) \mu(do) \\ &= \int_{o \in \Omega_O} p_o(o) \mu(do) \int_{c \in \Omega_E} C(\hat{e}(o), e) \cdot p_{E/O}(e/O=o) \mu(de) \end{aligned}$$

Puisque la distribution de probabilité du champ  $O$  est positive, l'estimateur bayésien optimal doit minimiser le risque conditionnel pour chaque réalisation  $o \in O$  :

$$\forall o \in O, \hat{e}(o) = \arg \max_{\hat{e} \in \Omega_E} \int_{e \in \Omega_E} C(\hat{e}, e) \cdot p_{E/O}(e/O=o) \mu(de) \quad (1.3)$$

Le choix de l'estimateur dépend du problème considéré. Voici les estimateurs les plus couramment utilisés en traitement des images

### 1.2.2.1 L'estimateur du MAP

L'estimateur du Maximum à Posteriori MAP pénalise de manière identique toutes les estimations différentes de la vraie solution  $s$ . Ainsi, deux estimations qui diffèrent respectivement d'un site et de la moitié des sites de la solution exacte auront la même pénalisation. La fonction coût est dans ce cas :

$$\begin{aligned} \forall (a, b) \in \Omega_E \times \Omega_E \\ C(a, b) = 1 - \delta(a, b) \text{ avec } \delta(a, b) = \begin{cases} 1 & \text{si } a = b \\ 0 & \text{si non} \end{cases} \end{aligned} \quad (1.4)$$

L'estimateur résultant est donné par :

$$\hat{e}^{Map}(o) = \arg \max_{e \in \Omega_E} [pr(E = e/O = o)] \quad (1.5)$$

### 1.2.2.2 L'estimateur du MPM

L'estimateur du Mode des Marginales à Posteriori MPM pénalise une solution proportionnellement au nombre de sites erronés. La fonction coût qui lui est associée est :

$$C(a, b) = \sum_{s \in S} (1 - \delta(a_s, b_s)) \quad (1.6)$$

L'estimateur résultant est donné par :

$$\begin{aligned}\hat{e}^{MPM}(o) &= \{\hat{e}_s^{MPM}, s \in S\} \text{ avec } \forall s \in S : \\ \hat{e}_s^{MPM}(o) &= \arg \max_{g \in G_E} [\Pr(E_s = g / O = o)]\end{aligned}\quad (1.7)$$

### 1.2.2.3 L'estimateur du champ moyen

Cet estimateur comptabilise l'erreur quadratique en chaque site. Sa fonction de coût est de la forme :

$$C(a, b) = \sum_{s \in S} (a_s - b_s)^2 \quad (1.8)$$

Ce qui donne l'estimateur :

$$\begin{aligned}\hat{e}^{MF}(o) &= \{\hat{e}_s^{MF}(o), s \in S\} \text{ avec } \forall s \in S : \\ \hat{e}_s^{MF}(o) &= \langle E_s / O = o \rangle_{G_E}.\end{aligned}\quad (1.9)$$

### 1.2.3. Lien statistique entre observation et étiquettes

Le terme de probabilité conditionnelle des observations par rapport aux étiquettes  $\Pr(O = o / E = e)$  dépend de la modélisation du processus d'observation. Le champ observé est censé contenir toute l'information disponible sur l'image, l'observation est soit directement accessible, soit elle résulte d'un pré-traitement plus ou moins complexe de l'image originale. Il est courant de modéliser le lien statistique entre observation et étiquettes par la relation [10] :

$$\forall s \in S, \quad O(s) = \Phi[F_s(e^*), B(s)] \quad (1.10)$$

Où  $s$  appartient à un sous-ensemble de sites de l'image  $s \in A \subseteq S$ . Dans la formule,  $e^*$  Représente le champ d'étiquettes réel que l'on cherche à estimer, et  $O(s)$  est l'observation (Considérée comme une variable aléatoire) au site  $s$ . Dans le cas de la restauration d'images,  $e^*$  n'est autre que l'image réelle  $i \in \Omega_I$  avant d'être corrompue par le bruit, et  $o$  est l'image dégradée récupérée par le système d'imagerie. Le fait que les observations soient définies pour un sous-ensemble  $A$  de l'image permet de modéliser le cas de données incomplètes. La fonction  $F_s$  qui est typiquement non inversible représente par exemple la réponse impulsionnelle du système d'imagerie. La fonction  $\Phi$  modélise quant à elle la réponse du détecteur : il s'agit d'une fonction

inversible qui couple localement aux observations un bruit de nature aléatoire, noté  $B(s)$ , avant d'appliquer une transformation locale sans effet mémoire. La fonction  $\Phi$  peut être linéaire (cas des caméras CCD) ou non linéaire (logarithmique pour l'oeil ou la plaque photo). Le bruit peut être couplé aux mesures de manière additive ou multiplicative. Sa statistique conditionnera directement celle du terme de vraisemblance ; elle peut être, par exemple poissonnienne ou gaussienne, selon la nature du phénomène observé.

A partir de ce modèle, on peut remonter au terme de vraisemblance recherché  $\Pr(O/E)$ . On a tout d'abord

$$B(s) = \Phi^{-1}[F_s(e^*), O(s)] \quad (1.11)$$

Si le bruit est décorréolé spatialement (i.e. si les variables  $B(s)$ ,  $s \in S$  représentent des variables aléatoires indépendantes) alors on peut écrire directement le terme de vraisemblance :

$$\Pr(O = o / E = e) = \prod_{s \in A} \Pr[B(s) = \Phi^{-1}(F_s(e), o(s))] \quad (1.12)$$

Par la suite on préférera la forme exponentielle :

$$\Pr(O = o / E = e) = \frac{1}{Z} \exp \left[ - \sum_{s \in S} \Psi_s(e, o(s)) \right] \quad (1.13)$$

Où l'on a posé de façon formelle :

$$\Psi_s(e, o(s)) = -Z I_A(s) \ln \Pr[B(s) = \Phi^{-1}(F_s(e), o(s))] \quad (1.14)$$

( $Z$  est une constante arbitraire et  $I_A$  est la fonction caractéristique de  $A \subseteq S$ ). Cette formulation s'avérera utile par la suite car elle fait apparaître la fonction de coût ou énergie de contrainte aux données observées  $U_{obs}(e, o) = \sum_{s \in S} \Psi_s(e, o(s))$  lors de

l'intégration dans la formule de Bayes (1.1) :

$$\Pr[o = o / E = e] = \frac{1}{Z} \exp[-U_{obs}(e, o)] \quad (1.15)$$

**Exemple 1 :** Restauration d'images dégradées sans prise en compte de bords.

Pour fixer les idées, prenons l'exemple plutôt simple de la restauration d'images dégradées sans prise en compte des bords [10]. Dans ce cas, le champ

d'observation  $O$  s'identifie avec le champ de niveaux de gris de l'image issu d'un système d'acquisition (objectif et caméra CCD par exemple). Le champ d'étiquettes  $E$  à reconstruire est tout simplement l'image originale non dégradée.

La figure 1.1 représente les diverses transformations de l'image originale pour aboutir au champ d'observation  $O$ . Le processus d'observation est bien modélisé par la formule (1.10). Dans le cas très simple d'un système d'imagerie sans flou ( $F=l$ ) et d'une réponse parfaitement linéaire du capteur, l'équation (1.10) devient tout simplement :  $o = e + n$ . Si l'on suppose en outre que le bruit est décorrélé spatialement, et qu'il est sur chaque site gaussien, centré et de variance  $\sigma^2$ , alors sur l'image entière le terme de vraisemblance s'écrit :

$$\Pr(O = o / E = e) = \prod_{s \in S} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(o(s) - e(s))^2}{2\sigma^2}\right) \quad (1.16)$$

La fonction d'énergie à minimiser est alors (à une constante additive près) :

$$U_{obs}(e, o) = \sum_{s \in S} \Psi(o, e_s) = \frac{1}{2\sigma^2} \sum_{s \in S} [e(s) - o(s)]^2 \quad (1.17)$$

La fonction  $\Psi(o, e_s) = \frac{1}{2\sigma^2} (e(s) - o(s))^2$  peut être vue comme un potentiel de clique binaire liant le champ d'observation avec le champ d'étiquettes (voir remarque au § 1.4.1).

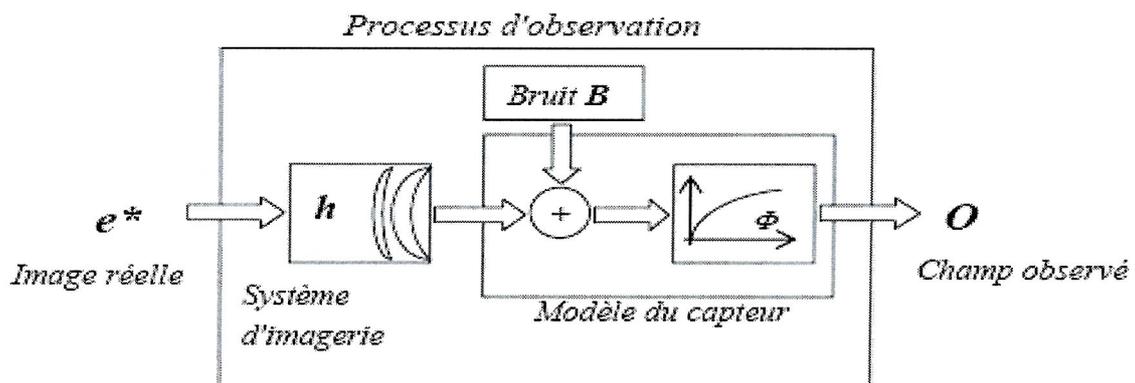


Figure 1.1 : Modèle du processus d'observation (acquisition d'image bruité)

#### 1.2.4. Probabilité a priori : modèle Markovien spatial

Si aucune information supplémentaire *a priori* n'est introduite,  $\Pr(E = e)$  est une constante par rapport à  $e$ , et la distribution *a posteriori* issue de la formule de Bayes s'identifie avec le terme de vraisemblance  $\Pr(O/E)$ . Le critère du MAP devient alors tout simplement le critère du Maximum de Vraisemblance (MV), puisque aucune contrainte n'est imposée quant à l'ensemble des solutions possibles du problème, la solution optimale correspond simplement à celle qui s'approche le mieux des données observées.

Il sera inutile par exemple de traiter par cette méthode la reconstruction d'images dégradées, car l'image observée est elle-même une solution acceptable du problème et qui en plus est la meilleure au sens du MV. Il faut donc introduire des contraintes *a priori* sur les solutions. Un champ de Markov en traitement d'images est un champ d'étiquettes à valeurs continues ou discrètes destiné à caractériser les informations sous-jacentes de l'image ou primitives (champ de vitesses dans le cas de l'estimation du mouvement, numéro de région dans le cas de la segmentation, etc.). La modélisation par champ de Markov rend compte des propriétés spatiales *a priori* des images observées. La propriété principale de la régularisation par champs de Markov est de conduire à des primitives homogènes par zones et automatiquement débarrassées du bruit [10]. D'un point de vue formel, la modélisation par champ de Markov présente deux intérêts majeurs :

- tout d'abord, il s'agit d'un modèle facilement paramétrable.
- comme on le verra au paragraphe suivant, à cette souplesse s'ajoute l'avantage d'une formulation immédiate du terme de probabilité conjointe du champ d'étiquettes  $\Pr(E)$  grâce au théorème de Hammersley-Clifford qui relie propriétés locales (marginales) et probabilités globales du champ.

Ce qui suit est destiné à rappeler la terminologie et les fondements de la théorie des champs des Markov aléatoires ou MRF (pour Markov Random Fields). Pour plus de détails, voir [2].

### 1.3. Champs de Markov

#### 1.3.1. Définitions

Le but de cette partie est de rappeler les principales définitions et propriétés relatives aux MRF utiles dans le cadre de la régularisation bayésienne. La caractéristique principale des MRF est de fournir un modèle probabiliste a priori du champ des primitives à estimer.

##### Définition

Un champ aléatoire  $E$  est un champ de Markov relativement à un voisinage  $V$  si et seulement si la propriété suivante est vérifiée :

$$Pr[E(s) = e(s) / E(r) = e(r), \forall r \in S, s \neq r] = pr[E(s) = e(s) / E(r) = e(r), r \in V(s)]$$

Ainsi, les propriétés statistiques locales d'un champ de Markov au site  $s$  ne dépendent que de l'état d'un voisinage plus ou moins étendu défini autour de ce site et non de l'état du champ tout entier. En traitement d'images bas - niveau, on est en général amené à se contenter de considérer des voisinages d'ordre 1 (4 voisins), d'ordre 2 (8 voisins) ou d'ordre 3 (12 voisins), voir figure 1.2

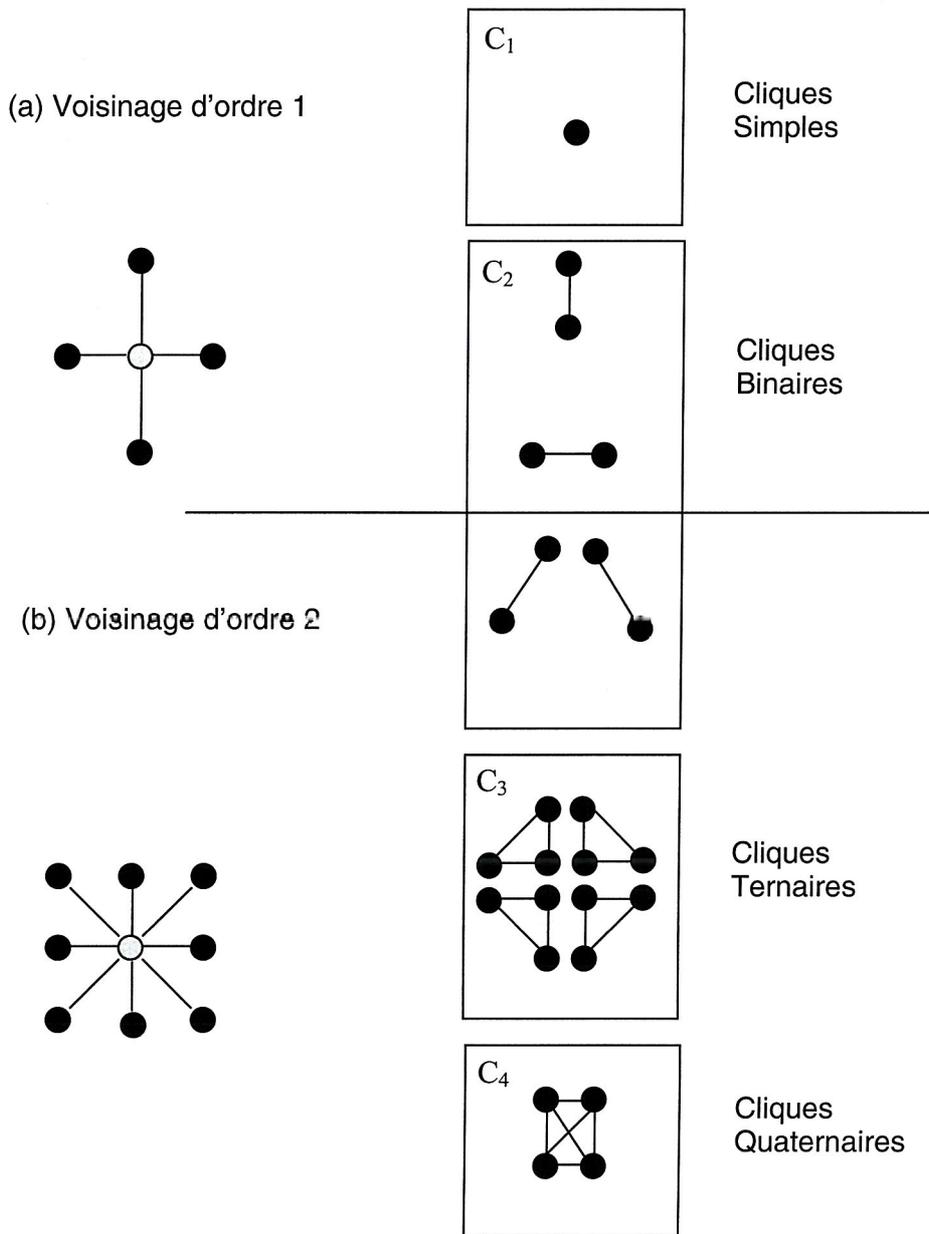


Figure 1.2 : Voisinages (a) d'ordre 1 et (b) d'ordre 2 et cliques correspondantes.

### 1.3.2. Théorème d'équivalence

La définition du champ de Markov à partir des seules distributions marginales ne semble offrir aucune issue permettant d'aboutir à la distribution conjointe recherchée (ou terme d'a priori)  $\Pr(E)$ . C'est pourquoi l'utilisation des champs de Markov en traitement d'images n'a été effective qu'à la découverte du théorème de Hammersley-Clifford qui établit l'équivalence entre champ de Markov et distribution de Gibbs :

**Théorème de Hammersley-Clifford** (ou théorème d'équivalence). Le théorème de Hammersley-Clifford [5] stipule que  $E$  est un champ de Markov si et seulement si la distribution de probabilités globale du champ  $\Pr[E = e]$  peut se mettre sous la forme Gibbsienne suivante :

$$\Pr[E = e] = \frac{1}{Z} \exp(-U_{reg}(e)) \quad (1.18)$$

avec  $Z = \sum_{e \in \Omega_E} \exp(-U_{reg}(e))$ , et  $U_{reg} = \sum_{c \in C} V_c(e)$ , où  $C$  désigne l'ensemble des cliques définie sur la grille  $S$  et  $V_c$  est une fonction de potentiel associé à la clique  $c \in C$ . Par la suite, la fonction  $U_{reg}$  sera indistinctement appelée *énergie spatiale*, *énergie a priori*, *énergie de régularisation* ou encore de *lissage spatial*. Elle ne dépend que des variables du champ à estimer et non du champ d'observation.

Un champ de Markov peut donc être défini en spécifiant les potentiels des cliques  $V_c$ ; les probabilités conditionnelles locales du champ s'écrivent alors :

$$\Pr[E(s) = e(s) / E(r) = e(r), \forall r \in S, r \neq s] = \frac{1}{Z_s} \exp[-U_s(e)] \quad (1.19)$$

Où  $Z_s = \sum_{g \in G_E} \exp(-U_s(e)) \Big|_{e(s)=g}$  et  $U_s(e) = \sum_{c \in C: s \in c} V_c(e)$ . (1.20)

La définition des potentiels élémentaires sera spécifique au problème à résoudre ; la force de l'approche markovienne réside justement dans le fait qu'il est beaucoup plus aisé de décrire et interpréter les interactions par le biais d'un potentiel que par l'écriture des distributions conditionnelles locales. De plus, la seule contrainte à respecter est que chaque potentiel  $V_c$  ne dépend que des pixels de la clique correspondante.

Ainsi, le terme a priori  $\Pr(E)$  que l'on cherchait pour compléter la règle de Bayes est tout simplement explicitée par la distribution de Gibbs de (1.18) ci-dessus.

Suite1 (exemple1) :

Si l'on s'intéresse toujours à la restitution d'images bruitées et si l'on modélise l'image par un MRF défini par rapport à un voisinage d'ordre 1, alors il est classique de modéliser la contrainte d'homogénéité spatiale de l'image par un potentiel portant sur les cliques binaires de la forme :

$$V_c(e) = V_c[e(s), e(r)] = \beta_s \cdot (e(s) - e(r))^2 \text{ avec } \beta_s > 0 \quad (1.21)$$

où  $c$  est la clique binaire comprenant les sites voisins  $s$  et  $r$ . Les configurations localement favorisées sont celles pour lesquelles les sites ont le même niveau de gris [17], [8], [13]. La forme quadratique du potentiel spatial ressemble dans ce cas à celle du potentiel de contrainte  $\Psi$  trouvé lors de la modélisation du processus d'observation (dans le cas de variables binaires, le potentiel ainsi défini devient un potentiel à deux niveaux 0,  $\beta_s$  symétrisé en  $\pm\beta_s$  par addition d'une constante). L'énergie spatiale (ou terme de régularisation, ou encore de lissage spatial) s'écrit finalement :

$$U_{reg}(e) = \sum_{c \in C_2} V_c(e) = \frac{\beta_s}{2} \sum_{s \in S} \sum_{r \in V_s} (e(s) - e(r))^2 \quad (1.22)$$

La valeur du paramètre  $\beta_s$  dépend des caractéristiques spatiales de l'image réelle non bruitée, le paramètre sera plutôt faible pour une image ayant une forte granularité ou des textures fines, et plutôt fort pour une image homogène. En fait  $\beta_s$  peut être estimé pour une image et un système de voisinages donnés, à condition de définir un critère d'erreur (de reconstruction) [16].

### 1.3.3. Le problème de l'estimation des paramètres

La formulation correcte d'un problème particulier en termes de fonction d'énergie à minimiser passe par le choix d'un certain nombre de paramètres définissant aussi bien les potentiels des cliques spatiales que le modèle d'observation. Dans la majorité des cas, ces paramètres sont fixés empiriquement ;

ils peuvent être déterminés une fois pour toutes ou être réactualisés périodiquement en fonction des nouvelles images acquises.

## 1.4. Estimateur MAP et minimisation de l'énergie

### 1.4.1 Construction de la fonction de coût

En remplaçant dans la formule de Bayes (1.1) le terme de vraisemblance par son expression dans (1.15), et le terme a priori par la formule de (1.18), on obtient finalement l'expression de la probabilité a posteriori :

$$\Pr(E = e / O = o) = \frac{1}{\Pr(O = o)} \cdot \frac{1}{Z} \cdot \exp[-U_{obs}(e, o) + U_{org}(e)] \quad (1.23)$$

L'introduction de la modélisation markovienne dans le cadre de l'estimation bayésienne permet, comme annoncé, d'avoir recours à des distributions gibbsiennes. Le critère du MAP, c'est à dire la maximisation de la probabilité a posteriori par rapport au champ des labels  $e$  étant donné un champ d'observations  $o$ , est alors équivalent à la minimisation d'une fonction de coût ou énergie totale comprenant deux termes :

$$U(e, o) = U_{obs}(e, o) + U_{reg}(e) \quad (1.24)$$

- le premier terme est un terme de rappel aux données qui provient du modèle du processus d'observation ;
- Le deuxième terme est un terme de régularisation qui intègre les connaissances a priori dans le modèle (champ de Markov).

La fonction d'énergie est une fonction à grand nombre de variables (i.e. le champ d'étiquettes). La nature des interactions et du processus d'observation peut conduire à des fonctions qui ne sont ni quadratiques ni convexes par rapport aux variables ; l'énergie peut comporter alors plusieurs minimums locaux. Par contre, dans le cas d'une fonction quadratique et convexe, on est assuré de l'existence d'un minimum global. Ce minimum peut être obtenu par des méthodes d'optimisation déterministes classiques de type descente de gradient (plus forte pente à pas optimal, gradients conjugués, méthode de Newton).

Remarque : La distribution de probabilités conjointe du champ d'étiquettes et du champ d'observation  $\Pr[E = e, O = o]$  s'écrit :

$$\Pr(E = e, O = o) = \frac{1}{Z} \cdot \exp[-U(e, o)] \quad (1.25)$$

Or, la fonction d'énergie globale  $U(e, o)$  peut s'écrire comme une somme de potentiels de cliques "généralisées"  $c \in C_X$  reliant le champ d'étiquettes et le champ d'observation :

$$U(e, o) = \sum_{c \in C_X} V_c(e, o) = \sum_{s \in S} V_s(e, o) + \sum_{c \in C} V_c(e). \quad (1.26)$$

Où  $V_s(e, o) = \Psi_s(e, o(s))$ . Autrement dit, et en vertu du théorème de Hammersley-Clifford, le champ couplé  $X = \{E, O\}$  est un champ de Markov.

Suite2 (exemple1) : Si l'on remplace dans la formule de Bayes le terme de vraisemblance et le terme d'a priori par les expressions correspondant au problème de la restauration d'images, alors l'énergie à minimiser prend la forme finale :

$$U(e, o) = \frac{1}{2\sigma^2} \sum_{s \in S} (e(s) - o(s))^2 + \frac{\beta}{2} \sum_{s \in S} \sum_{k \in \mathcal{V}(s)} (e(s) - e(k))^2. \quad (1.27)$$

Que l'on a coutume d'exprimer de la façon suivante :

$$U(e, o) = \sum_{s \in S} (e(s) - o(s))^2 + \lambda \sum_{s \in S} \sum_{k \in \mathcal{V}(s)} (e(s) - e(k))^2 \quad (1.28)$$

Le facteur  $\lambda$  permet de pondérer le terme de rappel aux données par rapport au terme de régularisation [7].

### 1.4.2. Remarque sur l'estimateur MPM et MF

L'estimation au sens du MPM ou du MF n'est pas immédiate, pour le premier, il manque les distributions marginales a posteriori, qui ne sont pas directement accessibles en général. La difficulté du MF provient aussi du fait que son calcul passe par la connaissance de la fonction de partition de la distribution a posteriori. Cependant, en raison du caractère moins brutal de la fonction de coût employée (La fonction de coût associé à l'estimateur MAP pénalise de la même façon toutes les configurations non optimales, alors que la fonction de coût du MPM pénalise chaque configuration par rapport au nombre de pixels différents de la configuration optimale.), l'estimateur MPM semble se comporter mieux que le MAP lorsque le processus de formation des observations introduit un bruit important [19]. En effet, le noyau de l'algorithme d'optimisation stochastique utilisé pour résoudre la minimisation de la fonction d'énergie (dans le cadre de l'estimateur MAP) nécessite la mise en place d'échantillonneurs de Metropolis ; alors que les estimés du MPM ou du MF peuvent consister à estimer les propriétés statistiques a posteriori de l'image à partir d'un large échantillon d'images (loi des grands nombres) issu d'une procédure d'échantillonnage. Ainsi, l'estimé du MF est tout simplement approché par la moyenne de tous les échantillons, et l'estimé du MPM (en chaque site) est approché par l'étiquette apparaissant le plus fréquemment dans les échantillons. Nous n'avons pas exploré ces possibilités ; seul l'estimateur du MAP a retenu toute notre attention, la raison principale étant la formulation mathématique rigoureuse du problème comme un problème d'optimisation.

## 1.5. Débruitage d'images binaires et modèle d'Ising

### 1.5.1. Définition de la fonction de coût ou énergie

Nous avons déjà présenté le problème de la restauration d'images à niveaux de gris. Intéressons nous maintenant au cas de la restauration d'images binaires. Supposons par ailleurs que le champ observé est lui-même binaire. Alors, la modélisation la plus simple du processus d'observation fait intervenir non pas un bruit gaussien, mais un bruit de canal de taux d'erreur  $\varepsilon$ , défini selon :

$$\Pr[O(s) = 1 / E(s) = 0] = \Pr[O(s) = 0 / E(s) = 1] = \frac{\varepsilon}{2} \quad (1.29)$$

$$\Pr[O(s) = 0 / E(s) = 0] = \Pr[O(s) = 1 / E(s) = 1] = \frac{1-\varepsilon}{2} \quad (1.30)$$

En notant  $\beta_0 = \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$ , on peut exprimer le terme de vraisemblance marginal par :

$$\Pr[O(s) = o(s) / E(s) = e(s)] = \frac{\varepsilon}{2} \exp[\beta_0 \cdot \delta(o(s) - e(s))] \quad (1.31)$$

Avec  $\delta(0) = 1$  et 0 sinon. La vraisemblance de l'observation  $O=o$  est alors :

$$\Pr(O = o / E = e) = \left(\frac{\varepsilon}{2}\right)^{n \times m} \exp\left(\beta_0 \sum_{s \in S} \delta(o(s) - e(s))\right) \quad (1.32)$$

L'énergie de rappel à l'observation est enfin (à une constante additive près, ce qui ne change pas la solution finale résultat de la recherche du minimum énergétique) :

$$U_{obs} = -\beta_0 \sum_{s \in S} \delta(o(s) - e(s)) \quad (1.33)$$

Soit finalement :

$$U_{obs} = -\frac{\beta_0}{2} \sum_{s \in S} (2o(s) - 1) \cdot (2e(s) - 1). \quad (1.34)$$

Par ailleurs, si l'on choisit un potentiel de régularisation spatial à deux niveaux  $\pm \beta_s$  :

$$\forall (r, s) \in S \times S, V_c(e(s), e(r)) = -\beta_s \cdot (2e(s) - 1) \cdot (2e(r) - 1) \quad (1.35)$$

Alors l'énergie de régularisation spatiale s'écrit :

$$U_{reg}(e) = -\frac{\beta_s}{2} \sum_{s \in S} \sum_{r \in V_s} (2.e(s) - 1).(2.e(r) - 1) \quad (1.36)$$

L'énergie *totale* à minimiser pour la restauration d'images binaires bruitées est finalement :

$$U(e, o) = -\beta_s \sum_{s \in S} \sum_{r \in V_s} (2e(s) - 1).(2e(r) - 1) - \beta_0 \sum_{s \in S} (2o(s) - 1).(2e(s) - 1) \quad (1.37)$$

On reconnaît dans le premier terme de la formule le Hamiltonien d'un système de spins en interaction mutuelle ou modèle d'Ising bidimensionnel [14]. Dans ce cas précis les interactions sont toutes positives (le coefficient de couplage vaut  $J_{rs} = \beta_s$ ) ce qui correspond au cas des matériaux ferromagnétiques. La configuration d'énergie minimale Pour le modèle ferromagnétique ou anti-ferromagnétique est connue d'avance il s'agit d'un champ d'étiquettes uniforme (+1 ou 0) pour le cas ferromagnétique et d'un champ en damier (direct et inverse) dans le cas antiferromagnétique. Ceci est possible car il n'y a pas de phénomène de "frustration" (autrement dit, les configurations d'énergie minimale locale sont compatibles avec la configuration d'énergie minimale globale). Le phénomène de frustration apparaît par exemple quand tous les coefficients de couplage n'ont pas le même signe (Verre de Spin), ou quand on couple le système ferromagnétique avec un champ magnétique externe non uniforme. L'image à restaurer  $O=o$  joue ici le rôle de ce champ magnétique externe non uniforme (deuxième terme de la formule). Expliciter alors la configuration du minimum énergétique est un tout autre problème très difficile.

### 1.5.2. Choix des paramètres $\beta_S$ et $\beta_O$

Si l'on définit la "granularité" (d'ordre 1) d'une image comme étant le rapport du nombre de paires de pixels voisins ayant la même étiquette au nombre total de pixels de l'image, alors on constate que pour des images générées selon la loi d'a priori du modèle, la granularité décroît avec  $\beta_S$  en suivant une courbe qui rappelle celle de la magnétisation d'un matériau ferromagnétique en fonction de la température. En particulier il existe une valeur critique qui fait chuter brutalement la granularité de l'image en faisant apparaître de domaines uniformes. Ce n'est pas par hasard : la température critique correspondant à une transition de phase pour l'Ising ferromagnétique bidimensionnel (défini par l'Hamiltonien du premier terme de (1.30) vaut à peu près  $T_{crit} \sim v \cdot \beta_S$  en utilisant l'approximation du champ moyen (ce qui implique en principe que le cardinal du voisinage soit  $v \gg 1$ ) ; dans le cas d'un voisinage 4, cette température peut être calculé exactement [9] : elle vaut  $T_c = 0,567 \cdot T_{cm} \approx 2,27 \cdot \beta_S$ . Or, la loi d'a priori du modèle de Markov  $\Pr(E)$  correspond à la distribution de Gibbs du modèle d'Ising pour  $\beta_S' = \beta_S/T$ . Il n'est donc pas étonnant que la granularité chute quand

$\beta_S' = \beta_S/T_c = 1/2,27 = 0,44$ . Prendre  $\beta_S$  en dessus de cette valeur signifie que les images à modéliser présentent principalement des régions homogènes ; inversement, choisir un paramètre  $\beta_S$  au dessous de cette valeur signifie que les images à traiter présentent naturellement une granularité élevée.

A un facteur multiplicatif près (ce qui n'a pas de conséquence sur le résultat de l'optimisation), la fonction de coût ne dépend en réalité que d'un seul paramètre, le Rapport  $\lambda = \beta_S / \beta_O$  :

$$U(e, o) = - \sum_{s \in S} (2o(s) - 1) - \lambda \sum_{s \in S} \sum_{r \in V_s} (2e(s) - 1) \cdot (2e(r) - 1) \quad (1.38)$$

Le facteur  $\lambda$  traduit la confiance (ou plutôt la méfiance) que l'on peut faire au champ de données issu de l'observation : si le bruit de canal est nul l'observation correspond exactement à l'image originelle, alors le facteur  $\lambda$  est lui aussi nul

( $\beta_O \rightarrow \infty$ ) et il n'y aura pas d'effet de régularisation : le minimum de l'énergie correspond tout simplement à l'image observée. Si par contre le bruit atteint 50% (le champ d'observation ne contient alors aucune information pertinente,  $\beta_O \rightarrow 0$ ), le terme de contrainte est négligeable, traduisant le fait que seule la connaissance a

priori que l'on a sur les images est à prendre en considération (le résultat de l'optimisation correspond à une image entièrement).

Remarque : Si l'on choisit d'utiliser l'estimateur MPM, alors le résultat de l'optimisation en présence d'un bruit important correspond à une image aléatoire de granularité correspondant à celle de l'image que l'on aurait voulu reconstruire (à condition que le paramètre  $\beta_S$  ait été bien estimé au préalable). C'est pourquoi à notre avis l'estimateur du MPM semble se comporter mieux en présence d'un bruit important.

Enfin, le facteur  $\lambda$  peut être estimé de façon plus ou moins heuristique (cf.§3.3).

Nous avons testé deux façons de le faire :

1. En imposant des contraintes sur le résultat de l'optimisation (c'est une version simplifiée de la technique d'Azencott [3]) L'idée consiste à rendre improbables certaines configurations locales ; ainsi pour un voisinage d'ordre 2, on voudra éviter qu'un pixel "s'aligne" avec la donnée de l'observation locale, si plus de 4 pixels du voisinage lui "indiquent" le contraire. Ceci se traduit exactement par l'inéquation (à condition d'utiliser l'estimateur du MPM) :

$$(+5.\beta_S - 3.\beta_S) > \beta_0, \text{ c.à.d. : } \lambda > 1/2.$$

2. Plus simplement on a tenté d'ajuster "à l'oeil" le paramètre  $\lambda$  pour un type spécifique d'images données.

## 1.6. Conclusion

Dans cette partie nous avons présenté une approche probabiliste générale permettant de régulariser les problèmes mal posés rencontrés en vision bas-niveau. Ceci est possible grâce à l'utilisation d'estimateurs bayésiens, dont par exemple l'estimateur du maximum a posteriori (MAP). Nous avons eu l'occasion de voir comment l'utilisation conjointe d'un modèle de Markov (MRF) chargé de représenter le champ de primitives, et d'un modèle probabiliste approprié du processus d'observation, permettait d'explicitier les termes de la formule de Bayes. Nous avons montré ensuite que l'estimation au sens du MAP est équivalente à la recherche de l'argument minimisant une fonction de coût ou énergie spécifique au problème, fonction que l'on a pu exprimer comme étant la somme de deux termes exerçant des influences "complémentaires" sur le résultat de l'optimisation (terme de contrainte aux données et terme de lissage spatial).

L'objet de la partie suivante sera donc de présenter la technique dite du recuit simulé, méthode d'optimisation stochastique permettant en principe de venir à bout de n'importe quel problème d'optimisation difficile à condition toutefois d'avoir suffisamment de temps disponible.

# Partie 2

## 1.7. Recuit Simulé

### 1.7.1. Introduction

Nous avons vu à la partie précédente comment l'estimation bayésienne associée à une modélisation markovienne du champ des primitives permet d'exprimer formellement un certain nombre de problèmes en analyse d'images bas niveau comme un problème d'optimisation, consistant à minimiser une fonction d'énergie ou coût. Cette énergie intègre un terme de régularisation qui exprime les connaissances a priori que l'on a sur la solution - telle que l'homogénéité spatiale de l'image et un terme de rappel aux données qui résulte de la modélisation probabiliste du processus d'observation

La formulation d'un grand nombre de problèmes en traitement d'images bas niveau gagne ainsi en rigueur mathématique, mais on n'est pas pour autant beaucoup plus près de leur résolution pratique. En effet, le calcul du minimum de la fonction d'énergie ne peut pas être réalisé par une recherche exhaustive sur l'ensemble de solutions possibles : le cardinal de cet ensemble est immensément grand (égal à  $\Omega=L^{n \times m}$ , avec en général  $L$ ,  $n$  et  $m$  de l'ordre de 256). Par ailleurs, la fonction d'énergie n'est en général ni quadratique ni convexe par rapport aux variables du problème et la recherche du minimum global est loin d'être un problème trivial d'un point de vue analytique. Ceci est vrai aussi des fonctions dont les variables prennent des valeurs discrètes (niveaux de gris discrets) dans ce cas, même des fonctions qui seraient quadratiques et convexes pour des variables continues présentent non pas un seul minimum global, mais un nombre considérable de minima locaux

La minimisation de ce genre de fonctions est en fait un problème d'optimisation combinatoire difficile, Les algorithmes stochastiques apportent une réponse à ces ardues problèmes d'optimisation.

### 1.7.2. Les algorithmes d'optimisation : généralités

Nous allons présenter par la suite les fondements théoriques des procédures de relaxation stochastique. Le but final est de bien saisir les raisons qui nous ont conduit à adopter l'algorithme stochastique dit de *recuit simulé* (Simulated Annealing) pour notre application, plutôt qu'un autre algorithme, de type déterministe par exemple ; il en existe en effet deux sortes d'algorithmes :

- Les algorithmes déterministes : Ils sont sous-optimaux d'un point de vue théorique, c'est-à-dire que la convergence de la solution n'est garantie que vers le minimum local le plus proche de la configuration initiale des variables. La solution est donc extrêmement sensible au choix de la configuration initiale. Mais ils ont à leur avantage une rapidité bien plus importante que celle des algorithmes stochastiques.
- Les algorithmes de relaxation stochastiques : de type recuit simulé ([10], [15]) sont théoriquement les meilleurs car la convergence vers un minimum global de la fonction d'énergie est théoriquement assuré, quelle que soit la configuration de départ. Mais il s'agit bien d'une preuve théorique de convergence : le minimum global de la fonction d'énergie n'est vraiment atteint qu'au prix d'un temps de calcul infini... Ils ne sont pas pour autant inutilisables : en pratique, les méthodes stochastiques "tronquées" et modifiées, bien que toujours assez lentes, permettent de trouver des solutions bien meilleures que celles proposées par les techniques déterministes. De plus, la qualité de la solution est indépendante de la configuration initiale des variables.

## 1.8. Algorithmes stochastiques

### 1.8.1. Introduction

Les algorithmes de relaxation stochastique de type recuit simulé partent d'une configuration initiale donnée du champ d'étiquettes et s'arrangent ensuite pour générer à partir de celle-ci une suite qui s'approche plus ou moins rapidement du mode ou état fondamental de la fonction d'énergie  $U$  à minimiser. En termes plus formels, ces algorithmes, qui s'inspirent fortement de la physique statistique, peuvent être vus comme la combinaison de deux procédures que l'on détaillera par la suite :

**a-** Une procédure dite d'échantillonnage dynamique visant à atteindre l'équilibre thermodynamique des variables du champ d'étiquettes à chaque palier de température. Plus précisément encore, la procédure d'échantillonnage vise la distribution de Gibbs à température  $T$  :

$$G_T(E = e) = \frac{1}{Z_T} \exp\left(-\frac{U(e)}{T}\right) \quad (1.39)$$

**b-** Une procédure (ou schéma) de refroidissement plus ou moins optimale, le recuit simulé à proprement parler, cette expression désigne une loi de décroissance continue ou par paliers de la température permettant d'atteindre une assez bonne solution du problème en un temps raisonnable.

L'idée sous-jacente est que si pour toute température l'équilibre thermodynamique peut être atteint grâce à la procédure d'échantillonnage (a), alors à très basse température seules les configurations d'énergie minimale du système auront une probabilité non négligeable d'apparaître (la distribution de Gibbs favorise toujours les configurations qui minimisent la fonction d'énergie ; à la limite de température nulle, seuls les modes fondamentaux du système ont une probabilité non nulle d'apparaître). Malheureusement il est très coûteux « voire impossible » de générer directement un échantillon suivant la distribution de Gibbs à très basse température. On procède alors de façon progressive, par recuit simulé (b). L'idée, comme on l'a dit, est directement inspirée de la physique (et de la métallurgie) : lorsqu'un matériau cristallin est recuit lentement, il est possible d'obtenir des états très ordonnés qui correspondent à des énergie proches de l'état fondamental. Au contraire, si le refroidissement est brutal, alors il est très vraisemblable que le cristal présente des défauts et que son énergie potentielle soit piégée dans un minimum

local. Nous allons par la suite discuter en détail des points (a) et (b), dans le §1.8.2 et §1.10 respectivement.

### 1.8.2. Les procédures d'échantillonnage dynamiques

Les méthodes d'échantillonnage ou méthodes de Monte Carlo [11] s'attaquent au problème suivant : étant donné une mesure de probabilité  $\pi$  dans l'espace des configurations d'étiquettes  $\Omega_E$ , comment générer des échantillons  $e \in \Omega_E$  (réalisations du champ d'étiquettes) suivant cette loi ?

Les méthodes de Monte Carlo peuvent être rangées dans deux catégories : les méthodes statiques et les méthodes dynamiques. Les premières consistent à générer une séquence d'échantillons statistiquement indépendants de la loi recherchée. Elles sont pourtant inadaptées à la plupart de cas pratiques rencontrés en physique par exemple (systèmes à très grand nombre de variables couplées suivant la distribution de Gibbs).

Les méthodes de Monte Carlo dynamiques procèdent de façon plus subtile et à l'aide d'un processus stochastique (n'ayant pas nécessairement d'équivalent réel) est inventé pour le besoin du problème, de telle sorte que seule la distribution  $\pi$  recherchée en soit la loi d'équilibre correspondante. Le processus stochastique est simulé sur ordinateur en partant d'une configuration aléatoire : au fur et à mesure de la simulation, le système tend naturellement vers son état d'équilibre. En général, le processus stochastique choisi est une chaîne de Markov (à ne pas confondre avec le champ de Markov vu dans la partie précédente, qui, lui, rend compte des propriétés spatiales du champ d'étiquettes).

#### 1.8.2.1 Chaînes de Markov. Théorème d'échantillonnage dynamique

Un théorème très important montre que si le processus stochastique choisi est une chaîne de Markov homogène, irréductible et apériodique, admettant une distribution stationnaire  $\pi$ , alors la chaîne converge naturellement vers cette distribution quelle que soit la configuration initiale. Précisons ces concepts par un bref rappel des propriétés des chaînes de Markov (pour plus de détails, voir [10]).

**Définition 1 :**

Une chaîne de Markov discrète dans l'espace de configurations  $\Omega_E$ , est une suite de variables aléatoires  $\{E_0, E_1, E_2, \dots\}$  à valeurs dans  $\Omega_E$  telle que la probabilité d'avoir  $E_{n+1} = e_{n+1}$  ne dépend que de la valeur du champ à l'instant précédant  $E_n = e_n$ .

Plus précisément, une chaîne de Markov est définie par :

- Une distribution de probabilités initiale

$$p^0(e) = \Pr(E_0 = e) = p^0(e), \forall e \in \Omega_E \quad (1.40)$$

- Une suite de matrices de transition

$$P^n = (P_{xy}^n) \quad P^n = (P_{xy}^n) \quad (1.41)$$

telles que  $P_{xy}^n = \Pr(E_{n+1} = y / E_n = x)$ , où  $x$  et  $y$  appartiennent à  $\Omega_E$ .

Les éléments de la matrice de transition  $P^n$  (appelés probabilités de transition ou de passage) sont positifs ou nuls, et l'on a nécessairement pour toute configuration de départ  $x \in \Omega_x$  la relation :

$\sum_{y \in \Omega_E} P_{xy}^n = 1$  (on dit que la matrice  $P^n$  est stochastique).

**Définition 2 :**

Une chaîne de Markov est dite homogène si la matrice de transition  $P^n$  est indépendante de  $n$  (elle sera noté simplement  $P$ ). Nous allons nous intéresser uniquement aux chaînes de Markov homogènes. La matrice de transition gère l'évolution de la distribution de probabilités du champ (noté  $P^n$  à l'étape  $n$ ) ; ainsi par exemple, la distribution de probabilités à l'étape  $n+1$  est liée à celle de l'étape  $n$  par la relation :

$$\forall y \in \Omega_E, \Pr(E_{n+1} = y) = \sum_{x \in \Omega_x} \Pr(E_n = x) \cdot p_{xy} \quad (1.42)$$

L'expression matricielle équivalente est :  $P^{n+1} = P P^n$ , (on considère ici  $P^n$  comme un vecteur à valeurs réelles de dimension égale au cardinal de  $\Omega_E$ ). On définit également la probabilité de passage d'une configuration à une autre en un nombre  $k$  de pas par :

$$p_{xy}^{(k)} = \Pr(E_{n+k} = y / E_n = x) \quad (1.43)$$

La matrice de transition ou de passage correspondante est  $P^{(k)} = P^k$  ; la distribution de probabilités du champ à l'étape  $n+k$  est lié à celle de l'étape  $n$  selon :

$$P^{n+k} = P^k P^n .$$

Définition 3 :

Une distribution de probabilité  $\pi$  sera dite stationnaire pour la chaîne considérée si et seulement si :

$$\forall y \in \Omega_E \quad \pi(y) = \sum_{x \in \Omega_x} \pi(x) \cdot p_{xy} \quad (1.44)$$

Matriciellement  $\pi = P \cdot \pi$ , c'est à dire que la distribution stationnaire est vecteur propre de la matrice de transition.

Remarque : Une chaîne de Markov est dite stationnaire si les probabilités a priori  $p^n$  des états à l'instant  $n$  sont indépendants de  $n$ . Soulignons alors que l'homogénéité et la stationnarité d'un processus stochastique sont deux caractères distincts : le processus peut avoir une évolution régulière (et donc être homogène) et être tel que cette évolution l'éloigne de sa configuration initiale (le processus n'est pas stationnaire).

Définition 4 :

Une distribution  $\pi$  sera a fortiori stationnaire si elle vérifie la condition de bilan détaillé :

$$\forall (x, y) \in \Omega_E \times \Omega_E \quad \pi(x) p_{xy} = \pi(y) p_{yx} \quad (1.45)$$

Remarque : La condition de bilan détaillé traduit la symétrie par inversion du temps dans la plupart des systèmes physiques (réversibilité macroscopique). Dans un tel système à l'équilibre thermodynamique, la transition de  $x$  vers  $y$  se fait, en moyenne, le même nombre de fois que de  $y$  vers  $x$  (on peut dire aussi en appliquant l'hypothèse d'ergodicité que dans un ensemble composé d'un grand nombre de systèmes identiques à l'équilibre, le nombre moyen de systèmes effectuant la transition de  $x$  vers  $y$  à l'instant  $t$  est égal à celui effectuant la transition inverse). Cette condition implique bien la stationnarité des grandeurs moyennes du système, dont en particulier l'énergie.

Définition 5 :

Une chaîne de Markov est dite irréductible s'il est toujours potentiellement possible de passer d'une configuration du système à une autre en un nombre fini d'étapes :

$$\forall (x, y) \in \Omega_E \times \Omega_E \quad \exists n \in \mathbb{N}^* : p_{xy}^n > 0 \quad (1.46)$$

Définition 6 :

On définit la période de la configuration particulière  $x \in \Omega_E$  comme étant le plus grand commun dénominateur (PGCD) des nombres  $n$  tels que  $p_{xx}^{(n)} > 0$ ; on dit que la chaîne de Markov est apériodique si elle n'admet pas de période quelle qu'elle soit la configuration choisie (i.e. pour tout  $x \in \Omega_E$ , la période est égale à 1). En particulier, une chaîne de Markov admettant les transitions nulles ( $p_{xx}^{(1)} > 0$ ) est apériodique.

On est enfin en mesure d'énoncer le théorème fondamental qui va nous permettre de construire un algorithme d'échantillonnage dynamique :

Théorème de "stationnarisation" Soit  $\{E_n\}_{n \in \mathbb{N}}$  une chaîne de Markov homogène, irréductible et apériodique. Si elle admet une distribution stationnaire  $\pi$ , alors celle-ci est non seulement unique mais de plus la distribution de probabilité de la chaîne converge vers la distribution d'équilibre  $\pi$ , et cela indépendamment de la distribution initiale [22] :

$$\forall x \in \Omega_E, \lim_{n \rightarrow \infty} p_{xy}^{(n)} = \pi(y) \quad (1.47)$$

(la chaîne de Markov se "stationnarise" de façon asymptotique sur la distribution  $\pi$ .)

Remarquons que la distribution initiale peut être de la forme :  $p^0(E = e) = \delta(e = e_0)$ , c.à.d. que l'on peut initialiser la chaîne avec une configuration bien déterminé au départ. Alors le théorème ci-dessus prouve que la distribution de probabilité à l'instant  $n$  converge vers  $\pi$  indépendamment du choix de la configuration initiale.

On voit alors comment il faut procéder pour obtenir un échantillonneur dynamique de Monte Carlo : il suffit de construire, à partir de la distribution  $\pi$  recherchée, une matrice de transition  $\pi$  satisfaisant aux conditions du théorème. La

simulation du comportement dynamique de la chaîne ne présente ensuite aucune difficulté de principe.

On présente dans ce qui suit deux techniques permettant de construire une matrice de transition  $P$  satisfaisant les conditions du théorème. La première donne lieu à la procédure d'échantillonnage dynamique de Metropolis ; la seconde à la procédure d'échantillonnage de Gibbs.

### 1.8.2.2 Echantillonneur de Metropolis (ou dynamique de Metropolis)

Une méthode très générale pour construire des matrices de transition satisfaisant aux conditions du théorème a été esquissée originellement par [18] dans le cadre de l'étude des propriétés thermodynamiques des systèmes à plusieurs particules.

Condition de stationnarité et d'homogénéité. Dans cette méthode, la stationnarité de la distribution est assurée par la condition suffisante quoique plus restrictive du bilan détaillé. Cette condition est néanmoins plus facile à vérifier : la méthode de Metropolis- Hastings [12] est une technique générale permettant de s'y tenir à partir de n'importe quelle matrice stochastique. L'idée est la suivante : soit  $Q = (q_{xy})_{(x,y) \in \Omega \times \Omega}$  une quelconque matrice de transition sur  $\Omega_E$ . Cette matrice sera dite de proposition, puisqu'elle servira à proposer des transitions qui seront ou non acceptées suivant les probabilités tabulées dans la matrice d'acceptation  $A = (a_{xy})_{(x,y) \in \Omega \times \Omega}$ . Plus précisément, une transition quelconque proposée, par exemple de  $x$  vers  $y$  sera acceptée avec la probabilité  $0 \leq a_{xy} \leq 1$  et rejetée avec la probabilité  $1 - a_{xy}$  (auquel cas on procède à une transition nulle  $x \rightarrow x$ ).

Les éléments de la matrice de transition  $P = (p_{xy})_{(x,y) \in \Omega \times \Omega}$  ainsi définie s'écrivent alors

$$\forall (x, y) \in \Omega_E \times \Omega_E, \begin{cases} p_{xy} = q_{xy} a_{xy} \\ p_{xx} = q_{xx} + \sum_{y \neq x} q_{xy} (1 - a_{xy}) \end{cases} \quad (1.48)$$

Il est facile de voir que la condition de bilan détaillé est observée si et seulement si :

$$\forall (x, y) \in \Omega_E \times \Omega_E, \frac{a_{xy}}{a_{yx}} = \frac{\pi_y q_{yx}}{\pi_x q_{xy}} \quad (1.49)$$

Il suffira, pour satisfaire cette condition, de prendre :  $a_{xy} = f\left(\frac{\pi_y q_{yx}}{\pi_x q_{xy}}\right)$  avec  $f$  une fonction quelconque de  $R^+$  dans  $[0,1]$  vérifiant pour tout  $z \in R^+$  la condition :

$$\frac{f(Z)}{f(Z^{-1})} = Z \quad (1.50)$$

Remarque importante : Dans le cas qui nous intéresse, la distribution stationnaire recherchée  $\pi$  est la distribution de Gibbs à température  $T$  :

$$\pi(x) = G_T^U(x) = \frac{1}{Z_T} \exp\left(\frac{-U(x)}{T}\right) \quad (1.51)$$

On remarque alors que la fonction de partition  $Z_T$  a disparu de l'expression de la loi d'acceptation :

$$a_{xy} = f\left(\frac{G_T^U(y)q_{yx}}{G_T^U(x)q_{xy}}\right) = f\left(\exp\left[\frac{-\Delta U_x^y}{T}\right] \cdot \frac{q_{yx}}{q_{xy}}\right) \quad (1.52)$$

où  $\Delta U_x^y \equiv U(y) - U(x)$  (on appelle parfois force la quantité  $F_x^y \equiv -\Delta U_x^y$ ). Par la suite on confondra toujours  $\pi$  et  $G_T$ .

Si l'on prend en outre une matrice de proposition  $Q$  symétrique (ce qui n'est pas une nécessité, mais c'est le choix qui sera fait par la suite), alors l'expression de la probabilité d'acceptation se réduit encore plus :

$$a_{xy} = f\left(\exp\left[\frac{-\Delta U_x^y}{T}\right]\right) \quad (1.53)$$

Enfin, plusieurs choix sont possibles pour la fonction  $f$  le choix de la fonction maximale satisfaisant la condition énoncé plus haut :  $f(Z) = \min(1, Z)$  est attribué à Metropolis [18], et donne lieu à la dynamique (d'évolution) de Metropolis. La probabilité d'acceptation de la transition proposée  $x \rightarrow y$  vaut alors (voir fig1.3 et procédure P.1) :

$$a_{xy} = \begin{cases} 1 & \text{si } U(y) < U(x) \\ \exp\left[-\frac{\Delta U_x^y}{T}\right] & \text{dans le cas contraire} \end{cases} \quad (1.54)$$

Un autre choix consiste à prendre :  $f(z) = \frac{z}{1+z}$  ; il donne lieu à la dynamique de Barker [11]. Si la distribution à échantillonner est la distribution de Gibbs à température  $T$ , et si l'on choisit une matrice de tentation symétrique, alors l'application de la dynamique de Barker donne lieu à ce que l'on appelle algorithme du thermostat (ou Heat-Bath), voir Procédure P.2). Dans ce cas, la probabilité d'acceptation devient :

$$a_{xy} = \frac{1}{1 + \exp\left[\frac{\Delta U_x^y}{T}\right]} \tag{1.55}$$

Il s'agit de la loi sigmoïde

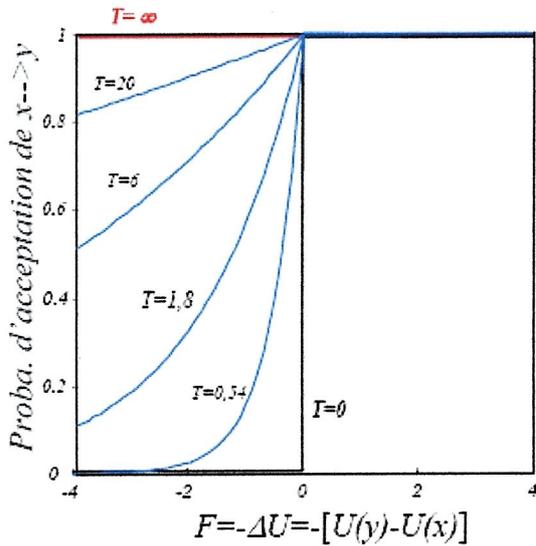


Figure1.3 : Probabilité d'acceptation pour la dynamique de Metropolis.

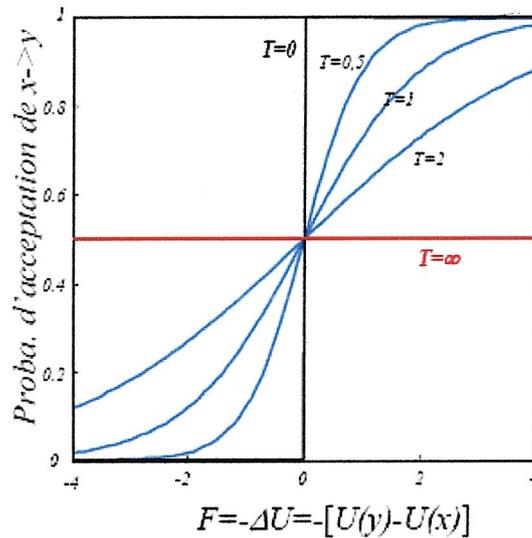


Figure1.4 : Probabilité d'acceptation sigmoïde pour la dynamique de Barker (loi du Thermostat). Remarque : plus la force est grande, et plus elle "tire" vers le nouvel état proposé.

Condition d'irréductibilité et d'apériodicité. On a donc vu comment la méthode de Metropolis-Hasting permet de construire une matrice de transition vérifiant la condition de bilan détaillé relativement à une distribution donnée (et donc d'assurer la stationnarité de cette distribution par rapport à la chaîne de Markov correspondante) à partir d'une matrice de transition quelconque ; cependant, pour appliquer le théorème de stationnarisation, il ne faut pas oublier la condition d'irréductibilité et

d'apériodicité de la matrice de transition. Une façon de s'assurer de ces propriétés consiste à procéder comme suit: considérons tout d'abord les transitions ne modifiant la configuration courante qu'en un seul site  $s$ , Soit la matrice de proposition locale  $Q_s$  dont les éléments sont définis par :

$$\forall (x, y) \in \Omega_E \times \Omega_E, q_{xy}^s = \begin{cases} \frac{1}{G_E}, & \text{si } x \text{ et } y \text{ coïncide sur toute la grille } S \text{ sauf en } s, \\ 0 & \text{si non} \end{cases} \quad (1.56)$$

Etant donnée une configuration courante  $x$ , la matrice de tentation locale  $Q^s$  propose donc de façon équiprobable toutes les transitions vers les configurations coïncidant avec  $x$  sur toute la grille  $S$ , sauf sur le site  $s$  en question.

#### Dynamique de Metropolis pour la mise à jour du site $s$ :

Soit  $x$  l'état courant du champ et  $g_x$  le label courant au site  $s$  ;

- Proposer un nouveau label  $g_y$  choisi au hasard selon une loi uniforme sur  $G_E$ .
- Calculer la force locale  $F(s)$  définie selon :

$$F(s) = -\Delta U(s) = -[U(x/x(s) = g_y) - U(x/x(s) = g_x)] \quad (1.57)$$

- Si  $F(s) \geq 0$ , alors on accepte le changement :  $x(s) \leftarrow g_y$ .
- Si  $F(s) < 0$ , alors on accepte le changement avec une probabilité  $\exp(F(s))$ .

P.1 : Procédure de mise à jour locale selon la dynamique de Metropolis (matrice de tentation locale choix équiprobable du label).

En appliquant la méthode de Metropolis-Hasting on peut construire une matrice de transition locale  $P^s$  qui sera stationnaire vis à vis de  $\pi$  (voir les procédures P.1 et P.2 pour la mise à jour de sites selon la dynamique de Metropolis et du Thermostat respectivement) ; mais il est manifeste que cette matrice n'est pas irréductible (elle ne permet pas de passer d'un état  $x$  à un autre état quelconque  $y$  différent de  $x$  pour un site autre que  $s$ ). Pour cela il faut choisir une stratégie de parcours des sites, définissant une matrice de transition globale irréductible.

Dynamique du Thermostat pour la mise à jour du site  $s$  :

Soit  $x$  l'état courant du champ et  $g_x$  le label courant au site  $s$  :

- Proposer un nouveau label  $g_y$  choisi au hasard selon une loi uniforme sur  $G_E$ .
- Calculer la force locale  $F(s)$  définie selon :

$$F(s) = -\Delta U(s) = -[U(x/x(s) = g_y) - U(x/x(s) = g_x)] \quad (1.58)$$

- Accepter ce changement avec une probabilité (loi sigmoïde) :

$$\Pr[x(s) \leftarrow g_y] = \frac{1}{1 + \exp\left[\frac{-F(s)}{T}\right]} \quad (1.59)$$

P.2 : Procédure de mise à jour locale selon la dynamique du Thermostat. (matrice de tentation locale choix équiprobable du label).

Voici deux stratégies de parcours possibles :

- 1- Stratégie de parcours aléatoire. Si  $N=nxm$  est le nombre de sites de la grille  $S$ , alors on peut définir la matrice de transition :

$$P = \frac{1}{N} \sum_{s \in S} P^s \quad (1.60)$$

Cela revient dans la pratique à réaliser un tirage aléatoire du site concerné par la transition selon une distribution équiprobable sur l'ensemble des sites de la grille. La matrice  $P$  est irréductible, puisqu'elle permet de réaliser (potentiellement) la transition entre deux configuration quelconques au plus en  $N$  opérations. La distribution  $\pi$  est toujours stationnaire, puisqu'elle

l'est pour chacune des matrices de transition locales. Enfin, elle est apériodique puisque toutes les transitions nulles sont permises. Les conditions du théorème de stationnarisation sont remplies. La procédure d'échantillonnage complète est explicitée dans l'encadré P.3.

Procédure Echantillonnage\_Aléatoire  $(x_0, T)$ .

- On part de la configuration initiale d'étiquettes  $x_0 \in \Omega_E$ ;
- On répète indéfiniment :
  - Choisir un site  $s$  au hasard.
  - Mettre à jour ce site selon la dynamique de Metropolis (P.1) ou du Thermostat (P.2) à température  $T$ .

P.3 : Procédure d'échantillonnage de Metropolis (ou du Thermostat) à parcours aléatoire des sites pour la construction d'une chaîne de Markov convergeant vers la distribution de Gibbs à température  $T$  ("thermalisation").

2- Stratégie de parcours séquentiel périodique Une autre stratégie consiste à parcourir la grille  $S$  selon une séquence prédéterminée  $\{s_0, s_1, s_2, \dots, s_N\}$  sans omettre aucun site. La séquence est souvent périodique et de période égale au nombre de sites de l'image : on parle dans ce dernier cas de balayage des sites. La chaîne de Markov associée n'est plus homogène puisque la matrice de transition  $P^n$  (créée à partir de  $Q^n$  par la procédure de Metropolis-Hasting) dépend du site considéré  $s_n$  à l'étape  $n$ . Par contre, la sous-chaîne extraite en prenant un échantillon de la chaîne à chaque fin de balayage (ou un multiple quelconque) est homogène. On vérifie alors que la matrice de transition résultante :

$$p = \prod_{n=0}^{N-1} p^{s_n} \quad (1.61)$$

est apériodique, irréductible et admet  $\pi$  comme distribution stationnaire (la condition de bilan détaillé n'est pas nécessairement remplie pour la matrice de transition globale, mais  $\pi$  est stationnaire car il s'agit d'un vecteur propre du produit des matrices de transition locales). Le balayage le plus simple suit en général un ordre lexicographique, mais d'autres parcours sont possibles pour éviter autant que possible les artefacts de convergence car dans la pratique la chaîne est tronquée on parle de chaînes de Markov à horizon fini.(balayage chaîné par exemple - lexicographique en aller-retour). On appellera itération la mise à jour complète du champ d'étiquettes (balayage de l'image toute entière).

Procédure Echantillonnage\_Parcours\_Périodique  $(x_0, T)$ .

- On part de la configuration initiale d'étiquettes  $x_0 \in \Omega_E$  ;
- On répète indéfiniment :
- Choisir un site après l'autre selon un parcours séquentiel de l'image.
- Mettre à jour ce site selon la dynamique de Metropolis (P.1) ou du

Thermostat (P.2) à température  $T$ .

P.4 : Procédure de Metropolis (ou du Heat-Bath) avec balayage séquentiel des sites pour la construction d'une chaîne de Markov convergeant à l'infini vers la distribution de Gibbs à température  $T$ .

## 1.8.2.3 L'échantillonneur de Gibbs

Condition de stationnarité et d'homogénéité. Introduit par les frères Geman [10], l'échantillonneur de Gibbs (Gibbs Sampler) ne s'appuie pas sur la condition de bilan détaillé, mais plutôt sur la distribution conditionnelle locale caractérisant le champ de Markov dont est constituée la chaîne de Markov. On considère que ces propriétés statistiques "intrinsèques" du champ sont indépendantes de  $n$ . En effet, les éléments de la matrice de transition locale  $P^s$  sont définis par :

$$\forall (x, y) \in \Omega_E \times \Omega_E \quad p_{xy}^s = \Pr(E(s) = y(s) / E(r) = x(r), r \neq s) \quad (1.62)$$

si  $x$  et  $y$  ne diffèrent que par le site  $s$ , et 0 sinon. On remarque que la probabilité de transition locale ne dépend plus de l'état courant du site  $s$ , et que la chaîne ainsi construite est homogène, ce qui n'était pas le cas directement dans la méthode de Metropolis (la probabilité d'acceptation  $a_{xy}$  dépendait du quotient  $\pi(x)/\pi(y)$ , ou de la différence d'énergie  $U(x) - U(y)$  entre l'état courant  $x$  et le nouvel état proposé  $y$  dans le cas où  $\pi$  s'assimile à la distribution de Gibbs).

Si l'on s'intéresse au problème de régularisation d'images par champs de Markov et en vertu du théorème de Hammersley-Clifford (§3.2) l'expression de la probabilité conditionnelle locale (et donc  $P_{xy}^s$ ) prend alors la forme Gibbsienne suivante (le paramètre de température  $T$  est introduit pour pouvoir réaliser le recuit) :

$$P_{xy}^s = \frac{1}{Z_s^T(y, o)} \exp\left(-\frac{U_s(y, o)}{T}\right) \equiv G_U^T(E(s) = y(s) / E(r) = x(r), r \neq s) \quad (1.63)$$

où  $Z_s^T$  et  $U_s$  sont respectivement la fonction de partition et l'énergie locales du champ de Markov couplé  $X = \{E, O\}$  (voir remarque au §4.1), associées au site  $s$  :

$$U_s(y, o) = \sum_{c \in C_X / s \in c} V_c(y, o) \text{ et } Z_s^T(y, o) = \sum_{o \in G_E} \exp\left[-\frac{U_s(y, o |_{y(s)=a})}{T}\right] \quad (1.64)$$

Dans ces conditions, la distribution de Gibbs (à température  $T$ ) est bien stationnaire pour la chaîne de Markov engendrée (et cela même si la condition de bilan détaillée n'est pas respectée). En effet, supposons que la distribution à l'étape  $n$  soit la distribution de Gibbs à température  $T$ , i.e :  $\forall x \in \Omega_E, \Pr(E_n = x) = G_U^T(x)$ . Après la mise à jour locale du site  $s$ , on aura :

$$\Pr(E_{n+1} = x) \Pr(E_{n+1}(s) = x(s) / E_n(r) = x(r), r \neq s) \cdot \Pr(E_n(r) = x(r), r \neq s), \quad (1.65)$$

$$= G_U^T(E(s) = x(s) / E(r) = x(r), r \neq s) \cdot G_U^T(E(r) = x(r), r \neq s) = G_U^T(x) \quad (1.66)$$

Car la mise à jour des sites se fait précisément selon la loi de distribution de Gibbs conditionnelle locale. On vient donc de montrer que la distribution de Gibbs est vecteur propre de la matrice de transition de l'échantillonneur de Gibbs : elle est donc stationnaire par rapport à la chaîne de Markov.

La procédure de mise à jour locale pour l'échantillonneur de Gibbs sur champ de Markov est esquissée dans l'encadré P.5.

Procédure de Gibbs pour la mise à jour du site  $s$  à température  $T$  :

Soit  $x$  l'état courant du champ d'étiquettes ;

- Pour chaque  $a_i \in G_E$ , calculer la probabilité conditionnelle locale  $p_i$  :

$$p_i = \Pr[E(s) = a_i / E(r) = x(r), r \in V(s)] = \frac{1}{Z_s^T(y, o)} \exp\left[-\frac{U_s(y, o)}{T}\right] \quad (1.67)$$

Où  $y(s) = a_i$  et  $y(r) = x(r)$  pour  $r \neq s$ .

- Choisir au hasard la nouvelle étiquette pour  $s$  suivant cette distribution ; en pratique cela revient à tirer un nombre  $\mu \in [0, 1]$  selon une loi uniforme et la nouvelle étiquette attribuée  $s$  sera  $a_j$  correspondant au plus petit indice  $j$  tel

$$\text{que } \sum_{n=1}^j p_n > \mu$$

P.5 : Remise à jour locale par l'échantillonneur de Gibbs dans le cas d'un champ d'étiquettes  $E$  markovien. On remarquera que la probabilité de transition locale ne dépend pas de la valeur du label courant au site  $s$ .

Condition d'irréductibilité et d'apériodicité. Pour construire une matrice de transition globale qui soit irréductible et apériodique, il suffit de procéder comme auparavant, on choisissant soit une stratégie de parcours aléatoire, soit un balayage séquentiel périodique (voir encadré P.6).

Procédure Thermalisation\_Gibbs  $(x_0, T)$ .

- On part de la configuration initiale d'étiquettes  $x_0 \in \Omega_E$  ;
- On répète indéfiniment :
- Choisir les sites selon un parcours séquentiel ou aléatoire.
- Mettre à jour ce site selon la procédure de Gibbs (P.5) à température  $T$ .

P.6 : Procédure de thermalisation (ou échantillonnage) de Gibbs avec balayage séquentiel ou aléatoire des sites.

Enfin, il est intéressant de noter que si le champ de Markov est à valeurs binaires ( $G_E = \{0, 1\}$ ) alors la matrice de transition dans le cas de l'échantillonneur de Gibbs coïncide avec celle de l'algorithme du Thermostat. En particulier, la force ne dépend

pas de la valeur courante du site considéré, mais seulement de celles de ses voisins (exemple débruitage d'images binaires au §1.2.3).

Nous allons désormais porter toute notre attention au cas particulier des champs d'étiquettes binaires, dont la dynamique est suffisante pour traiter le modèle d'Ising 2D, le problème de la restauration d'images binaires.

#### 1.8.2.4 Régime transitoire

Quel que soit l'état initial du champ d'étiquettes, on peut être assuré que si la procédure d'échantillonnage (Metropolis ou Gibbs) est conforme aux prescriptions énoncées dans le théorème de stationnarisation, alors la chaîne de Markov engendrée converge vers la distribution d'équilibre... au bout d'un nombre infini d'itérations! Dans la pratique toutefois, la chaîne de Markov sera nécessairement tronquée. Il faut pouvoir en mesurer les conséquences ; en particulier, si le champ n'est pas initialement à l'équilibre, alors il serait utile de connaître la durée du régime transitoire (ou de façon équivalente la rapidité de convergence) avant qu'il n'atteigne une configuration proche de l'équilibre et pouvoir fixer ainsi un nombre d'itérations raisonnable qui ne soit ni trop petit ni trop grand.

Remarque : Le problème est en vérité très délicat ; sans rentrer dans les détails (on pourra trouver une discussion intéressante dans [23]), on montre que l'on peut obtenir au mieux une borne supérieure pour la durée du régime transitoire à partir de la connaissance du temps exponentiel d'autocorrelation défini selon :

$$\tau_{\text{exp}} = \sup_f \tau_{\text{exp},f}, \text{ avec } \tau_{\text{exp},f} = \limsup_{t \rightarrow \infty} \frac{t}{-\log(\rho_{ff}(t))} \quad (1.68)$$

où  $f$  est une observable quelconque du système à valeurs réelles (par exemple l'énergie) et  $\rho_{ff}$  sa fonction d'autocovariance normalisée. Cette dernière décroît typiquement de façon exponentielle (selon  $e^{-t/\tau}$ ) ;  $\tau_{\text{exp}}$  représente donc la constante de temps de l'observable "la plus lente" du système (pour des chaînes à temps discret,  $\tau_{\text{exp}}$  correspond à un nombre d'itérations entier). Par exemple, si l'on attend  $10 \cdot \tau_{\text{exp}}$  alors le biais entre la valeur de n'importe quelle observable et sa valeur moyenne à l'équilibre est au plus égale à  $e^{-10}$  ( $\approx 4,5 \times 10^{-5}$ ). Malheureusement la connaissance exacte de  $\tau_{\text{exp}}$  s'avère en général impossible. On peut essayer tout au plus d'estimer

l'autocovariance normalisée sur quelques observables qui devraient a priori fluctuer lentement mais le choix des observables peut s'avérer maladroit.

D'autre part, il se peut bien évidemment que la convergence réelle soit beaucoup plus rapide qu'il n'en résulte de l'estimation précédente. On procède alors de façon empirique, en traçant plusieurs observables en fonction du temps et en notant la longueur apparente du régime transitoire. C'est ce que nous avons fait pour le modèle de la restauration d'images binaires.

#### 1.8.2.5 Parallélisation des procédures d'échantillonnage

Il est possible d'adopter deux stratégies complémentaires de Parallélisation pour accélérer la convergence des algorithmes d'échantillonnage :

- Chaînes de relaxation en parallèle. Cette classe regroupe des algorithmes à gros grain, où plusieurs unités de calcul exécutent indépendamment des processus de relaxation séquentiels [1]. Les différents processus explorent de façon concurrente l'espace des configurations et échangent périodiquement des informations afin d'accélérer (ou d'améliorer) la convergence. Ce genre d'algorithme se prête bien à l'implantation sur des machines parallèles à usage général de type MIMD (Multiple Instructions, Multiple Data).
- Réactualisation simultanée des sites. Les algorithmes appartenant à cette classe correspondent à un niveau de parallélisme à grain fin où plusieurs de sites sont mis à jour en parallèle (voire tous les sites en même temps). Une seule chaîne est générée, qui converge plus rapidement que son homologue séquentiel. Ces algorithmes sont adaptés à la mise en oeuvre sur des machines possédant un grand nombre de processeurs à connexions locales fonctionnant en mode SIMD (Single Instruction, Multiple Data).

Commençons par une remarque intéressante. La localité des transitions proposées (par l'échantillonneur de Metropolis ou de Gibbs) est susceptible de rendre également local le calcul de la différence d'énergie :

$$F_x^y = F(s) = -\Delta U_x^y = -\Delta U(s) = -[U(x/x(s) = g_x) - U(x/x(s) = g_y)] \quad (1.69)$$

À condition que l'étendue des interactions entre les différentes variables de la fonction d'énergie  $U$  à minimiser soit elle même réduite (on parlera alors de force

locale). C'est précisément le cas pour l'énergie d'un champ de Markov à voisinage réduit d'ordre 1 ou 2 (§3).

Exemple : Force locale dans le cas du débruitage d'images binaires. L'énergie associée au problème de la restauration d'images binaires

$$U(e, o) = -\frac{\beta_s}{2} \sum_{s \in S} \sum_{r \in \mathcal{V}(s)} (2e(s) - 1)(2e(r) - 1) - \frac{\beta_o}{2} \sum_{s \in S} (2o(s) - 1)(2e(s) - 1) \quad (1.70)$$

La différence d'énergie ou force locale vaut alors (en supposant que l'étiquette courante du site  $s$  soit  $e(s)=1$  et celle du nouvel état proposé soit  $e(s)=0$ ) :

$$F_0^1(s) = -\Delta U(s) = -U(e, o / e(s) = 1) + U(e, o / e(s) = 0) = F_o(s) + F_s(s) \quad (1.71)$$

- Le premier terme ou force locale de contrainte aux données vaut alors :

$$F_o(s) = \beta_o (2o(s) - 1) \quad (1.72)$$

- De même, le deuxième terme ou force spatiale de régularisation vaut :

$$F_s(s) = \beta_s \sum_{r \in \mathcal{V}(s)} (2e(r) - 1) \quad (1.73)$$

- La force locale qui "tire" de l'état courant 1 vers le nouvel état proposé 0 s'écrit finalement :

$$F_0^1(s) = \beta_o (2o(s) - 1) + \beta_s \sum_{r \in \mathcal{V}(s)} (2e(r) - 1) \quad (1.74)$$

La Parallélisation massive de la mise à jour des sites n'est pas pour autant immédiate. En effet, le calcul de chaque matrice de transition locale  $P^n$  (i.e. le calcul de la probabilité de mise à jour et la mise à jour du site, les procédures de mise à jour locales P.1, P.2 et P.5), bien qu'il puisse être effectué localement, est dépendante de la réalisation particulière du champ à l'étape  $n-1$ .

## 1.9. Conclusion

Tout d'abord il est intéressant de remarquer que les méthodes d'échantillonnage constituent à elles seules un outil permettant d'ajuster les paramètres d'un modèle particulier en traitement d'images. En effet, les techniques présentées dans cette partie permettent d'appréhender de façon "expérimentale" les propriétés statistiques globales émergeant d'un champ de Markov dont on aurait spécifié de façon arbitraire les potentiels locaux des cliques.

Néanmoins, et comme il a été expliqué brièvement au §1.8.2 l'intérêt fondamental des techniques d'échantillonnage de Monte Carlo est que, utilisées conjointement avec une procédure de refroidissement (i.e. une loi adéquate de décroissance de la température) celles-ci vont permettre d'approcher la configuration d'étiquettes minimisant une fonction de coût donnée. C'est ce que l'on verra avec un peu plus de détail à la partie suivante.

# Partie 3

## 1.10. La procédure de refroidissement ou Recuit Simulé

Les méthodes de mise à l'équilibre thermodynamique décrites précédemment peuvent être opérées à une température  $T$  quelconque. Il serait tentant de choisir directement une température proche de zéro pour favoriser d'un seul coup les modes fondamentaux de l'énergie. Malheureusement, on montre [24] que le temps de calcul à l'approche de l'équilibre croît exponentiellement avec  $1/T$ . L'algorithme de recuit simulé, en faisant décroître la température de manière contrôlée, est chargé de contourner cet inconvénient. La chaîne de Markov est tout d'abord rapidement amenée à l'équilibre thermodynamique à haute température ; puis, à chaque itération (i.e. mise à jour complète du champ d'étiquettes), la température est diminuée, et la chaîne est maintenue dans un état proche d'équilibre par l'exécution plus ou moins longue (mais jamais infinie!) d'une procédure d'échantillonnage. Le principe du recuit est que à haute température le paysage énergétique est rapidement exploré et le système sort aisément des minima locaux ; puis l'exploration se fait de moins en moins rapide mais le système rôde déjà plus probablement autour du minimum global (ou au moins d'un minimum global, s'il y en a plusieurs). Le fait que la température ne soit jamais nulle implique que si l'on attend suffisamment longtemps, il existe une probabilité non nulle que le système sorte de n'importe quel minimum local, aussi profond soit-il. Toutefois ce temps croît de façon exponentielle avec  $1/T$ , si bien qu'à basse température l'algorithme se comporte de façon de plus en plus déterministe et seuls les changements diminuant l'énergie seront acceptés.

On distingue deux classes différentes d'algorithmes de relaxation correspondant à des schémas de descente en température continus ou par paliers.

### 1.10.1 Recuit continu

Dans ce genre d'algorithmes, la température est abaissée entre deux itérations (mises à jour) consécutives. La chaîne de Markov qui en résulte est inhomogène. Le système n'est jamais totalement à l'équilibre thermodynamique ; cependant, il est démontré formellement [10] que si la température tend vers zéro et que à chaque itération  $n$  on a :

$$T_k \geq \frac{\alpha}{\log(n)}, \text{ où } \alpha = \text{card}(\Omega_E) \cdot \left\{ \max_{e \in \Omega_E} [U(e)] - \min_{e \in \Omega_E} [U(e)] \right\} \quad (1.75)$$

Alors la configuration converge vers un mode fondamental de l'énergie (il peut en avoir plusieurs). Malheureusement cette décroissance logarithmique reste dans la pratique beaucoup trop lente pour être effectivement respectée. En pratique on préfère des lois de décroissance exponentielles plus rapides de type  $T_{n+1} = r.T_n$ , avec  $r < 1$  et proche de 1 ; mais le système risque d'être piégé dans des minima locaux.

### 1.10.2. Algorithmes avec descente en température par paliers

Pour éviter autant que possible ces pièges, une tactique consiste à maintenir la valeur de la température un peu plus longtemps avant de la changer, dans l'espoir que l'équilibre soit un peu mieux atteint à la fin de chaque palier de température (voir procédure P.7). La difficulté consiste alors à déterminer la longueur minimale du palier. Par la suite on se référera à la quantité  $nscan(T)$ , définie comme le nombre de balayages complets (ou scan) de l'image effectués pendant toute la durée du palier à température constante  $T$ . Il n'est pas nécessaire que  $nscan(T)$  soit beaucoup plus grand que le nombre d'itérations nécessaires pour dépasser le régime transitoire de la procédure d'échantillonnage toute seule là précisément réside l'intérêt du recuit simulé.

### 1.10.3. Critère d'arrêt : définition heuristique du "recuit sérieux"

Dans la pratique, l'algorithme stochastique doit être stoppé au bout d'un nombre fini d'itérations. Pour simplifier l'algorithme, nous avons fixé le nombre maximal d'itérations par palier quelle que soit la température.

Alors la Procédure de Recuit Simulé comme suite :

### Algorithme de Recuit Simulé Parallèle

(avec décroissance de la température par paliers)

Paramètres du recuit :

**Ti** : température initiale.

**Tf** : température finale

**r** : facteur de décroissance de la température

**nscan** : nombre d'itérations par palier de température.

Début recuit simulé

$T \leftarrow Ti$

Tant que  $r.T > Tf$ , faire :

Répéter **nscan** fois :

- Mettre à jour des sites
- Calculer la *force locale*  $F(s)$  qui "tire" vers l'état  $e(s)=1$  définie selon

$$F(s) = -\Delta U_0^1(s) = - \sum_{c \in C / s \in c} [V_c(e(s)=1) - V_c(e(s)=0)]$$

(cette quantité est indépendante de l'état courant du site).

- Accepter ce changement avec la probabilité (loi sigmoïde) :

$$\Pr[e(s) = 1] = \frac{1}{1 + e^{-\frac{F(s)}{T}}}$$

Fin mise à jour du champ d'étiquettes

Fin thermalisation pour le palier de température courant

$T \leftarrow r \times T$  (refroidissement)

Fin procédure de recuit simulé

P.7 : Procédure de Recuit Simulé Parallèle avec décroissance de la température par paliers.

## 1.11. Algorithmes déterministes

### 1.11.1. Introduction

Bien que théoriquement sous-optimaux, les algorithmes déterministes offrent une alternative aux méthodes stochastiques souvent trop coûteuses en temps de calcul (même pour des algorithmes parallélisés). En première approximation et à la lumière de ce qui vient d'être expliqué au paragraphe précédent- les algorithmes déterministes sont des techniques d'échantillonnage qui se déroulent à température nulle : à chaque itération, la transition se fait de façon déterministe depuis la configuration courante vers une configuration d'énergie strictement inférieure (les remontées d'énergie ne sont jamais acceptées). Ces algorithmes partent d'une configuration initiale et convergent vers le premier minimum local rencontré dans le paysage énergétique (fig.1.5).

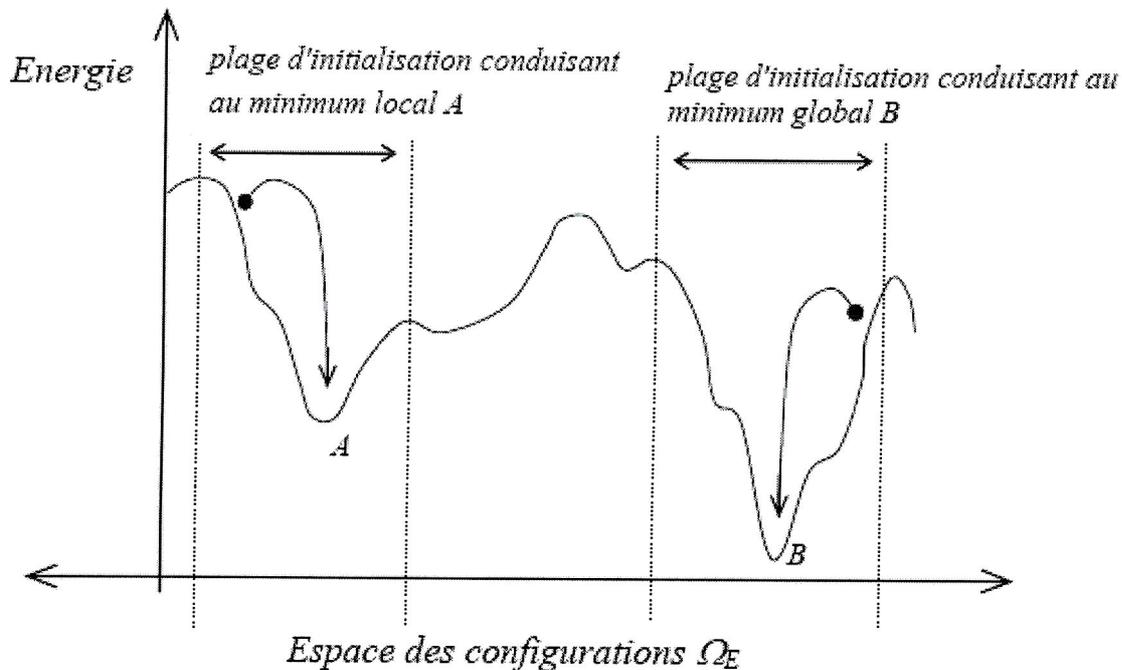


Figure. 1.5 : Paysage énergétique et importance de l'initialisation dans le cas de la relaxation déterministe (Pour simplifier la visualisation on a représenté un espace des configurations unidimensionnel).

Si la configuration d'étiquettes initiale est proche d'une solution optimale, alors on peut espérer converger vers ce minimum global ; si ce n'est pas le cas, l'algorithme est irrémédiablement piégé dans un minimum local de l'énergie. D'où la nécessité de choisir avec soin la configuration de départ. Dans le cas des problèmes inverses en traitement d'images, une possibilité consiste à prendre comme configuration initiale

l'estimée du maximum de vraisemblance (c'est à dire le champ d'observation lui même).

Parmi les méthodes de relaxation déterministes largement utilisées figurent :

- L'algorithme ICM (pour Iterated Conditional Modes) et proposé par [6]. Il consiste à remettre à jour chaque site du champ avec l'étiquette qui minimise le gradient local de l'énergie. La convergence vers une solution stable est très rapide. L'arrêt de la procédure peut être décidée par un nombre fixe d'itérations, ou lorsque l'on a atteint un équilibre global sur l'ensemble des sites. L'algorithme ICM n'est autre que l'échantillonneur de Gibbs à température nulle.
- Le recuit par champ moyen ou MFA (pour Mean Field Annealing, [21]) fait l'hypothèse qu'un système composé d'un grand nombre de variables en interaction peut être vu comme un ensemble de systèmes à une seule variable, pour lesquels les voisins sont gelés dans leur état moyen (en pratique on remplace les étiquettes voisines du site traité par leur état le plus probable calculé pendant la relaxation).
- L'algorithme du GNC (pour Graduated Non Convexity), introduit par Blake et Zisserman [7] consiste à construire une famille de fonctions convexes approchant de plus en plus la fonction d'énergie originale à minimiser. La minimisation de la fonction originale à lieu par minimisations successives de chaque fonction convexe (à minimum unique facile à atteindre par l'algorithme du gradient par exemple). Cette technique cherche à pallier l'inconvénient du piégeage dans un minimum local de l'algorithme ICM.
- Une autre approche astucieuse consiste à réduire l'espace de configurations dans un premier abord pour l'élargir convenablement ensuite au fur et à mesure de la recherche du minimum par relaxation déterministe : on s'assure ainsi d'une bonne configuration initiale à chaque étape de la relaxation. Ces méthodes portent le nom de méthodes de relaxation multirésolution, et ont été introduites en traitement d'images par [20]. L'initialisation est alors en général moins cruciale que dans le cas où l'on s'attaque directement à l'image dans sa résolution maximale.

## 1.12. Conclusion

Au Chapitre nous avons étudié quelques problèmes inverses en traitement d'images bas niveau ; nous avons montré que moyennant une modélisation mathématique adéquate du champ de d'étiquettes (i.e. les primitives de l'image recherchées), leur résolution passait par la minimisation d'une fonction d'énergie à grand nombre de variables. Cette minimisation est, dans la plupart des cas, un problème d'optimisation difficile. Nous avons exposé dans ce chapitre la procédure de recuit simulé, méthode d'optimisation stochastique générale capable de fournir des solutions très proches d'une solution optimale pour ce genre de problèmes. Le seul inconvénient est la lenteur extrême de convergence quand la procédure est opérée sur un ordinateur séquentiel. En effet, il faut d'une part (a) mettre à jour une à une les variables de la fonction d'énergie à minimiser, en calculant à chaque fois le gradient local de l'énergie, nécessaire pour effectuer le tirage aléatoire du nouvel état selon une loi de probabilité bien déterminée (contrôlé globalement par le paramètre de température) ; et (b) d'autre part, il faut répéter cette procédure un grand nombre de fois en réduisant très lentement la température du recuit.

Mais le problème que les ordinateurs séquentiels ne sont pas disponibles et très chère alors la solution que nous faisons est une hybridation de champs de Markov avec l'analyse multirésolution pour accélérer le processus

Alors au chapitre suivant nous présenterons tout d'abord un panorama de l'analyse temps-fréquence et leurs propriétés fondamentales En suite nous rappellerons les théories de la transformée en ondelettes et de l'analyse multirésolution

# LA TRANSFORMEE EN ONDELETTES

## 2.1. Introduction

Le traitement du signal est devenu une composante essentielle de l'activité scientifique et technologique contemporaine. Le traitement du signal est par exemple utilisé dans les télécommunications ou l'analyse des images dans un but de traitement (exemple la restauration des images de télédétection, obtenues par des capteurs aéroportés ou spatioportés), d'identification, de compression, de transmission ou de compréhension.

La plus célèbre et la plus ancienne des transformations pour étudier les différents points énumérés ci-dessus est la transformée de Fourier. Lors de cette transformation le signal est décomposé sur un ensemble de fonctions de base qui sont le cosinus, le sinus ou l'exponentielle imaginaire.

Très tôt dans l'histoire du traitement du signal, il est apparu que la décomposition obtenue n'était pas toujours la plus satisfaisante lorsque l'on était confronté à des phénomènes non-stationnaires.

Une nouvelle transformation nommée transformation en ondelettes est née de la convergence des travaux théoriques déjà anciens, notamment ceux de Haar (1910), de Littlewood et Paley (1930), de Zygmund (1930), de Gabor (1940), puis vers 1960 de Calderon, et des idées récentes mises en avant propos; pour le traitement numérique de certains signaux par Morlet (le premier à avoir proposé le nom d'ondelettes, 1982), ou pour le développement d'outils mathématiques utilisés en physique théorique par Grossmann (1983).

Dés lors de nombreux chercheurs apportèrent des bases mathématiques solides en faisant apparaître la notion de base orthogonale (Meyer, 1985), d'analyse multirésolution (Mallat, 1989), et d'ondelettes à support compact (Daubechies, 1988). Pour obtenir plus de renseignements sur l'histoire des ondelettes le lecteur peut se référer aux ouvrages [25] et [26].

### 2.1.1. Analyse de Fourier

L'analyse temporelle d'un signal doit s'accompagner très souvent d'une analyse fréquentielle. La première représentation nous renseigne sur la durée du signal et ses discontinuités, la deuxième présente l'intérêt de renseigner sur la périodicité du signal.

#### 2.1.1.1 Transformée de Fourier

La méthode dite de décomposition en série de Fourier, consistant en la décomposition d'un signal en somme de fonctions sinusoïdales de fréquences différentes, est l'outil le plus connu et utilisé pour le passage du domaine temporel au domaine fréquentiel [27].

Les séries de Fourier sont utilisées pour l'étude de signaux périodiques (par exemple, les signaux carrés, triangulaires etc....). Ceux-ci peuvent en effet être représentés comme une superposition d'une onde sinusoïdale fondamentale (dont la fréquence est appelée la fréquence fondamentale) et de diverses harmoniques de fréquences multiples de la fréquence fondamentale.

Pour les signaux non périodiques, il est nécessaire d'avoir recours à une intégrale de Fourier, c'est à dire à une somme continue. Cette méthode consiste à représenter le signal par une superposition d'ondes sinusoïdales de toutes les fréquences possibles. Les amplitudes associées à ces fréquences représentent, comme pour les séries de Fourier, les importances respectives des diverses ondes sinusoïdales.

Ces amplitudes forment alors une fonction de la fréquence  $f$  appelée "spectre continu des fréquences du signal" : c'est la transformée de Fourier du signal. Elle est calculée à l'aide de l'intégrale de Fourier :

$$\hat{s}(w) = \int_{-\infty}^{+\infty} s(t) e^{-j\omega t} dt \quad \omega = 2\pi f \quad (2.1)$$

La transformée inverse permet de reconstruire le signal à partir des sinusoïdes qui le Constituent :

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{s}(w) e^{j\omega t} d\omega \quad (2.2)$$

Pour que la transformée de Fourier existe, le signal doit être de carré sommable, c'est à dire d'énergie finie. Pour les signaux réels, cette condition est toujours remplie puisque la mesure est faite sur un temps fini.

L'analyse de Fourier suppose implicitement que le signal est identique à lui-même en dehors de l'intervalle de mesure. La fonction  $\hat{s}(w)$  étant périodique de période  $T$ , il est d'usage de limiter son intervalle de définition à  $[-T/2, T/2]$ .

Voici une des propriétés de la transformée de Fourier [27], dite identité de Parseval, utile pour la suite :

$$\int_{-\infty}^{+\infty} |s(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{s}(w)|^2 dw \quad (2.3)$$

D'une façon générale, on peut définir la transformée de Fourier comme une application linéaire qui associe à  $N$  valeurs  $s(0), \dots, s(k), \dots, s(N-1)$ ,  $N$  autres valeurs  $\hat{s}_0, \dots, \hat{s}_n, \dots, \hat{s}_{N-1}$ , définies par :

$$\hat{s}_n = \sum_{k=0}^{N-1} s(k) e^{-j2\pi \frac{kn}{N}}, \quad n \in \{0, \dots, N-1\} \quad (2.4)$$

où  $N$  représente le nombre minimum d'échantillons à prendre pour reconstruire le signal  $s(t)$ . Dans ce cas on nommera cette transformée, transformée de Fourier discrète. La transformée de Fourier discrète inverse s'écrit :

$$s(k) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{s}_n e^{j2\pi \frac{kn}{N}}, \quad k \in \{0, \dots, N-1\} \quad (2.5)$$

### 2.1.1.2 Echantillonnage

Par définition, un signal  $s(t)$  d'énergie finie dont la transformée de Fourier a un support borné  $[-B, B]$  est entièrement défini par ses échantillons  $s(kT_e)$  prélevés à la fréquence d'échantillonnage  $f_e \geq 2B$ .

En multipliant le signal  $s(t)$  par la fonction d'échantillonnage  $e(t)$ , le signal Échantillonné suivant est obtenu :

$$\tilde{s}(t) = s(t)e(t) = \sum_{k=-\infty}^{+\infty} s(kT_e) \delta(t - kT_e) \quad (2.6)$$

La transformée de Fourier s'écrit alors :

$$\tilde{\hat{s}}(f) = \frac{1}{T_e} \sum_{n=-\infty}^{+\infty} \hat{s}\left(f - \frac{n}{T_e}\right) \quad (2.7)$$

$\tilde{\hat{s}}(f)$  est donc périodique de période  $1/T_e$ . Afin d'éviter que le spectre ne se replie par rapport aux axes de symétrie (Figure 2.1), il est nécessaire que le signal ne comporte aucune fréquence supérieure à  $f_e/2$ .

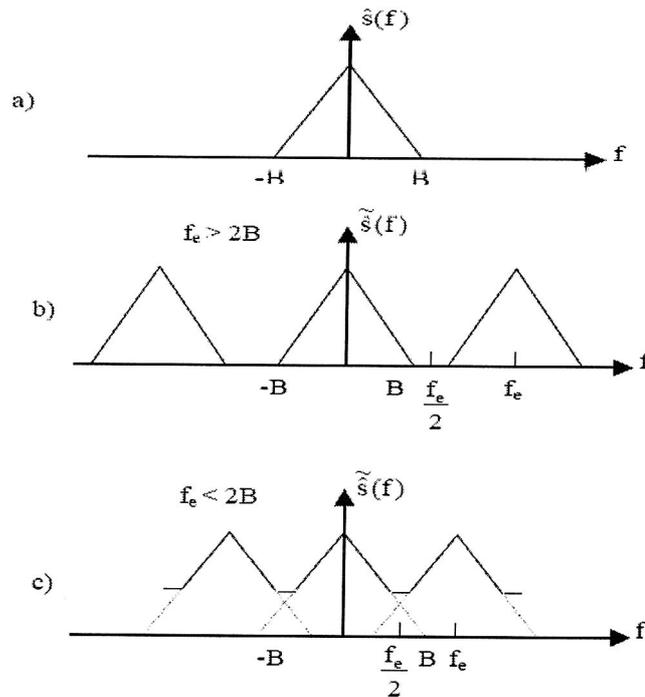


Figure 2.1 a) transformée de Fourier d'un signal  $s(t)$ ,  
 b) bon échantillonnage,  
 c) repliement du spectre.

### 2.1.1.3 Inconvénients de l'analyse de Fourier

Malgré son immense succès, cette technique a plusieurs défauts :

Manque de localisation temporelle. En effet, l'analyse de Fourier permet de connaître les différentes fréquences excitées dans un signal, c'est à dire son spectre, mais ne permet pas de savoir à quels instants ces fréquences ont été émises. Cette analyse donne une information globale et non locale, car les fonctions d'analyse utilisées sont des sinusoides qui oscillent indéfiniment sans s'amortir. Cette perte de localité n'est pas un inconvénient pour analyser des signaux stationnaires, mais le devient pour des signaux non stationnaires.

Quand on a besoin de la localisation dans le temps des composantes spectrales, on utilise une transformée qui fournit une représentation temps-fréquence du signal.

### 2.1.2. Analyse temps fréquence

Gabor dans les années 1940 [28] découvre la première forme de représentation temps-fréquence. Sa technique consiste à découper le signal en différentes plages ou fenêtres de longueur fixée. Chaque plage est alors étudiée séparément des autres par l'analyse traditionnelle de Fourier.

#### 2.1.2.1 Le fenêtrage

L'idée consiste à tronquer le signal on ne le considérant que sur un intervalle fini  $[-\tau, \tau]$ . Ceci revient à multiplier le signal  $s(t)$  par une fonction de fenêtrage  $\phi(t)$  qui est égale à zéro en dehors de l'intervalle d'étude. Le produit :

$$s(t)\phi(t-b) =: s_b(t) \quad (2.8)$$

Contient alors les informations de  $s(t)$  autour de  $t = b$ . En particulier si  $\phi(t) = X_{[-\tau, \tau]}(t)$ , voir Figure 2.2, alors :

$$s_b(t) = \begin{cases} s(t) & t \in [b-\tau, b+\tau] \\ 0 & \text{sin on} \end{cases} \quad (2.9)$$

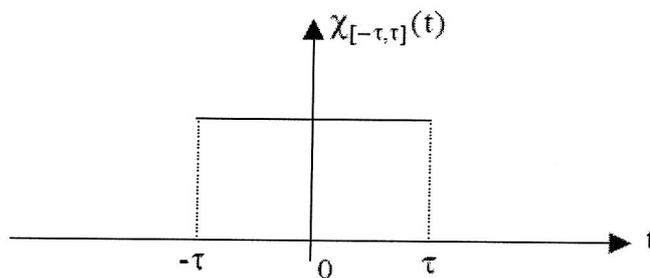


Figure 2.2 la fonction  $\phi(t) = X_{[-\tau, \tau]}(t)$

En changeant le paramètre  $b$ , la fenêtre se déplace le long de l'axe des temps. Une fonction de fenêtrage est essentiellement définie par son centre :

$$t^* := \frac{1}{\|\phi\|^2} \int_{-\infty}^{+\infty} t |\phi(t)|^2 dt \quad (2.10)$$

et sa largeur, elle même définie par l'écart type :

$$\Delta\phi := \frac{1}{\|\phi\|} \left\{ \int_{-\infty}^{+\infty} (t - t^*)^2 |\phi(t)|^2 dt \right\}^{1/2} \quad (2.11)$$

La fonction  $\phi(t)$  décrite précédemment est une fonction de fenêtrage en temps. Une fonction de fenêtrage en fréquence  $\hat{\phi}(w)$  peut tout aussi bien être définie par son centre :

$$w^* := \frac{1}{\|\hat{\phi}\|^2} \int_{-\infty}^{+\infty} w |\hat{\phi}(w)|^2 dw \quad (2.12)$$

et par son écart type :

$$\Delta\hat{\phi} := \frac{1}{\|\hat{\phi}\|} \left\{ \int_{-\infty}^{+\infty} (w - w^*)^2 |\hat{\phi}(w)|^2 dw \right\} \quad (2.13)$$

La fonction  $\phi(t)$  de la Figure 2.2 est bien définie en temps mais très mal définie en fréquence puisqu'il est facile de remarquer que  $w^* = 0$  et  $\Delta\hat{\phi} = +\infty$ .

Ainsi, le but est d'obtenir une fonction de fenêtrage correctement définie à la fois en temps et en fréquence. Un compromis entre résolution temporelle et résolution fréquentielle, donné par le principe d'incertitude de Heisenberg doit être envisagé (une démonstration de ce principe d'incertitude est donné dans [25]) :

$$\Delta\phi \Delta\hat{\phi} \geq \frac{1}{2} \quad (2.14)$$

L'inégalité ci-dessus devient une égalité si et seulement si  $\phi(t)$  est de type gaussien [29].

### 2.1.2.2 Transformée de Fourier à fenêtre glissante

En limitant le domaine d'intégration temporel à l'aide d'une fonction de fenêtrage  $\phi(t)$  (pour obtenir  $s(t)\phi(t-b) =: s_b(t)$ , que l'on fait glisser pour explorer le signal  $s(t)$  (on applique la transformée de Fourier sur  $s_b(t)$ , on obtient la transformée de Fourier à fenêtre glissante (STFT) définie de la manière suivante :

$$(G_\phi s)(b, \xi) := \int_{-\infty}^{+\infty} s(t)\overline{\phi(t-b)}e^{j\xi t} dt \quad (2.15)$$

Où

$$\phi_{b,\xi} := \phi(t-b)e^{j\xi t} \quad \text{et} \quad \xi = 2\pi\lambda \quad (2.16)$$

Le paramètre  $\lambda$  joue le rôle d'une fréquence localisée autour de l'abscisse  $b$  du signal temporel.  $(G_\phi s)(b, \xi)$  donne ainsi une indication sur ce qui se passe autour de l'abscisse  $t = b$  pour la fréquence  $\lambda$ . La fonction de fenêtrage  $\phi(t)$  peut être complexe et satisfait la condition :

$$\hat{\phi}(0) = \int_{-\infty}^{+\infty} \phi(t) dt \neq 0 \quad (2.17)$$

ce qui signifie que  $\hat{\phi}(w)$  se comporte comme un filtre passe-bas.

Cette transformée peut être interprétée comme une projection de  $s$  sur la base des fonctions fenêtrées glissantes  $\phi$  :

$$(G_\phi s)(b, \xi) = \langle s(t), \phi(t-b)e^{j\xi t} \rangle \quad (2.18)$$

La fonction  $\phi_{b,\xi}(t) = \phi(t-b)e^{j\xi t}$  se comporte comme une onde oscillante dans l'enveloppe de la fonction  $\phi(t)$ .

La Figure 2.3 présente la notion de fenêtrage temps-fréquence donnée par transformée de Fourier à fenêtre glissante. Cette transformée donne des informations sur la fonction  $s(t)$  dans la fenêtre temps-fréquence suivante [29] :

$$[t^* + b - \Delta\phi, t^* + b + \Delta\phi] \times [w^* + \xi - \Delta\hat{\phi}, w^* + \xi + \Delta\hat{\phi}] \quad (2.19)$$

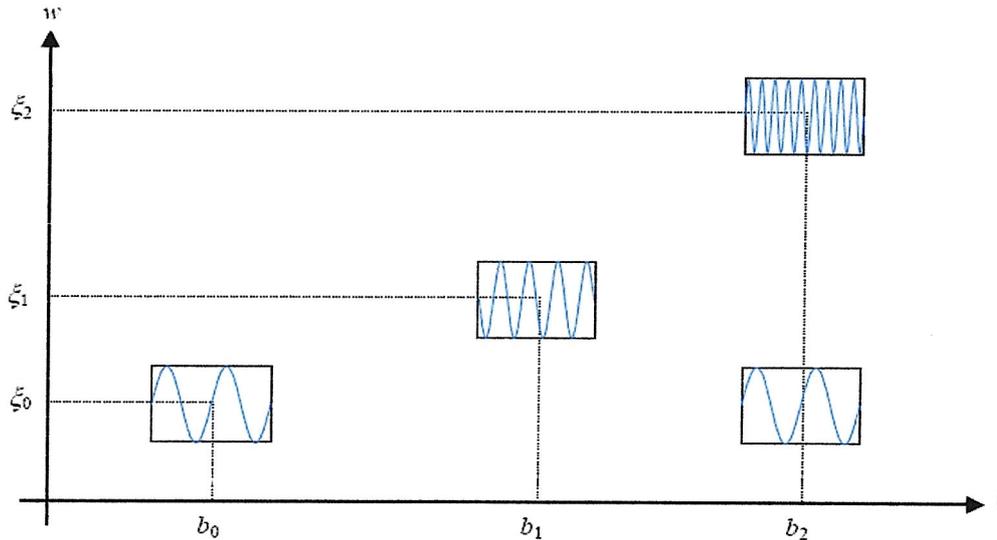


Figure 2.3 : Fenêtrage temps-fréquence de la STFT ( $t^* = w^* = 0$ ).

### 2.1.2.2.1 Propriétés de la STFT

- Linéarité :

Si  $s(t) = \alpha s_1(t) + \beta s_2(t)$  est une combinaison linéaire de deux fonctions  $s_1(t)$  et  $s_2(t)$  avec des coefficients  $\alpha$  et  $\beta$  indépendants de  $t$ , alors la STFT de  $s(t)$ , est la combinaison linéaire des STFT de  $s_1(t)$  et  $s_2(t)$  :

$$(G_\phi s)(b, \xi) = \alpha(G_\phi s_1)(b, \xi) + \beta(G_\phi s_2)(b, \xi) \quad (2.20)$$

- Translation temporelle :

Si  $s_0(t) = s(t - t_0)$  alors :

$$\begin{aligned} (G_\phi s_0)(b, \xi) &= \int_{-\infty}^{+\infty} s(t - t_0) \phi(t - b) e^{j \xi t} dt \\ &= \int_{-\infty}^{+\infty} s(t) \phi[t - (b - t_0)] e^{-j \xi t} e^{-j \xi t_0} dt \\ &= e^{-j \xi t_0} (G_\phi s)(b - t_0, \xi) \end{aligned} \quad (2.21)$$

L'équation (2.21) signifie que la translation de  $t_0$  de la fonction  $s(t)$  dans le domaine temporel entraîne dans le domaine temps-fréquence une translation de  $t_0$  suivant l'échelle des temps, pendant que la fréquence reste identique.

Un changement de phase dans le domaine temps-fréquence, directement proportionnel à la translation suivant le temps ( $t_0$ ) peut également être observé.

- Translation fréquentielle :

Si  $s_0(t) = e^{jw_0 t} s(t)$  alors :

$$\begin{aligned} (G_\phi s)(b, \xi) &= \int_{-\infty}^{+\infty} s(t) e^{jw_0 t} \phi(t-b) e^{-j \xi t} dt \\ &= (G_\phi s)(b, \xi - w_0) \end{aligned} \quad (2.22)$$

L'équation (2.22) signifie que l'amplitude et la phase de la transformée de Fourier à fenêtre glissante de  $s_0(t)$  sont identiques à celles de  $s(t)$ . Cependant, la localisation de la fonction  $s(t)$  dans le domaine temps-fréquence est translatée de  $w_0$  suivant l'échelle des fréquences.

### 2.1.2.3 STFT discrète

Comme pour le cas de la transformée de Fourier discrète, l'intégrale (2.15) peut être évaluée sous forme de somme en échantillonnant convenablement la fonction  $s(t)$  ainsi que la fonction de fenêtrage  $\phi(t)$ . Comme pour le cas de la transformée de Fourier discrète, l'intégrale (2.15) peut être évaluée sous forme de somme en échantillonnant convenablement la fonction  $s(t)$  ainsi que la fonction de fenêtrage  $\phi(t)$ .

La transformée de Fourier à fenêtre glissante peut être représentée de la façon suivante :

$$(G_\phi s)(b, \xi) \approx h \sum_{k=0}^{N-1} s(t_k) \phi(t_k - b_n) e^{-j \xi_n t_k} \quad (2.23)$$

Où

$$t_k = bk = kh \quad k = 0, \dots, N-1$$

Et

$$\xi_n = \frac{2\pi n}{Nh} \quad n = -N/2, \dots, N/2$$

En particulier quand  $h = 1$ , on obtient :

$$(G_{\phi}s)(n, \xi_n) \approx \sum_{k=0}^{N-1} s(k)\phi(k-n)e^{-j\frac{2\pi kn}{N}} \quad (2.24)$$

#### 2.1.2.4 Bilan sur la STFT

Dans la méthode de Fourier, les “fonctions de bases” sont totalement concentrées en fréquence (impulsions de Dirac) et totalement réparties dans le temps (sinusoïdes non amorties s'étendant de  $-\infty$  à  $+\infty$ ). C'est une autre façon d'expliquer que le passage dans l'espace de Fourier donne le maximum d'informations sur la répartition des fréquences mais perd entièrement les informations relatives au temps.

Dans la méthode de la transformée de Fourier à fenêtre glissante, les informations temps-fréquences restent couplées par un compromis, limité par la relation d'incertitude, sur la localisation à la fois en temps et en fréquence. Ceci met en évidence les avantages de cette méthode sur celle de Fourier.

Cependant, la méthode de la transformée de Fourier à fenêtre glissante, présente l'inconvénient majeur d'avoir une fenêtre de longueur fixe. Il n'est donc pas possible d'analyser simultanément des phénomènes dont les échelles de temps sont différentes. De plus, le calcul de la STFT doit s'effectuer pour chaque changement de taille de la fenêtre d'où une charge de calcul importante.

Une autre méthode d'analyse, qui ne privilégie aucune échelle particulière mais qui généralise à toutes les échelles, l'analyse locale des fréquences obtenues par la STFT, est donc nécessaire. En 1982, Morlet ouvre la voie conduisant à la solution en construisant la transformation en ondelettes [30].

### 2.1.3. Les Ondelettes

#### 2.1.3.1 Introduction

Nous savons bien que la transformée de fourrier a fenêtre STFT présente un inconvénient majeur de résolution temporelle fixe.

La transformée en ondelettes va nous permettre de pallier cet inconvénient

Les ondelettes, famille de fonctions déduites d'une même fonction, appelée ondelette mère, par opérations de translations et dilatations, ont trouvé, de par la puissance de leur théorie, des applications dans de nombreux domaines aussi variés que les mathématiques (analyse [31], probabilités [8], fractales [33]), le traitement du signal (compression [34], astronomie [35], sismique [36]), la physique (mécanique quantique [37], turbulence [38]). En effet, cet outil permet la représentation de fonctions de  $L^2$ , dans une base bien localisée en espace et en fréquence, offrant les avantages de l'analyse de Fourier et s'affranchissant des inconvénients du manque de localisation de cette dernière.

#### 2.1.3.2 Transformée en ondelettes continue

A partir d'une fonction de base  $\psi$ , appelée ondelette mère, on construit une famille de fonctions analysantes :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a > 0 \quad (2.25)$$

où le paramètre  $a$  est le facteur d'échelle (dilatation), et  $b$  est le paramètre de translation.

On définit alors les coefficients de la transformée en ondelettes d'un signal  $s(t)$ , comme étant les produits scalaires :

$$C_s(a,b) := \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt \quad (2.26)$$

Ces coefficients mesurent, en un certain sens, les fluctuations du signal  $s(t)$  autour du point  $t = b$ , à l'échelle fournie par  $a$ . D'après l'équation (2.25), en

diminuant  $a$ , le support de  $\psi_{a,b}$  réduit en temps et donc couvre une plage en fréquence plus grande et vice versa. Donc  $1/a$  est proportionnel à une fréquence.

Pour un facteur d'échelle assez grand, la représentation des coefficients d'ondelettes en fonction de  $b$ , la position, donne une représentation de "la forme générale de la fonction". Par contre un facteur d'échelle faible correspond à une représentation des singularités. Cette propriété de "microscope" est très utile pour l'étude de la régularité d'une fonction [39].

La transformée en ondelettes, comme il sera démontré par la suite, est un opérateur linéaire, invariant par translation, et par dilatation. Quelle que soit l'échelle et quel que soit l'endroit, l'analyse du signal se fait avec la même fonction. La transformée en ondelettes d'un signal n'est pas unique, elle dépend de l'ondelette mère utilisée (la suite montrera qu'il existe plusieurs types d'ondelettes).

On notera que la norme de  $\psi_{a,b}$  est conservée lors du changement de facteur d'échelle :

$$\begin{aligned} \|\psi_{a,b}\| &= \int_{-\infty}^{+\infty} \frac{1}{a} \left| \psi\left(\frac{t-b}{a}\right) \right|^2 dt \\ &= \frac{1}{a} \int_{-\infty}^{+\infty} |\psi(x)|^2 a dx \\ &= \|\psi\|^2 \end{aligned} \tag{2.27}$$

L'ondelette mère  $\psi(t)$  devra avoir une bonne localisation, donc à l'infini, une convergence rapide vers 0, et devra être oscillante. On demande que l'intégrale de  $\psi(t)$  soit nulle et qu'il en soit de même pour les  $m$  premiers moments de  $\psi$ . Cela s'écrit :

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \quad \text{pour } 0 \leq k \leq m \tag{2.28}$$

$$\int_{-\infty}^{+\infty} \psi(t) dt = \hat{\psi}(0) = 0 \tag{2.29}$$

On peut montrer [16] que si la fonction analysante (l'ondelette) est correctement choisie, la transformation en ondelettes est inversible. Le signal  $s(t)$

peut être reconstruit après double intégration suivant le facteur d'échelle  $a$  et le paramètre de translation  $b$  :

$$s(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} C_s(a,b) \psi_{a,b}(t) da db \quad (2.30)$$

Cette possibilité reste théorique car le calcul n'est possible que numériquement et sa convergence peut être très lente. Le coefficient  $C_\psi$  est une constante qui dépend du choix de l'ondelette et est donné par :

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(w)|^2}{|w|} dw < +\infty \quad (2.31)$$

Cette condition se ramène le plus souvent à la condition exprimée par l'équation (2.29).

### 2.1.3.3 Fenêtre temps-fréquence

La définition en temps pour le centre  $t^*$  et l'écart type  $\Delta\psi$  de la fenêtre reste inchangée par rapport au paragraphe (1.2.1) avec  $\phi(t)$  remplacée par  $\psi(t)$ .  $C_s(a,b)$  contient les informations du signal  $s(t)$  sur l'intervalle de temps :

$$[at^* + b - a\Delta\psi, at^* + b + a\Delta\psi] \quad (2.32)$$

La définition en fréquence est différente puisque  $\hat{\psi}(0) = 0$  ( $\hat{\psi}(w)$  a les caractéristiques d'un filtre passe-bande). Il existe donc deux centres et deux écarts types pour  $\hat{\psi}(w)$ . Dans la suite, seules les fréquences positives seront considérées. Le centre  $w^*$  et l'écart type  $\Delta\hat{\psi}$  sont alors définis sur l'échelle des fréquences positives de la façon suivante :

$$w^* := \frac{\int_0^{+\infty} w |\hat{\psi}(w)|^2 dw}{\int_0^{+\infty} |\hat{\psi}(w)|^2 dw} \quad (2.33)$$

$$\Delta \hat{\psi} := \left\{ \frac{\int_0^{+\infty} (w - w^*)^2 |\hat{\psi}(w)|^2 dw}{\int_0^{+\infty} |\hat{\psi}(w)|^2 dw} \right\}^{\frac{1}{2}} \quad (2.34)$$

Le principe d'incertitude est le suivant :

$$\Delta \psi \cdot \Delta \hat{\psi} > \frac{1}{2} \quad (2.35)$$

En appliquant l'identité de Parseval (2.3) à l'équation (2.26), afin d'obtenir les renseignements sur l'intervalle en fréquence occupé par la fenêtre, on obtient :

$$\begin{aligned} C_s(a, b) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt \\ &= \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{+\infty} \hat{s}(w) \overline{\hat{\psi}(aw)} e^{jbw} dw \end{aligned} \quad (2.36)$$

D'après l'équation (2.36), la fenêtre occupe l'intervalle fréquentiel suivant :

$$\left[ \frac{1}{a}(w^* - \Delta \hat{\psi}), \frac{1}{a}(w^* + \Delta \hat{\psi}) \right] \quad (2.37)$$

La surface occupée par la fenêtre =  $2a\Delta\psi \times \frac{2}{a}\Delta\hat{\psi} = 4\Delta\psi \cdot \Delta\hat{\psi} = \text{constant}$ .

La Figure 2.4 présente la notion de fenêtrage temps-fréquence pour la transformée en Ondelettes (l'ondelette de Morlet est présentée à l'intérieur de ces fenêtres). Plus la fréquence croît, plus la largeur de la fenêtre diminue en temps et vice versa. Pour une fréquence fixée, ( $1/a_0$  par exemple), les intervalles en temps et fréquences occupés par la fenêtre sont identiques. (Rappel : la fenêtre temps-fréquence pour la transformée de Fourier à fenêtre glissante est fixe quelle que soit la fréquence).

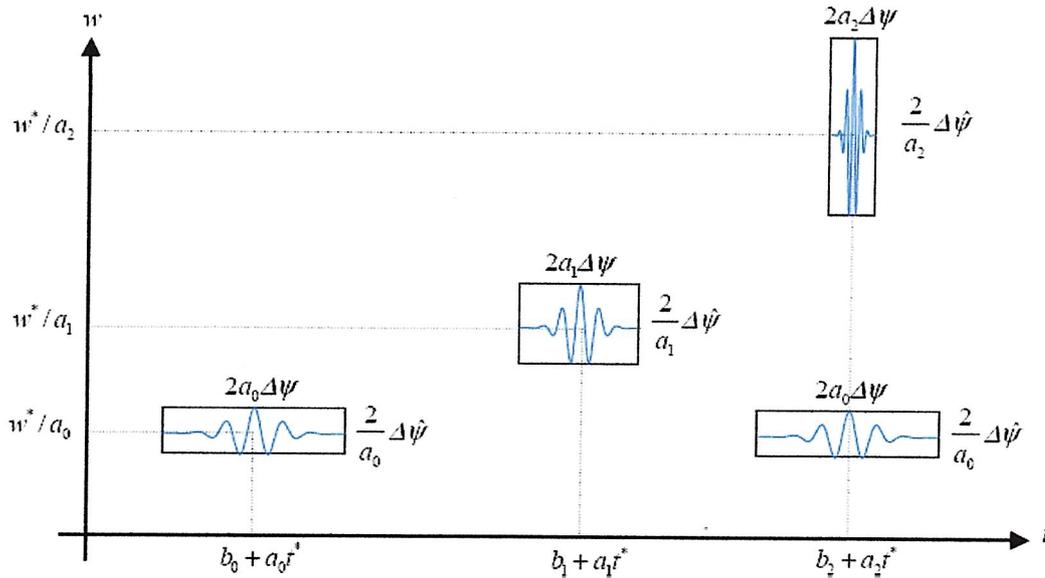


Figure 2.4 Fenêtrage temps-fréquence pour la transformée en ondelettes.

### 2.1.3.4 Quelques propriétés

- Linéarité :

La transformée en ondelettes est linéaire, c'est à dire :

$$C_{\alpha s_1 + \beta s_2}(a, b) = \alpha C_{s_1}(a, b) + \beta C_{s_2}(a, b) \quad (\alpha, \beta) \in \mathbb{V}^2, (f_1, f_2) \in L^2 \quad (2.38)$$

- Translation :

Une des propriétés importante de la transformée en ondelettes continue est l'invariance en translation :

$$C_{\tau_{b_0} s}(a, b) = C_s(a, b - b_0) \quad (2.39)$$

où  $\tau_{b_0} s$  désigne la translation de  $s$  par le vecteur  $b_0$ , c'est-à-dire  $(\tau_{b_0} s)(t) = s(t - b_0)$

- Dilatation :

La transformée en ondelettes a également une propriété de dilatation. En effet, si la famille d'ondelettes est déterminée par l'équation (2.25) (avec une normalisation  $L^2$ ), on obtient la relation suivante :

$$C_{\delta_k s}(a, b) = k^{-\frac{1}{2}} C_s(ka, kb) \quad \text{avec} \quad \delta_k s(t) = s(kt) \quad (2.40)$$

Afin de s'affranchir du facteur de dilatation  $k^{-\frac{1}{2}}$ , la normalisation  $L^1$  de la famille d'ondelettes peut s'avérer extrêmement importante. En effet, ceci conduit à une propriété d'invariance en dilatation :

$$C_{\delta_k s}(a, b) = C_s(ka, kb) \quad (2.41)$$

### 2.1.3.5 Transformée en ondelettes discrète

L'intérêt de ce paragraphe est de présenter très brièvement la transformée en ondelettes discrète pour montrer la relation avec la transformée en ondelettes continue.

Les valeurs discrètes du facteur d'échelle  $a$  et du paramètre de translation  $b$  seront considérées sous la forme :  $a = 2^{-j}$  et  $b = k2^{-j}$ . Avec ces valeurs de  $a$  et  $b$ , l'équation (2.26) devient :

$$C_s(2^{-j}, k2^{-j}) = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} s(t) \psi(2^j t - k) dt \quad (2.42)$$

Si la fonction  $s(t)$  est discrétisée, en supposant une période d'échantillonnage égale à 1, pour des raisons de simplicité, l'équation (2.42) s'écrit alors :

$$C_s(2^{-j}, k2^{-j}) = 2^{\frac{j}{2}} \sum_n s(n) \psi(2^j n - k) \quad (2.43)$$

Pour calculer la transformée en ondelettes d'une fonction en différents points dans le plan temps-échelle, il n'est pas nécessaire de connaître les valeurs de la fonction sur l'ensemble de l'axe des temps. Il suffit simplement de connaître les valeurs de la fonction aux temps où l'ondelette est différentes de zéro.

D'après l'équation (2.43), la transformée en ondelettes discrète d'une fonction translatée dans le temps est différente de la transformée en ondelettes discrète de la fonction originale. Pour l'expliquer, supposons :

$$s_m(t) = s(t - t_m) \quad (2.44)$$

La transformée en ondelettes de  $s_m(t)$  s'écrit : La transformée en ondelettes de  $s_m(t)$  s'écrit :

$$C_s(2^{-j}, k2^{-j}) = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} s_m(t) \psi(2^j t - k) dt \quad (2.45)$$

$$\begin{aligned}
&\approx 2^{\frac{j}{2}} \sum_n s(n-m)\psi(2^j n - k) \\
&= 2^{\frac{j}{2}} \sum_n s(n)\psi[2^j n - (k - m2^j)] \\
&\approx C_s (2^{-j}, (k - m2^j)2^{-j})
\end{aligned} \tag{2.46}$$

Rappel : La transformée de Fourier d'une fonction translatée dans le temps apparaît comme un changement de phase dans le domaine fréquentiel. La transformée de Fourier à fenêtre glissante d'une fonction translatée dans le temps apparaît comme la translation dans le temps d'un même facteur de translation.

## 2.2. Analyse multirésolution et algorithmes

### 2.2.1. Présentation de l'analyse multirésolution

L'idée de l'analyse multirésolution a été développée par Meyer [26, 41] et Mallat [42, 43]. Dans [44], L'analyse multirésolution est une théorie définissant des opérateurs linéaires permettant d'analyser un signal à différentes échelles. Nous pourrions dire que la construction d'une multirésolution permet de regarder un signal de "très près" ou de "très loin". Ce Zoom est effectué à l'aide d'une fonction d'échelle, qui se dilate à travers les échelles. Le signal projeté sur cette fonction donne une représentation de notre signal d'origine à l'échelle supérieure. Cette représentation (coefficients de projection) provoque un zoom arrière de notre signal d'origine, d'où le terme approximation. Afin de reconstruire notre signal, à partir des coefficients d'approximation, nous devons également projeter notre signal originel sur un espace perpendiculaire (conservation de toute l'information). La fonction générant ce deuxième espace vectoriel sera une ondelette. Finalement, le signal est projeté sur une fonction d'échelle créant une approximation du signal et sur une ondelette pour récupérer l'information perdue lors de la première projection. Cette deuxième projection contient les détails du signal d'origine.

Les bases  $\phi_{j,k}(t)$  et  $\psi_{j,k}(t)$  sont respectivement nommées fonctions d'échelles et ondelettes.

### 2.2.1.1 Définition d'une analyse multirésolution

Une analyse multirésolution de  $L^2(\mathbb{R})$  est, par définition, une suite croissante  $V_j$ , avec  $j \in \mathbb{Z}$ , de sous espaces vectoriels fermés de  $L^2(\mathbb{R})$  ayant les propriétés suivantes :

- $V_j \subset V_{j+1}$  ; pour tout  $j, V_j$  est un sous espace de  $V_{j+1}$
- $s(t) \in V_j \Leftrightarrow s(2t) \in V_{j+1}$  ;  $V_{j+1}$  est l'image de  $V_j$  par une dilatation d'un facteur 2.
- $s(t) \in V_j \Leftrightarrow s(t - 2^{-j}k) \in V_j, \forall k \in \mathbb{Z}$  ;  $V_j$  est invariant par translation de  $2^{-j}$ .
- $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$  ; la réunion des  $V_j$  est dense dans  $L^2$ .
- $\bigcap_{j \in \mathbb{Z}} V_j = \{0\} = \lim_{j \rightarrow -\infty} V_j$  ; l'intersection de  $V_j$  est réduite à 0 dans  $L^2$ .
- il existe une fonction d'échelle  $\phi$  et  $V_0$  telle que la suite  $\{\phi(t - k)\}, k \in \mathbb{Z}$ , soit une base de Riesz pour  $V_0$

les sous espaces  $W_j$  sont définis comme étant les complément de  $V_j$  dans  $V_{j+1}$

$$V_j \oplus W_j = V_{j+1} \tag{2.47}$$

$$V_j \cap W_j = \{0\} \tag{2.48}$$

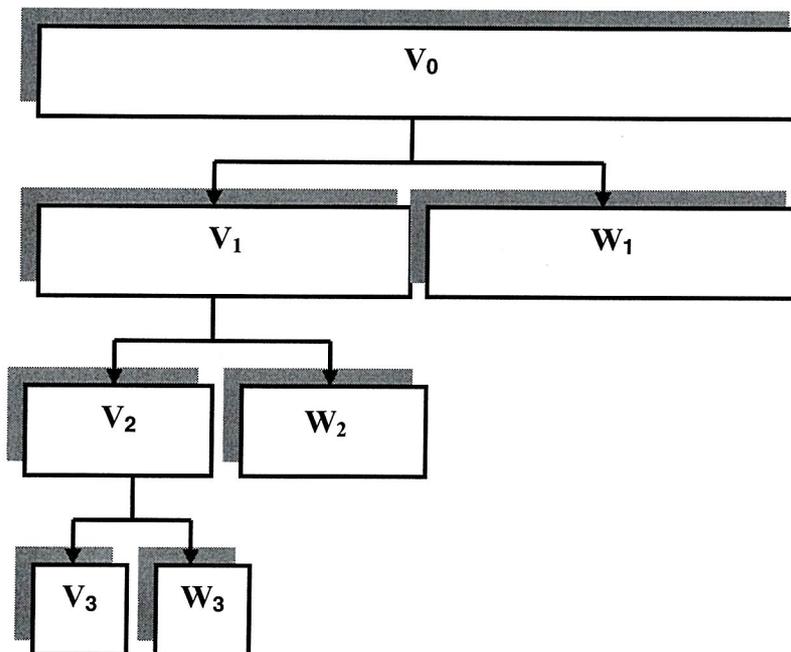


Figure 2.5 : représente la nature hiérarchique de l'analyse multirésolution .

### 2.2.1.2 Relations de reconstruction et de décomposition

D'après la définition de l'analyse multirésolution,  $\phi \in V_0 \subset V_1$  et  $\psi \in W_0 \subset V_1$ , il existe deux séquences  $\{h_1[k]\} \in l^2$  et  $\{g_1[k]\} \in l^2$  telles que la fonction d'échelle  $\phi(t)$  et l'ondelette associée  $\psi(t)$ , à un niveau donné, peuvent être construites à partir de la fonction d'échelle au niveau juste au dessus. Ces équations seront nommées : relations de reconstruction.

$$\phi(t) = \sum_k h_1[k] \phi(2t - k) \quad (2.49)$$

$$\psi(t) = \sum_k g_1[k] \phi(2t - k) \quad (2.50)$$

En général,  $\forall j \in \mathbb{Z}$ , la relation entre les sous espaces  $V_j, W_j$  et  $V_{j+1}$  est engendrée par les deux équations suivantes :

$$\phi(2^j t) = \sum_k h_1[k] \phi(2^{j+1} t - k) \quad (2.51)$$

$$\psi(2^j t) = \sum_k g_1[k] \phi(2^{j+1} t - k) \quad (2.52)$$

En prenant la transformée de Fourier des équations (2.49) et (2.50), on obtient :

$$\hat{\phi}(w) = H(z) \phi\left(\frac{w}{2}\right) \quad (2.53)$$

$$\hat{\psi}(w) = G(z) \hat{\phi}\left(\frac{w}{2}\right) \quad (2.54)$$

Où :

$$H(z) := \frac{1}{2} \sum_k h_1[k] z^k, \quad z = e^{\frac{-jw}{2}} \quad (2.55)$$

$$G(z) := \frac{1}{2} \sum_k g_1[k] z^k, \quad z = e^{\frac{-jw}{2}} \quad (2.56)$$

On admettra que  $\hat{\phi}(0) = 1$  (pour plus de détails voir [45]). En appliquant ce résultat à l'équation (2.53), on déduit :

$$H(1) = \frac{1}{2} \sum_k h_1[k] = 1 \quad \Rightarrow \quad \sum_k h_1[k] = 2 \quad (2.57)$$

L'ondelette se comportant comme un filtre passe-bande,  $\hat{\psi}(0) = 0$  (voir équation (2.28)), à l'aide de l'équation (2.54), on obtient :

$$G(1) - \frac{1}{2} \sum_k g_1[k] = 0 \quad \Rightarrow \quad \sum_k g_1[k] = 0 \quad (2.50)$$

L'analyse multirésolution permet d'écrire :  $V_1 = V_0 + W_0$ ,  $\phi(2t) \in V_1, \phi(2t-1) \in V_1$ , il existe donc deux séquences  $\{h_2[k]\} \in l^2$  et  $\{g_2[k]\} \in l^2$  telles que :

$$\phi(2t) = \sum_k [h_2[-2k]\phi(t-k) + g_2[-2k]\psi(t-k)] \quad (2.59)$$

$$\phi(2t-1) = \sum_k [h_2[1-2k]\phi(t-k) + g_2[1-2k]\psi(t-k)] \quad (2.60)$$

En combinant ces deux relations, on obtient :

$$\phi(2t-n) = \sum_k [h_2[n-2k]\phi(t-k) + g_2[n-2k]\psi(t-k)] \quad (2.61)$$

Dans le cas général :

$$\phi(2^{j+1}t-n) = \sum_k [h_2[n-2k]\phi(2^j t-k) + g_2[n-2k]\psi(2^j t-k)] \quad (2.62)$$

Cette équation est nommée : relation de décomposition.

### 2.2.1.3 Algorithmes

Deux algorithmes principaux ont été mis en évidence : l'algorithme à trous [46] et l'algorithme de Mallat [42, 43, 47]. Le premier concerne des analyses multirésolution non orthogonales, le second est pratiquement le seul utilisé dans le cas des analyses multirésolution orthogonales et biorthogonales. Seul l'algorithme de Mallat

### 2.2.1.3.1 Algorithme de décomposition

L'algorithme de décomposition (analyse) est très employé en traitement du signal. Il est aussi bien utilisé pour la compression de données .L'algorithme est basé sur la relation de décomposition (2.62).

Si la fonction des  $\phi_{j,k}$  et  $\psi_{j,k}$ ,  $j$  et  $k$  décrivant  $R^2$ , est une base orthonormée de  $L^2(R)$ , alors un signal  $s(t) \in L^2(R)$  se décompose suivant les différentes échelles  $j$  en une fonction d'approximation  $x_j(t)$  et une fonction de détail  $y_j(t)$  de la façon suivante :

$$x_{j+1}(t) \in V_{j+1} \quad \Rightarrow \quad x_{j+1} = \sum_k a_{j+1}[k] \phi_{j+1,k}(t) \quad (2.63)$$

$$x_j(t) \in V_j \quad \Rightarrow \quad x_j(t) = \sum_k a_j[k] \phi_{j,k}(t) \quad (2.64)$$

$$y_j(t) \in W_j \quad \Rightarrow \quad y_j(t) = \sum_k d_j[k] \psi_{j,k}(t) \quad (2.65)$$

Avec

$$a_j[k] = 2^{j/2} \int_{-\infty}^{+\infty} s(t) \overline{\phi(2^j t - k)} dt \quad \text{et} \quad d_j[k] = 2^{j/2} \int_{-\infty}^{+\infty} s(t) \overline{\psi(2^j t - k)} dt \quad (2.66)$$

L'analyse multirésolution exige que :

$$V_{j+1} = V_j + W_j \quad (2.67)$$

Donc :

$$x_{j+1}(t) = x_j(t) + y_j(t) \quad (2.68)$$

$$\sum_k a_{j+1}[k] \phi(2^{j+1} t - k) = \sum_k a_j[k] \phi(2^j t - k) + \sum_k d_j[k] \psi(2^j t - k) \quad (2.69)$$

On substitue la relation de décomposition suivante

$$\phi(2^{j+1} t - n) = \sum_k \{h_2[n - 2k] \phi(2^j t - k) + g_2[n - 2k] \psi(2^j t - k)\} \quad (2.70)$$

Dans l'équation (2.69) pour obtenir une équation où toutes les bases sont à l'échelle  $j$ . Après avoir interchangé l'ordre des sommations et comparé les coefficients de  $\phi(2^j t - k)$  et  $\psi(2^j t - k)$  de chaque côté de l'équation, la décomposition suivant l'algorithme de Mallat [42, 43] est obtenue :

$$a_j[k] = \sum_n h_2[n-2k]a_{j+1}[n] \quad (2.71)$$

$$d_j[k] = \sum_n g_2[n-2k]a_{j+1}[n] \quad (2.72)$$

Les équations (2.71) et (2.72) montrent que les coefficients de la fonction d'échelle et de l'ondelette à l'échelle  $j$ , se calculent à partir de ceux de l'échelle immédiatement supérieure. En répétant ce processus, on obtient la décomposition en cascade suivante :

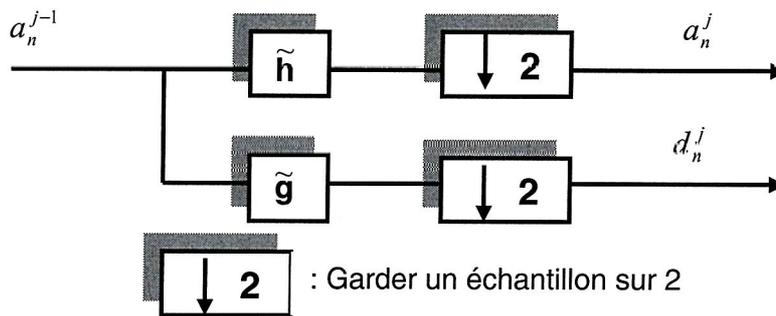


Figure 2.6 : Décomposition en cascade.

### 2.2.1.3.2 Algorithme de reconstruction

Il existe une unique transformation en ondelettes discrète inverse (synthèse), telle que le signal original peut être reconstruit parfaitement à partir de ses composantes à différentes échelles. L'algorithme de reconstruction est basé sur les relations de reconstruction (2.51) et (2.52). En substituant ces deux équations dans (2.68), on obtient :

$$\sum_k a_j[k] \sum_n h_1[n-2k] \phi(2^{j+1}t-n) + \sum_k d_j[k] \sum_n g_1[n-2k] \phi(2^{j+1}t-n) = \sum_n a_{j+1}[n] \phi(2^{j+1}t-n) \quad (2.73)$$

En comparant les coefficients de  $\phi(2^{j+1}t-n)$  dans les deux membres de l'équation (2.73), la reconstruction suivant l'algorithme de Mallat est obtenue :

$$a_{j+1}[n] = \sum_k \{h_1[n-2k]a_j[k] + g_1[n-2k]d_j[k]\} \quad (2.74)$$

L'algorithme de reconstruction peut être schématisé de la façon suivante :

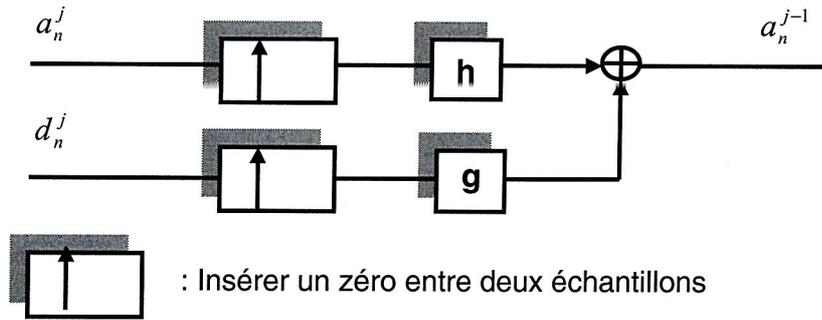


Figure 2.7 : Reconstruction unidimensionnelle par ondelettes.

### 2.2.2. Bilan

L'analyse multirésolution a permis d'introduire et de mieux comprendre les algorithmes rapides de décomposition et de reconstruction. Les relations de décomposition et de reconstruction sont essentielles pour le développement de ces algorithmes.

Pour connaître les filtres  $h_1[k]$ ,  $g_1[k]$ ,  $h_2[k]$  et  $g_2[k]$  dans le cas de bases orthogonales, il suffit de connaître seulement le filtre  $h_1[k]$  puisque les trois autres filtres se calculent à partir de ce dernier :

$$g_1[k] = (-1)^k h_1[1-k] \tag{2.75}$$

$$h_2[k] = \frac{1}{2} h_1[-k] \tag{2.76}$$

$$g_2[k] = \frac{1}{2} (-1)^k h_1[k+1] \tag{2.77}$$

Les algorithmes de décomposition et de reconstruction s'appliquent à tous les types de fonctions d'échelles et d'ondelettes (orthogonales, biorthogonales, semiorthogonales)

Un signal  $s(t) \in V_0$ , correspondant à l'échelle  $J_0$  (Figure 2.5), s'écrira à l'aide de l'analyse multirésolution ( $J \in \mathbb{R}^2$  étant l'échelle la plus basse) :

$$s(t) = \sum_{k \in \mathbb{Z}} a_J[k] \phi(2^J t - k) + \sum_{j=J}^{J_0} \sum_{k \in \mathbb{Z}} d_j[k] \psi(2^j t - k) \tag{2.78}$$

### 2.3. Exemples d'ondelettes

Les premières ondelettes qui sont nées des travaux de Meyer [41] et Mallat [42], sont les ondelettes orthogonales. Ils existent un certain nombre de familles d'ondelettes orthogonales couramment utilisées. Les plus connues sont sans doute les ondelettes de Daubechies [40].

Les familles d'ondelettes non orthogonales ont reçu l'appellation d'ondelettes biorthogonales. En effet, le préfixe "bi" rappelle que deux bases d'ondelettes sont utilisées, une pour la décomposition (la base duale) et une pour la reconstruction [48, 49]. Le fait que la fonction d'échelle et la fonction ondelette soient biorthogonales ne signifie pas nécessairement que l'analyse multirésolution ne soit pas orthogonale. Si les fonctions d'échelles et ondelettes biorthogonales engendrent une analyse multirésolution orthogonale, on les appelle semiorthogonales [45].

#### 2.3.1. Ondelettes orthogonales

Ondelette de Haar :

Un premier exemple d'ondelette orthogonale est l'ondelette de Haar. La fonction d'échelle dans ce cas est  $\phi_{haar}(t) = X_{[0,1]}(t)$  et l'ondelette correspondante :

$$\psi_{haar}(t) = X_{[0,1/2]}(t) - X_{[1/2,1]}(t) \quad (2.79)$$

Dans la pratique, l'ondelette de Haar n'est pas beaucoup utilisée en raison de ses discontinuités.

Ondelette de Shannon :

L'ondelette de Shannon :

$$\psi_{SH} = \frac{\sin 2\pi t - \sin \pi t}{\pi t} \quad (2.80)$$

dont la fonction d'échelle est  $\phi_{SH} = \text{sinc}(\pi t)$ , n'est pas beaucoup utilisée en raison de sa très faible décroissance à l'infini. En effet,  $\phi_{SH}$  est très mal localisé en temps ( $\Delta\phi_{SH} = \infty$ ). La raison de cette mauvaise localisation en temps vient du fait que dans le domaine fréquentiel,  $\hat{\phi}_{SH}(\omega)$  comporte deux discontinuités à  $-\pi$  et  $\pi$ . En conséquence, dans le domaine temporel, la fonction décroît en  $1/t$  et donc  $\Delta\phi_{SH} = \infty$

Ondelette de Meyer :

Meyer [26, 41] a construit une fonction d'échelle (Figure 2. 8) de telle sorte que sa transformée de Fourier soit lisse aux endroits de discontinuités de  $\hat{\phi}_{SH}(\omega)$  (voir ondelette de Shannon). En temps, cela se traduit par une décroissance plus rapide à l'infini par rapport à Shannon. La fonction d'échelle et l'ondelette sont symétriques respectivement par rapport à 0 et -1/2. Les ondelettes de Meyer sont des ondelettes indéfiniment dérivables, de support infini. Leur implémentation se fait plutôt dans le domaine fréquentiel.

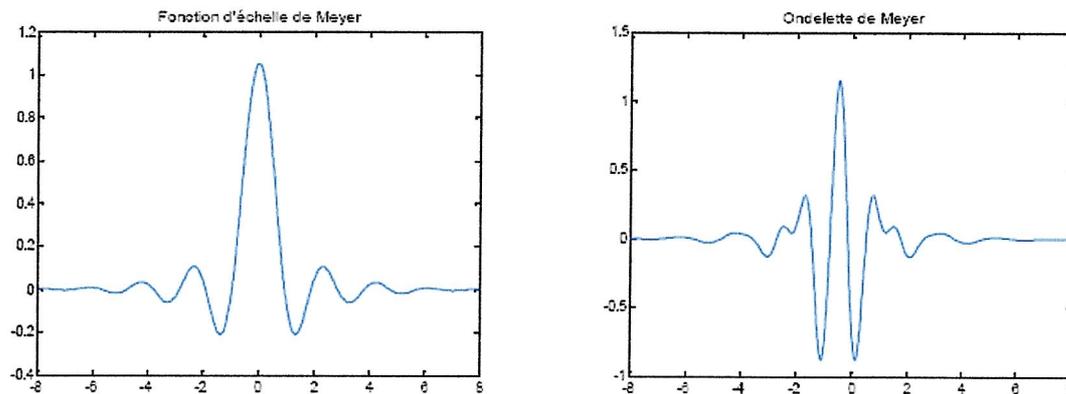


Figure 2.8 : Fonction d'échelle et ondelette de Meyer.

Ondelette de Daubechies :

Les ondelettes de Daubechies [40] sont probablement les plus utilisées en ce qui concerne les ondelettes orthogonales. Elles sont à support compact ( les filtres  $H$  et  $G$  ont une réponse impulsionnelle finie donc les filtres  $h_1$  et  $g_1$  sont finis). Ces ondelettes seront notées  $dbN$ , où  $db$  est le symbole donné pour Daubechies, et  $N$  est le nombre de moments nuls de l'ondelette. Les ondelettes de Daubechies sont supportées sur un intervalle de longueur  $2N-1$ . Ces ondelettes présentent l'inconvénient de ne pas être symétriques ou antisymétriques, excepté quand  $N = 1$  ce qui correspond à l'ondelette de Haar. La Figure 2.9 représente les fonctions d'échelles et ondelettes pour  $N = 2$ ,  $N = 4$ , et  $N = 8$ .

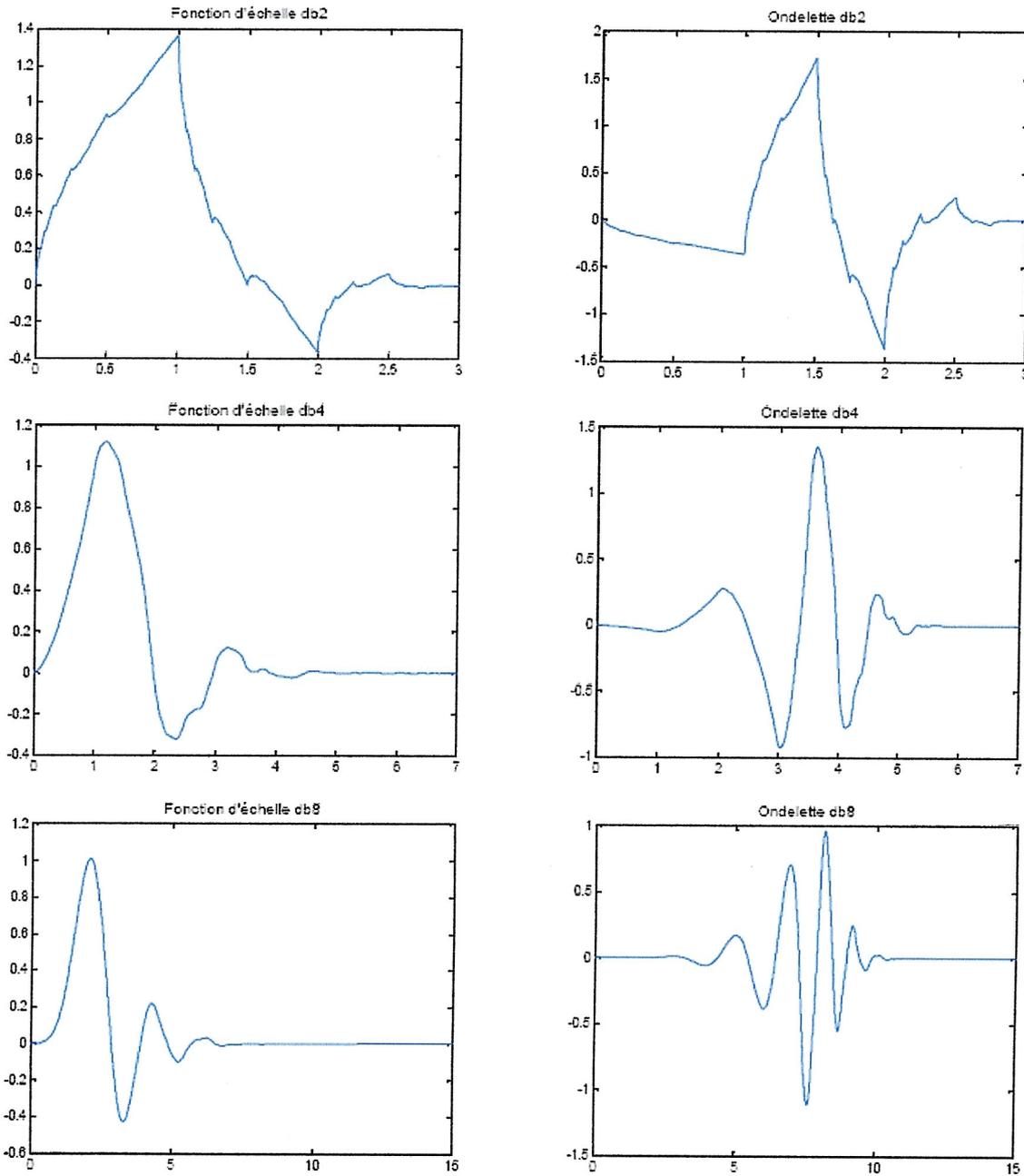


Figure 2.9 : Fonctions d'échelles et ondelettes de Daubechies pour  $N = 2, 4, 8$ .

**Symlets :**

Pour obtenir une ondelette symétrique ou antisymétrique, le filtre  $h_1$  doit être symétrique ou antisymétrique par rapport au centre de son support ( $\hat{h}_1(\omega)$  a une phase linéaire). Les symlets sont des ondelettes de Daubechies construites de telle

sorte que la phase de  $\hat{h}_1(\omega)$  soit la plus linéaire possible. Le support des symlets est  $2N+1$ . La Figure 2.10, représente la fonction d'échelle et l'ondelette pour  $N = 8$ . Une meilleure symétrie par rapport à l'ondelette de Daubechies ( $N = 8$ ) peut être remarquée.

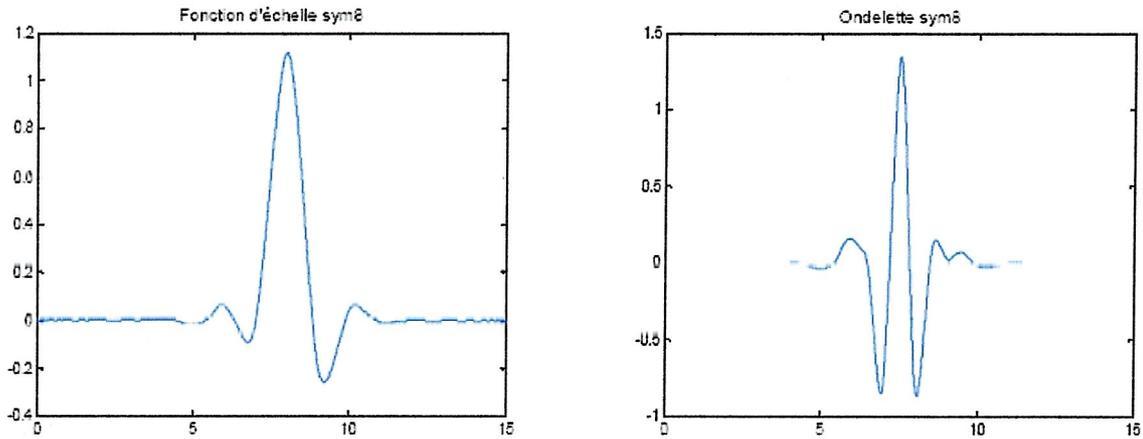


Figure 2.10 : Fonction d'échelle et ondelette Symlets pour  $N = 8$ .

Coiflets :

Pour une application en analyse numérique Coifman a demandé à Daubechies [50] de construire une famille d'ondelettes avec  $N$  moments nuls et un support de taille minimum, et dont la fonction d'échelle vérifie :

$$\int_{-\infty}^{+\infty} \phi(t) dt = 1 \quad \int_{-\infty}^{+\infty} t^k \phi(t) dt = 0 \quad \text{pour } 1 \leq k \leq N \quad (2.81)$$

Le résultat est l'ondelette coiflets dont la taille du support est  $3N-1$  au lieu de  $2N-1$  pour une ondelette de Daubechies. La Figure 2.11, représente la fonction d'échelle et l'ondelette pour  $N = 5$ .

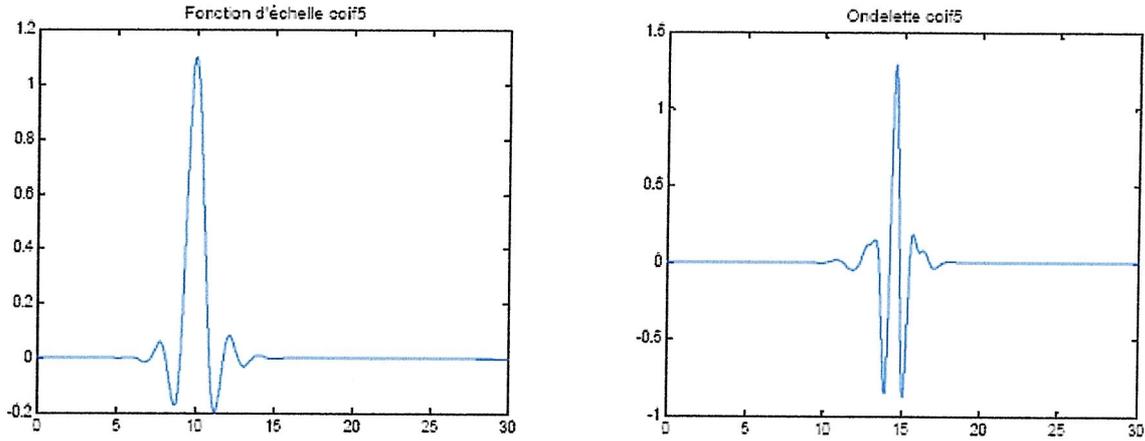


Figure 2.11 : Fonction d'échelle et ondelette Coiflets pour  $N = 5$ .

## 2.4. Bases d'ondelettes a deux dimensions

### 2.4.1. Principe

Une base d'ondelettes orthonormales à deux dimensions de  $L^2(\mathbb{R}^2)$  est construite par produit tensoriel d'une fonction d'échelle  $\phi$  et de l'ondelette correspondante  $\psi$  générées par la base d'ondelette orthonormale de  $L^2(\mathbb{R}^2)$ .

Par définition,  $V_j^2 = V_j \oplus V_j$ , et  $W_j^2$  est l'espace de détail au complément orthogonal de l'espace d'approximation  $V_j^2$  dans  $V_{j+1}^2$  :

$$V_j^2 = V_j \oplus V_j \tag{2.82}$$

$$V_j^2 \oplus W_j^2 = V_{j+1}^2 \tag{2.83}$$

L'équation (2.83) s'écrit également :

$$(V_j \otimes V_j) \oplus W_j^2 = V_{j+1} \otimes V_{j+1} \tag{2.84}$$

D'après l'équation (2.47),  $V_{j+1} = V_j + W_j$ . ainsi après simplification (2.83) s'écrit :

$$W_j^2 = (V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j) \tag{2.85}$$

Sachant que la famille des  $\phi_{j,k}$ , et  $\psi_{j,k}$   $j$  et  $k$  décrivant  $\mathbb{R}^2$ , est une base orthonormée respectivement de  $V_j$  et  $W_j$ , d'après les équations (2.82) et (2.83), on déduit :

$$\{\phi_{j,n}(x)\phi_{j,m}(y)\}_{(m,n)\in\mathbb{Z}^2} \text{ est une base orthonormale de } V_j^2 \quad (2.86)$$

Et

$$\{\phi_{j,n}(x)\psi_{j,m}(y), \psi_{j,n}(x)\phi_{j,m}(y), \psi_{j,n}(x)\psi_{j,m}(x)\}_{(n,m)\in\mathbb{Z}^2} \quad (2.87)$$

Est une base orthonormale de  $W_j^2$

Connaissant les formes analytiques de  $\phi_{j,k}(x)$  et  $\psi_{j,k}(x)$ , on obtient pour l'analyse multirésolution à 2D, une fonction d'échelle à 2D (approximation) et trois ondelettes à 2D (détails) pour chaque échelle  $j \in \mathbb{R}^2$  :

$$\phi_{j,n,m}(x,y) = \phi_{j,n}(x)\phi_{j,m}(y) = 2^j \phi(2^j x - n)\phi(2^j y - m) \quad (2.88)$$

$$\psi_{j,n,m}^1(x,y) = \phi_{j,n}(x)\psi_{j,m}(y) = 2^j \phi(2^j x - n)\psi(2^j y - m) \quad (2.89)$$

$$\psi_{j,n,m}^2(x,y) = \psi_{j,n}(x)\phi_{j,m}(y) = 2^j \psi(2^j x - n)\phi(2^j y - m) \quad (2.90)$$

$$\psi_{j,n,m}^3(x,y) = \psi_{j,n}(x)\psi_{j,m}(y) = 2^j \psi(2^j x - n)\psi(2^j y - m) \quad (2.91)$$

En se référant à l'analyse d'images, les trois ondelettes  $\psi^1, \psi^2$  et  $\psi^3$  sont respectivement des ondelettes horizontales, verticales et diagonales et servent à détecter les détails horizontaux, verticaux et diagonaux de l'image.

#### 2.4.2. Transformée en ondelettes rapide à 2D

Supposons que le signal d'entrée soit une image de taille  $N \times N$ . La décomposition en ondelettes à deux dimensions s'effectue d'abord en décomposant chaque ligne de la matrice en utilisant la décomposition en ondelettes à une dimension (paragraphe 2.1.3.1). Deux matrices rectangulaires de taille  $N \times N/2$  sont alors obtenues. Ces deux matrices sont ensuite décomposées suivant leurs colonnes pour obtenir quatre matrices de taille  $N/2 \times N/2$ . Cette procédure peut être répétée un nombre arbitraire de fois et le nombre total de coefficients après chaque décomposition est toujours égal au nombre initial de coefficients, c'est-à-dire  $N^2$ .

Pour des raisons de simplification d'écriture le produit de deux filtres  $h[n]$  et  $g[m]$  est écrit de la manière suivante  $hg[n,m] = h[n]g[m]$ . Notons également par  $\tilde{i}[n,m]$ , l'image de taille deux fois plus grande que  $\tilde{i}[n,m]$ , obtenue en insérant un zéro entre chaque ligne et colonne de la matrice  $\tilde{i}[n,m]$ .

A l'aide de la séparabilité des bases d'ondelettes et de l'algorithme de Mallat, il peut être vérifié que :

$$a_j[n, m] = a_{j+1} * h_2 h_2[2n, 2m] \quad (2.92)$$

$$d_j^1[n, m] = a_{j+1} * h_2 g_2[2n, 2m] \quad (2.93)$$

$$d_j^2[n, m] = a_{j+1} * g_2 h_2[2n, 2m] \quad (2.94)$$

$$d_j^3[n, m] = a_{j+1} * g_2 g_2[2n, 2m] \quad (2.95)$$

et

$$a_{j+1}[n, m] = \check{a}_j * h_1 h_1[n, m] + \check{d}_j^1 * h_1 g_1[n, m] + \check{d}_j^2 * g_1 h_1[n, m] + \check{d}_j^3 * g_1 g_1[n, m] \quad (2.96)$$

Les équations (2.92) à (2.95) et l'équation (2.96) représentent respectivement l'algorithme de décomposition et de reconstruction en ondelettes à deux dimensions (Figure 2.12).

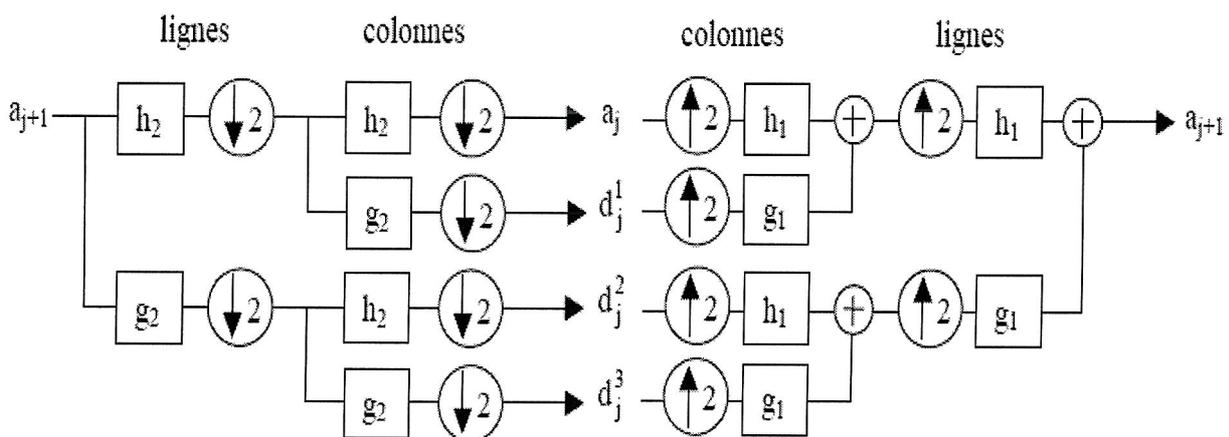


Figure 2.12 : Décomposition et reconstruction en deux dimensions.

Un signal à deux dimensions  $s(x, y) \in V_0^2$  s'écrira, à l'aide de l'analyse multirésolution à deux dimensions, de la manière suivante ( $J_0$  est l'échelle la plus haute et  $J$  l'échelle la plus basse) :

$$s(x, y) = \sum_{n, m} a_j[n, m] \phi_{j, n, m}(x, y) + \sum_{j=J}^{J_0} \sum_{n, m} (d_j^1[n, m] \psi_{j, n, m}^1(x, y) + d_j^2[n, m] \psi_{j, n, m}^2(x, y) + d_j^3[n, m] \psi_{j, n, m}^3(x, y)) \quad (2.97)$$

La Figure 2.13 représente les dispositions spatiales, pour deux niveaux de décomposition, des coefficients d'approximation et de détails de la représentation en ondelettes utilisée.

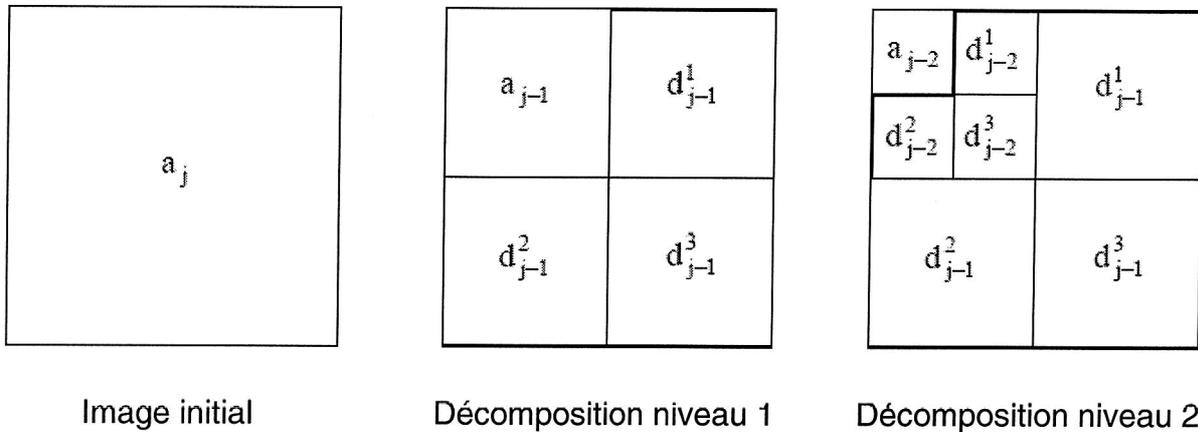


Figure 2.13 : Dispositions spatiales pour deux niveaux de décomposition.

### 2.4.3. Les critères de qualités des ondelettes utilisées en traitement d'image

On classe les bases d'ondelettes en terme de leurs performances dans différentes applications. Le problème majeur est de définir des critères nécessaires pour leurs classifications. On présent un nombre des propriétés souhaitées pour une ondelette dans le cadre d'une analyse multirésolution sont les suivantes :

- **Orthogonalité** : cette propriété est l'une des plus importantes, si les ondelettes sont orthogonales, l'analyse multirésolution est parfaite et la décomposition numérique est stable. Si l'analyse multirésolution est orthogonale (ceci inclut également les ondelettes semi orthogonales), les opérateurs de projection dans les différents sous-espaces conduisent à des approximations optimales au sens de  $L^2(R)$ .
- **Support compact** : si la fonction d'échelle  $\varphi$  et la fonction d'ondelette  $\psi$  sont à supports compacts, les filtres  $H$  et  $G$  sont des filtres à réponse impulsionnelle finie, ce qui évidemment facilite leur implantation. Si elles ne sont pas à supports compacts, une décroissance rapide est souhaitable de sorte que les filtres puissent être raisonnablement approchés par des filtres à réponse impulsionnelle finie.

- Coefficients rationnels : pour une implémentation informatique rapide, on souhaite que les coefficients des filtres  $h_k$  et  $g_k$  soient rationnels ou encore mieux, dyadiques. En effet, diviser par une puissance de 2 sur un ordinateur correspond à un simple décalage de bits.
- Symétrie : si la fonction d'échelle et l'ondelette sont antisymétriques, alors les filtres ont une phase linéaire. Sans cette propriété, il se produit une distorsion de phase lors de la reconstruction.
- Régularité : la régularité d'une ondelette est importante. Si la fonction originale est une image et que l'ondelette n'est pas assez régulière, l'erreur de reconstruction est facilement détectable par l'oeil humain. Notons que la régularité des fonctions de décomposition est plus importante que celle des fonctions de reconstruction. Plus de régularité implique aussi une meilleure localisation fréquentielle des filtres. Finalement, des bases de fonctions régulières sont souhaitées pour des applications d'analyse numériques impliquant des dérivées.
- Nombre de moments nuls : dans le cas d'ondelettes biorthogonales, le nombre de moments nuls de l'ondelette duale est fondamental pour la détection des singularités et pour la caractérisation d'espaces réguliers. Il détermine le taux de convergence d'approximations de fonctions régulières. Le nombre de moments nuls de l'ondelette duale est lié à la régularité de l'ondelette.
- Expression analytique : on ne dispose généralement pas de l'expression analytique de  $\varphi$  ou  $\psi$ . Elle est disponible dans certains cas et c'est très utile.
- Interpolation : si la fonction d'échelle satisfait  $\varphi(k) = \delta_k$ ,  $k \in \mathbb{Z}$ , alors de  $V_j$  qui interpole les données échantillonnées sur une grille de pas  $2^{-j}$ , puisque les coefficients sont simplement les échantillons.

Il est malheureusement pas possible de construire des ondelettes ayant toutes ces propriétés. Il faut donc nécessairement faire des compromis selon le type d'application concernée.

Régularité implique aussi une meilleure localisation fréquentielle des filtres. Finalement, des bases de fonctions régulières sont souhaitées pour des applications d'analyse numériques impliquant des dérivées.

#### 2.4.4. Choix de l'ondelette

Le choix d'ondelette adaptée n'est pas aisé. Il convient de bien concerner le problème à étudier et d'identifier le type de transformée à utiliser (continue ou discrète). En analyse d'image il est souvent utile d'avoir une certaine redondance pour avoir plus d'informations. L'utilisation de la transformée en ondelette continue ou discrète et alors conseillée. Pour une analyse multirésolution, on préfère une base d'ondelettes orthonormales. Et si on veut un calcul exact, alors les ondelettes à support compacte sont idéales. On peut également avoir besoin d'une ondelette indéfiniment dérivable, il faudra alors utiliser une des ondelettes de Meyer. Pour détecter les frontières, on préfère des ondelettes symétriques.

On veut donc qu'on ne peut parler d'une ondelette « idéale » adaptée à tous les cas. Chaque cas particulier correspond une ondelette particulière adaptée

#### 2.4.5. Exemple

La figure 2.14 est un exemple d'analyse multirésolution appliquée à l'image de Souk Ahras à l'aide de l'algorithme de Mallat, en utilisant comme base de décomposition l'ondelette de Haar, et pour des échelles différentes.

Image originale(128x128)



Image à la résolution 1

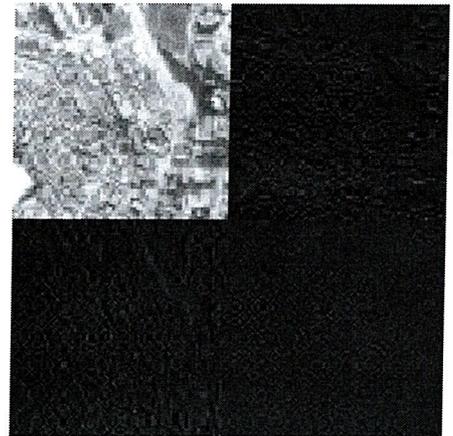


Image à la résolution 2

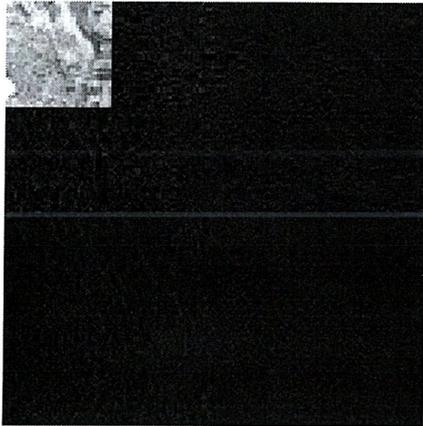


Image à la résolution 3



Figure 2.14 : Analyse multirésolution appliquée à l'image de Souk Ahras à l'aide de l'algorithme de Malat.



Image originale

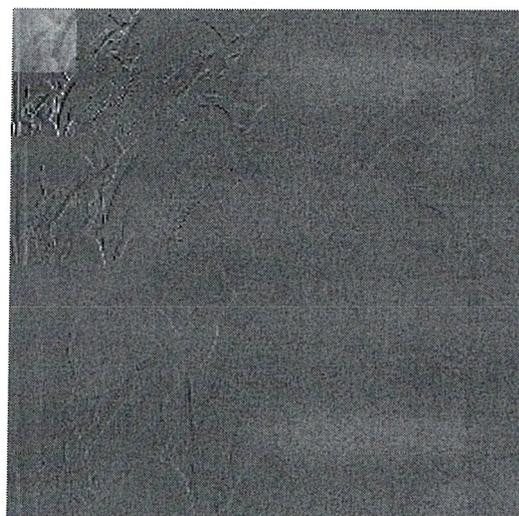


Image à la résolution 1

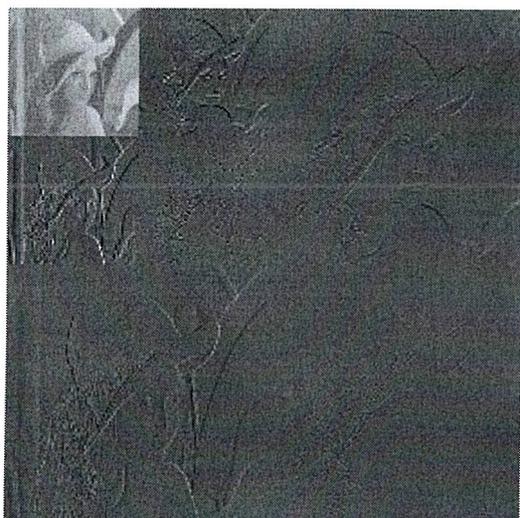


Image à la résolution 2

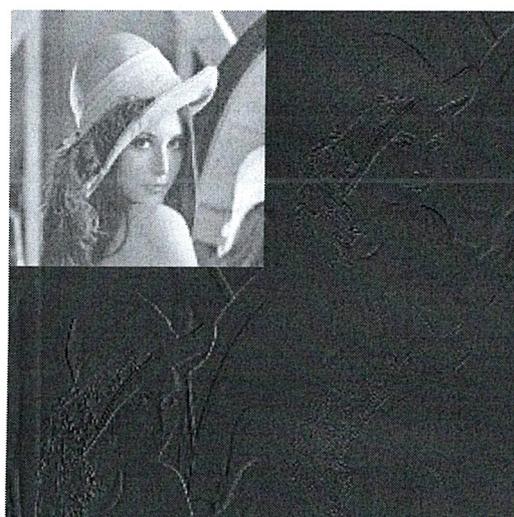


Image à la résolution 3

## 2.5. Conclusion

La transformée en ondelettes est une méthode d'analyse et de représentation des signaux qui est complémentaire à celle de Fourier. Les bases de Fourier n'étaient pas bien adaptées à l'analyse des phénomènes transitoires, tandis que les bases d'ondelettes le sont. La plupart des signaux nécessitent une analyse temps fréquence et les ondelettes fournissent une méthode simple et efficace pour une telle analyse.

L'analyse en ondelettes est caractérisée par l'utilisation de fonctions bien localisées, à la fois dans l'espace physique et dans l'espace spectral, engendrées les unes à partir des autres par translation et dilatation. L'analyse de Fourier quant à elle utilise des fonctions non localisées dans l'espace physique, mais bien localisées dans l'espace spectral, engendrées par les fonctions périodiques sinus et cosinus.

Pour effectuer une transformation en ondelettes, il faut d'abord choisir une fonction d'ondelette mère qui doit respecter la propriété principale d'admissibilité et à partir de cette ondelette mère est construite la famille d'ondelettes par application de dilatation et de la translation de l'ondelette mère, et ensuite calculer les coefficients d'ondelettes. Une bases d'ondelettes discrète est obtenue en discrétisant les paramètres de dilatation et translation.

Avec l'introduction de la fonction d'échelle et de la base de fonctions discrète qui en dérive, l'analyse d'un signal donné en produit une approximation de ce signal; les détails sont ensuite encodés par les ondelettes. Une analyse en multirésolution en a été alors déduite. Cette dernière effectuée à l'aide des ondelettes orthogonales permet de décrire le signal comme un ensemble de coefficients d'ondelettes et d'échelles. Les premiers donnent une représentation globale du signal et les coefficients d'échelles donnent le comportement le plus régulier du signal. L'ensemble des coefficients associés permet de reconstruire le signal.

# **Application à la restauration d'images**

### **3.1. Généralités**

#### **3.1.1. Application de l'imagerie satellitaire**

##### **3.1.1.1. Les spatiocartes**

Une spatiocarte numérique est un produit image obtenu à partir d'une ou plusieurs scènes originales (assemblées par mosaïquage numérique). Elle couvre un site précis quelles que soient les contraintes d'acquisition des données. Ces spatiocartes sont destinées aux personnes ayant besoin d'informations géographiques et sont donc présentées dans un découpage cartographique normalisé afin d'être accessibles à tous les utilisateurs.

##### **3.1.1.2. Les modèles numériques de terrain (MNT)**

Le modèle numérique de terrain est le fichier maillé des altitudes d'une région. Il est présenté sous la forme d'une grille régulière, une altitude tous les 20m par exemple. Les MNT, éventuellement croisés avec d'autres informations, permettent la réalisation de cartes de pentes, d'ensoleillement ou de ruissellement, utilisées dans le cadre de l'aménagement de sites (parcs de loisirs, prévention des risques naturels, ...). Ils peuvent aussi être utilisés pour prévoir les tracés des constructions linéaires (canaux, lignes de TGV, implantation d'autoroutes, ...) et calculer les cubages de matériaux pour les études d'impact, pour optimiser l'implantation d'émetteurs de télécoms, etc.

##### **3.1.1.3. Systèmes d'information géographique (SIG)**

Mis au point il y a une vingtaine d'années, le SIG est un système d'information qui intègre des données géographiques (ou à référence spatiale). Il permet de tirer profit des éléments géographiques présents dans les bases de données, de manière à visualiser, cartographier, comprendre et aider à la prise de décision.

Un SIG comprend trois composantes : une technologie (matériel et logiciel), une base de données et une structure (hommes, budgets, organisation et méthode). C'est un projet qui se déploie dans le temps.

### 3.1.1.4. Suivi des subsidences

La subsidence ou son corollaire l'inflation de terrain ont des causes diverses : exploitations minières souterraines, stockage de gaz souterrain, installations d'équipement souterrain, pompes de nappes...

L'impact sur l'environnement et la sécurité des ouvrages souterrains nécessitent une activité de surveillance. L'objectif de cette surveillance est de préciser le comportement à moyen et long terme des cavités, d'évaluer leurs répercussions en surface, de mettre au point des méthodes de surveillance des zones à risques et enfin de concevoir des procédés de réduction des risques.

Parmi les techniques utilisées, certaines sont fondées sur des méthodes de mesures directes (nivellement ou positionnement par satellite GPS, extensomètres...), d'autres sur des données obtenues à partir d'images satellitaires par interférométrie optique ou radar. Les performances de ces méthodes sont liées à la précision des mesures de déformations, à la densité des réseaux de capteurs et à la pertinence géostatistique de ces réseaux.

## 3.2. Acquisition des images satellites

Un satellite est constitué de deux grandes parties :

- la charge utile
- la plateforme

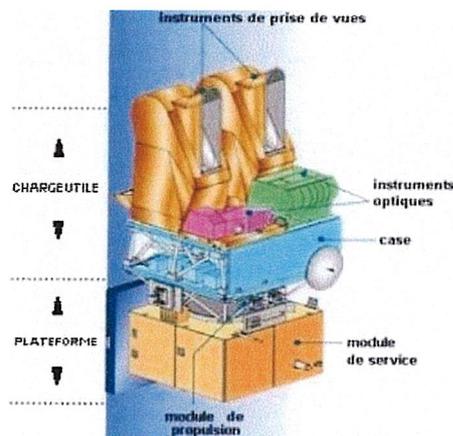


Figure 3.1 : Machinerie d'un satellite

### 3.2.1. La charge utile

La charge utile varie en fonction de la mission du satellite. On y trouve soit des instruments de télécommunication ou des instruments d'observations. Pour l'acquisition des images satellites se sont évidemment des instruments d'optiques. Il est constituée de :

- Une case

Il s'agit d'une structure en fibres de carbone assurant la stabilité nécessaire aux instruments. Elle accueille les équipements électroniques optiques de traitement d'images, de mémorisation et de transmission des données vers le sol. On peut y stocker les images après compression et formatage dans une mémoire de masse pouvant être des enregistreurs à bande magnétique pour les satellites anciens et une mémoire électronique pour les plus récents (capacité allant de 20 à 100 Go).

- Les instruments de prises de vues et optiques

Parmi les différents instruments de prise de vues et optiques, le système altimétrique constitue l'un des principaux et l'un des systèmes les plus perfectionnés de visualisation par satellite. En effet, pour obtenir certaines cartes (du relief, du niveau des océans,...), il est nécessaire de faire appel à des technologies différentes de l'acquisition numérique directe (comme la photo), qui n'est pas assez précise ou trop longue à traiter pour obtenir des images au jour le jour. Ainsi, le système d'altimétrie spatiale a été choisi par les concepteurs de satellites et fonctionne sur de nombreux satellites.

Un radar altimètre embarqué à bord d'un satellite émet un signal à très haute fréquence (plus de 1700 impulsions par seconde) à la verticale de celui-ci en direction du sol et reçoit en retour l'écho réfléchi par la surface du sol. L'analyse de l'écho permet d'extraire une mesure très précise du temps de trajet aller-retour entre le satellite et la surface terrestre. Toutefois les ondes électromagnétiques peuvent être ralenties pendant leur traversée dans l'atmosphère à cause du taux d'humidité ou du taux d'ionisation. Il faut donc appliquer certaines corrections prenant en compte ces phénomènes physiques (négligeables sur de petites distances mais pas sur des grandes). Ce temps est ensuite transformé en distance par simple

multiplication par la vitesse de la lumière, vitesse à laquelle se propagent les ondes électromagnétiques émises.

### 3.2.2. La plateforme

La plateforme est l'endroit où l'on trouve les équipements qui permettent d'assurer le bon fonctionnement du satellite. Parmi ces équipements il y a un ordinateur de bord, des réservoirs, des petits moteurs permettant de replacer le satellite sur son orbite en cas de dérive ainsi que des radios de liaison avec la terre.

### 3.3. Acquisition des images

#### 3.3.1. Le système d'acquisition

Le satellite destiné à l'acquisition d'images est équipé d'un radiomètre, appareil mesurant la quantité de lumière provenant de la partie de la planète visée. Ce radiomètre est constitué d'une grille de détecteurs CCD (Charge-Coupled Device, capteur à transfert de charge en français), grâce auxquels le rayonnement du terrain va être perçu (selon un fonctionnement identique à celui d'un appareil photo numérique).

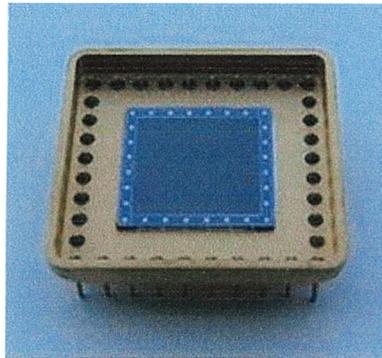


Figure 3.2 : Capteur à transfert de charge

Les capteurs CCD sont des composants microélectroniques sensibles qui convertissent l'intensité lumineuse (signal analogique) en un signal électrique. Pour

ce faire, le capteur CCD emmagasine les particules de lumière, les photons, dans un corps de silicium. Ce dernier a pour propriété de libérer des électrons lors de l'absorption de photons. Le détecteur relié à un contact électrique transmet donc une charge électrique plus au moins élevée selon la lumière perçue. Cette charge est convertie par un convertisseur Analogique/Numérique en un signal numérique codé en binaire

Dans le cas d'une image satellite acquise en couleur, chaque détecteur est composé de trois capteurs CCD qui perçoivent chacun l'intensité lumineuse des trois couleurs de base : rouge, vert et bleu. Par superposition des trois couleurs de base, on obtient une couleur précise, Le signal numérique va ensuite pouvoir être interprété pour obtenir une image.

Pour acquérir une image satellite d'une ligne complète au sol de la largeur du champ que l'on souhaite observer, on utilise une prise d'image dite en 'râteau' (push broom en anglais). Ce principe permet, grâce à des lignes de détecteurs de couvrir la largeur désirée d'un seul coup. Le balayage de la longueur est quant à lui obtenu par le défilement du satellite sur son orbite. Ainsi en recoupant les informations on obtient une image satellite de la zone à observer. (Le principe est similaire au fonctionnement d'un scanner c'est-à-dire un balayage avec une largeur fixée).

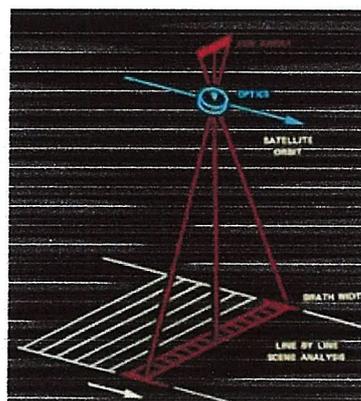


Figure 3.3 : Le principe du balayage en râteau ou push-broom

### 3.4. L'obtention de l'image satellite numérique

Suite à l'interprétation des signaux précédemment acquis, on obtient une image satellite numérique. Le radiomètre constitué d'une grille de détecteurs CCD va donc émettre une suite de nombres binaires de 8 chiffres chacun qui correspondent à l'intensité lumineuse perçue par chaque capteur CCD de la grille. Chaque nombre, ou octet, correspond :

- Dans le cas d'une photographie noir et blanc à un niveau de gris. Ainsi, l'octet 11111111 correspondra au blanc de valeur décimale 255, et l'octet 00000000 au noir de valeur décimale 0.
- Dans le cas d'une photo couleur, trois octets sont émis par un détecteur CCD pour chaque couleur de base : vert, rouge, et bleu.

La formation de l'image est alors possible. L'image numérique est également une grille, une matrice composée de lignes et de colonnes. Chaque carreau est nommé pixel (Picture élément). Chaque pixel contient son " compte numérique " déterminant sa couleur qui n'est autre que le nombre binaire issu de chaque détecteur CCD, et ses coordonnées en x et y. Les pixels réunis, on obtient une image identique à celle prise par le satellite.

La résolution d'une image étant la quantification du nombre de pixel pour une dimension donnée, exprimé en dpi ou en ppp.

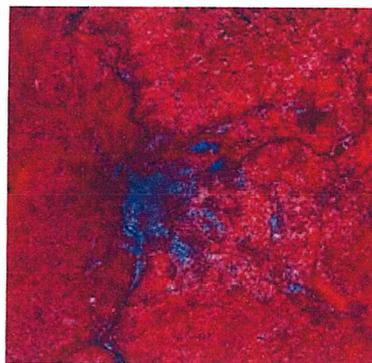


Figure 3.4. Exemple d'image satellite

### **3.5. Différents modes d'acquisition**

Les données envoyées par les satellites ne sont donc pas des images toutes faites. Elles doivent être composées à partir des informations fournies par ses instruments. Comme nous l'avons vu, les capteurs CCD collectent la lumière que renvoient la surface observée et traduit cette information en signal numérique. Avec un satellite récent différents modes d'acquisition existent : le mode multibande ou le mode panchromatique. Leur rendu et leur utilisation sont différents.

#### **3.5.1. Le mode multibande ou multispectral**

Elle est acquise par un capteur numérique qui mesure la réflectance dans de nombreuses bandes spectrales (typiquement, une dizaine). Ces multiples valeurs de réflectance se combinent pour créer des images couleur. Elle sert principalement à obtenir des informations à caractère " qualitatif " (ex. classification des types de végétation sur des parcelles).

#### **3.5.2. Le mode panchromatique**

Elle est acquise par un capteur numérique qui mesure la réflectance dans une large bande électromagnétique. Les données panchromatiques sont représentées sous forme d'images en noir et blanc. Elle sert principalement à obtenir des informations de type " géométrique " (formes, dimensions, surface).

#### **3.5.3. Les différents instruments**

Selon le mode d'acquisition choisi, les instruments diffèrent. Par exemple, l'instrument HRVIR (Haute résolution visible et infrarouge) de Spot servira pour acquérir des images en mode multi spectral. En effet, il peut observer dans différentes bandes spectrales

En revanche, si l'on veut observer en mode panchromatique, alors on utilisera un appareil adapté comme l'instrument HRS situé également sur Spot.

Il existe aussi des instruments spécifiques comme « Végétation » sur le satellite Spot et qui permet, comme son nom l'indique, une observation adaptée de zone de végétation.

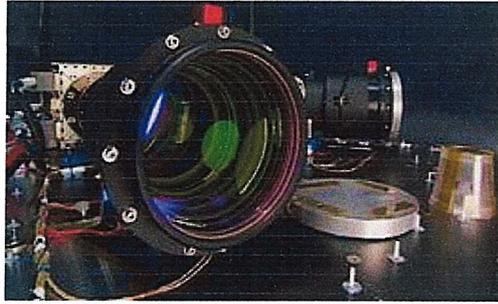


Figure 3.5. Instrument d'acquisition

### 3.6. Transfert des images du satellite vers la Terre

Les images satellites acquises sont transmises sur Terre soit en direct si une station de réception est en visibilité du satellite et est programmée, soit en différé (après stockage) lors du passage sur une station. Le débit de transmission varie selon le satellite entre 25 et 100 Mo par seconde.

Deux cas majeurs peuvent se présenter :

Le satellite est en visibilité d'une station de réception. Les images peuvent être envoyées en temps réel à cette station si celle-ci a été programmée. Le satellite n'est pas en visibilité d'une station. Les acquisitions programmées sont réalisées et les images sont stockées par des enregistreurs embarqués. Elles sont transmises plus tard, lorsque le satellite se trouve en visibilité d'une station de réception.

Pour transmettre entre des points éloignés ces signaux dont la portée est limitée, on utilise, le plus souvent, comme véhicule de liaison, une onde électromagnétique. Aujourd'hui, des dizaines de stations de réceptions de données satellites, et notamment d'images satellites, sont réparties à la surface de la planète. À chacune des stations de réception, on retrouve une ou plusieurs antennes paraboliques d'une dizaine de mètres de diamètre. Ce diamètre important permet une meilleure réception des signaux même les plus faibles. Pour permettre l'enregistrement des données transmises par le satellite, l'antenne de réception doit suivre et pointer directement vers le satellite.

### **3.7. Traitement et interprétation des images**

L'image que nous transmet le satellite est brute. Il faut donc effectuer un certain nombre de retouches pour la rendre utilisable.

Tout d'abord il faut effectuer une correction géométrique car lors de la prise de l'image, les reliefs, la rotation de la Terre sur elle-même ou encore le positionnement du satellite, affectent l'image et la déforme légèrement.

Une fois ces corrections effectuées, Alors par la suite nous intéressons a la partie de traitement c'est à dire la restauration d'image satellitaire bruité par l'application des de deux méthodes ICM et recuit simulé avec l'analyse multirésolution avant l'interprétation finale.

### 3.8. Application de la méthode ICM

Dans le cadre de l'ICM, la recherche du maximum a posteriori se fait de façon déterministe.

Cet algorithme, contrairement au recuit simulé, est très rapide (une dizaine de balayages permettent d'arriver à convergence) et peu coûteux en temps de calcul puisqu'il ne nécessite que le calcul des énergies conditionnelles locales. En contrepartie, ses performances dépendent très fortement de l'initialisation (par rapport à la forme du paysage énergétique) puisqu'il converge vers un minimum local. L'ICM s'apparente à une descente de gradient (on fait baisser l'énergie à chaque itération) ou à un recuit simulé gelé à température nulle, et peut donc rester bloqué dans le minimum énergétique local le plus proche de l'initialisation. Nous illustrons les résultats suivants.

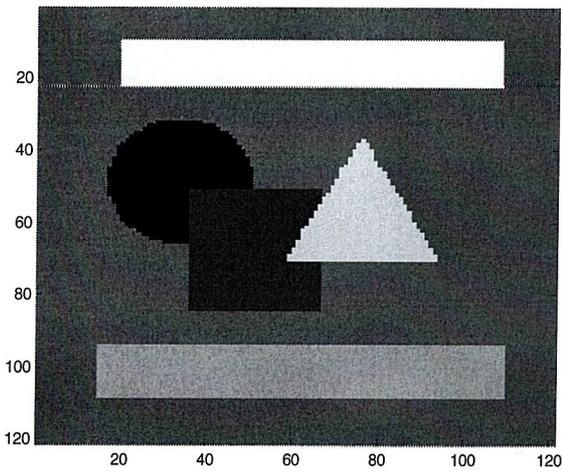


Image Test Original 121x121

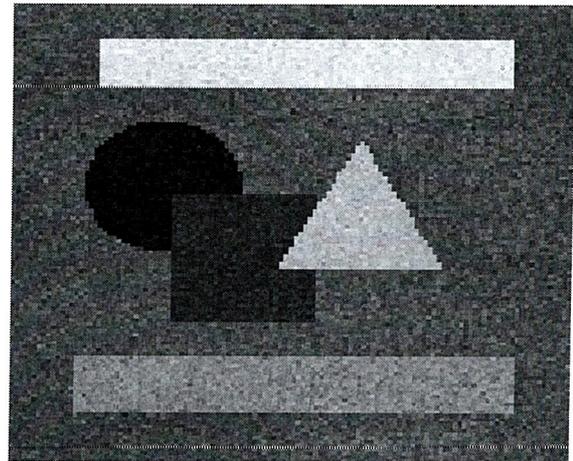


Image Test Bruité variance =10

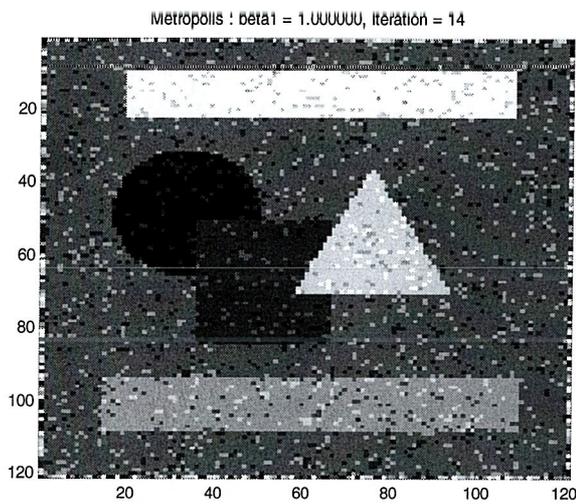


Image Test à l'itération 14

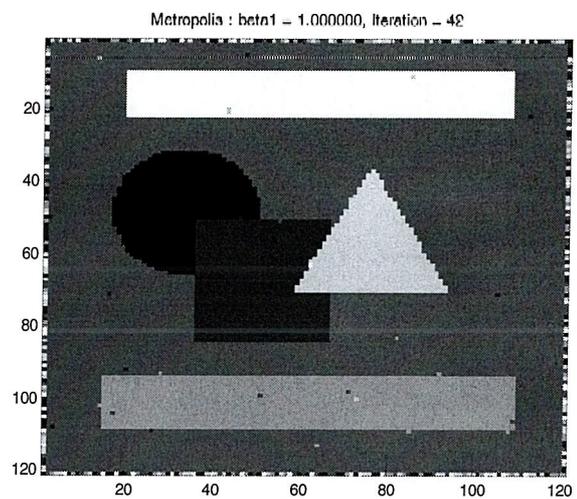


Image Test à l'itération 42

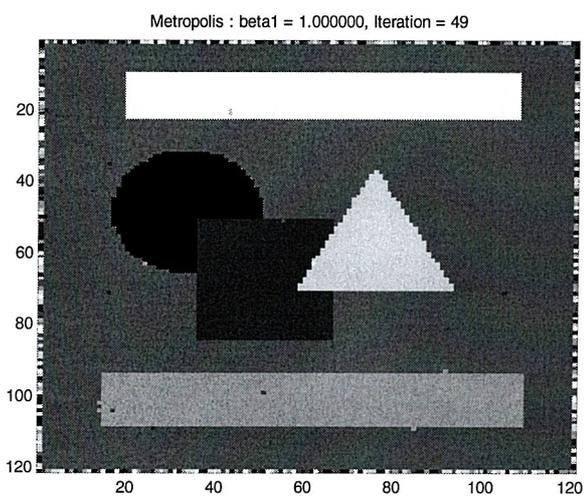


Image Test à l'itération 49

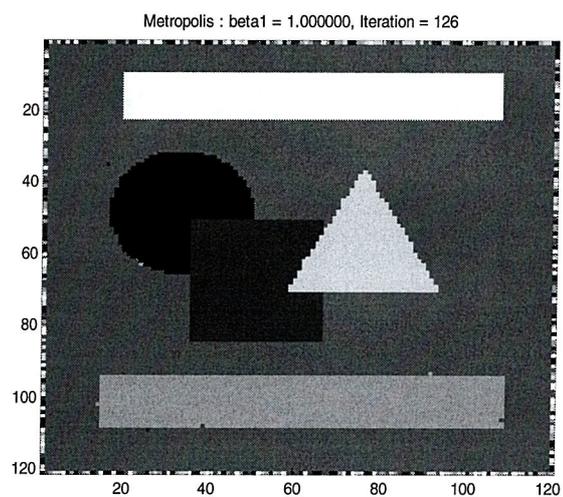


Image Test à l'itération 126  
SNR=29.3567

Figure 3.6 : Exemple de La méthode ICM

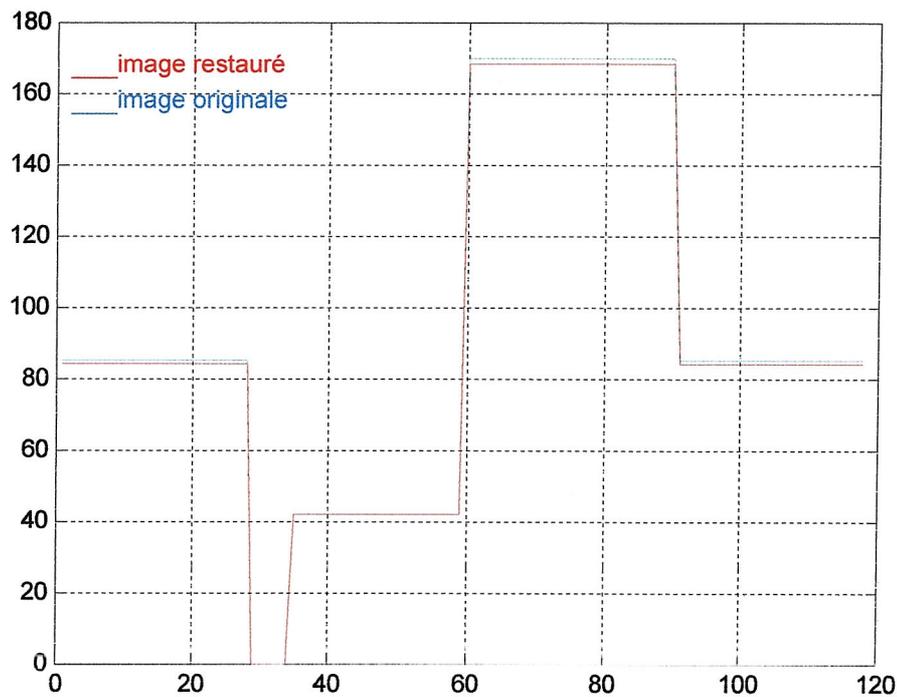


Figure 3.7 : Profil à la ligne 65

On prend comme pondération du terme de régularisation  $\beta = 1$ . On constate que, dès l'itération (42), l'ICM a convergé. En effet il n'y a plus d'évolution du résultat. On peut effectivement remarquer en itération (126) que, au bout de 80 itérations, le résultat n'a pas changé avec  $\text{SNR}=29.3567$ . Cependant, il reste des points mal classés, ce sont des minima locaux, ce que nous n'avons pas avec l'algorithme de recuit simulé comme on peut le voir à la suite, L'algorithme de recuit simulé met plus longtemps à converger que l'ICM mais converge vers les minima globaux.

### 3.9. Application de la méthode recuit simulé

L'algorithme de recuit-simulé est dédié à la recherche d'une configuration d'énergie minimale d'un champ de Gibbs, les images obtenues par recuit simulé sont uniques et doivent en théorie correspondre aux minima globaux de l'énergie. L'idée d'intégrer un paramètre de température et de simuler un recuit a été initialement proposée par Kirkpatrick et reprise par Geman et Geman. Intuitivement le recuit permet d'atteindre un optimum global car il accepte, grâce à ce paramètre de température, des remontées en énergie. Avec la décroissance de la température, ces sauts énergétiques sont progressivement supprimés au fur et à mesure qu'on se rapproche de l'optimum global. La descente en température doit donc se faire suffisamment lentement pour que l'algorithme ne reste pas piégé dans un minimum local de l'énergie, Nous illustrons les résultats suivants.

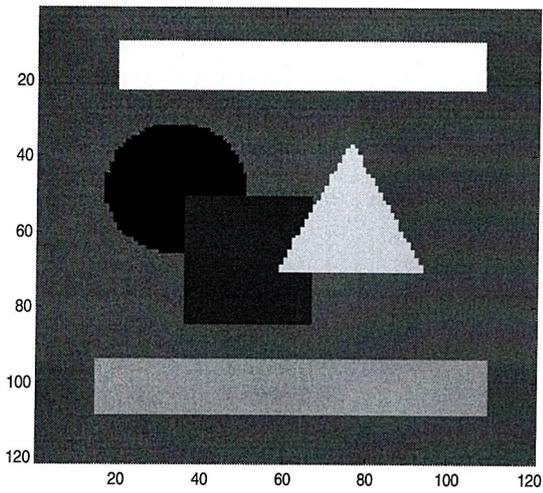


Image Test Original 121x121

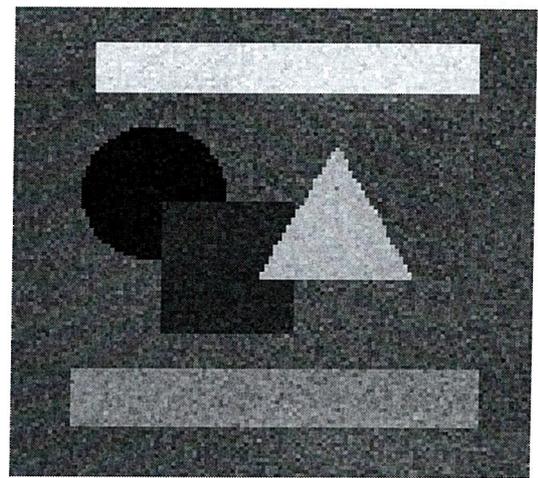


Image Test Bruité variance =10

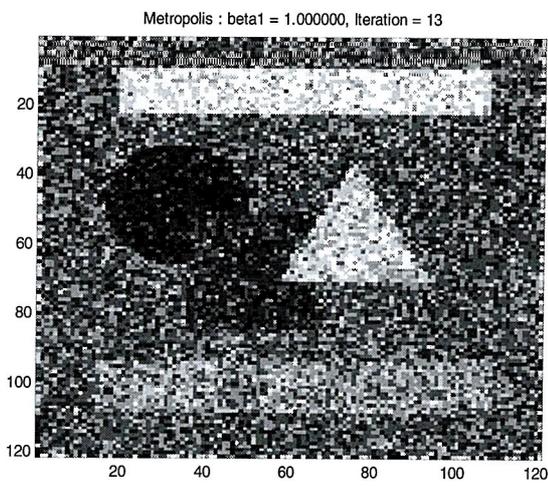


Image Test à l'itération 19

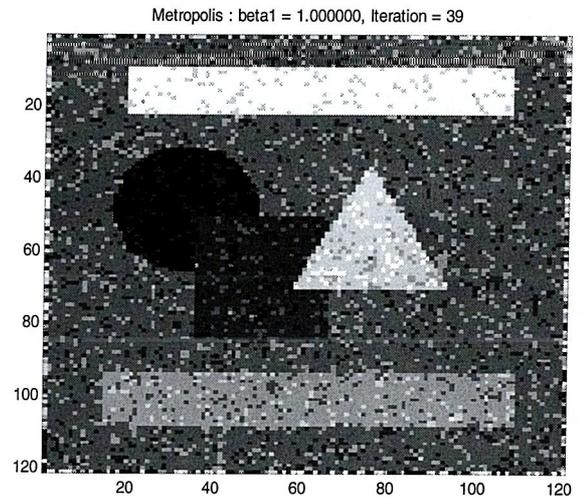


Image Test à l'itération 29

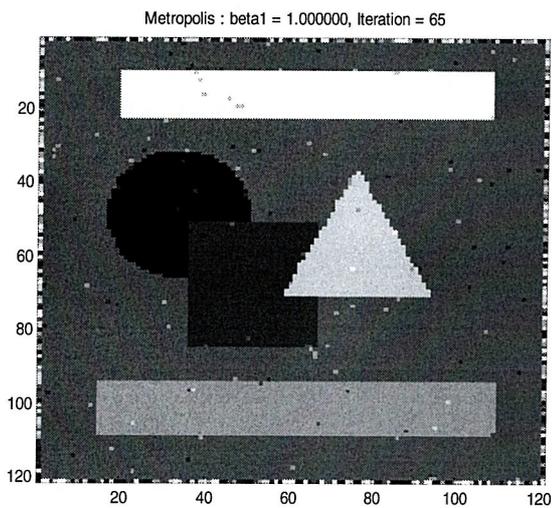


Image Test à l'itération 89

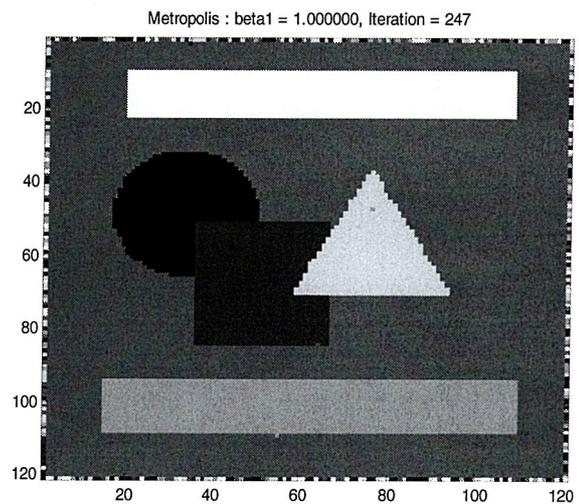


Image Test à l'itération 247  
Avec 39.7980

Figure 3.8 : Exemple de La méthode recuit simulé

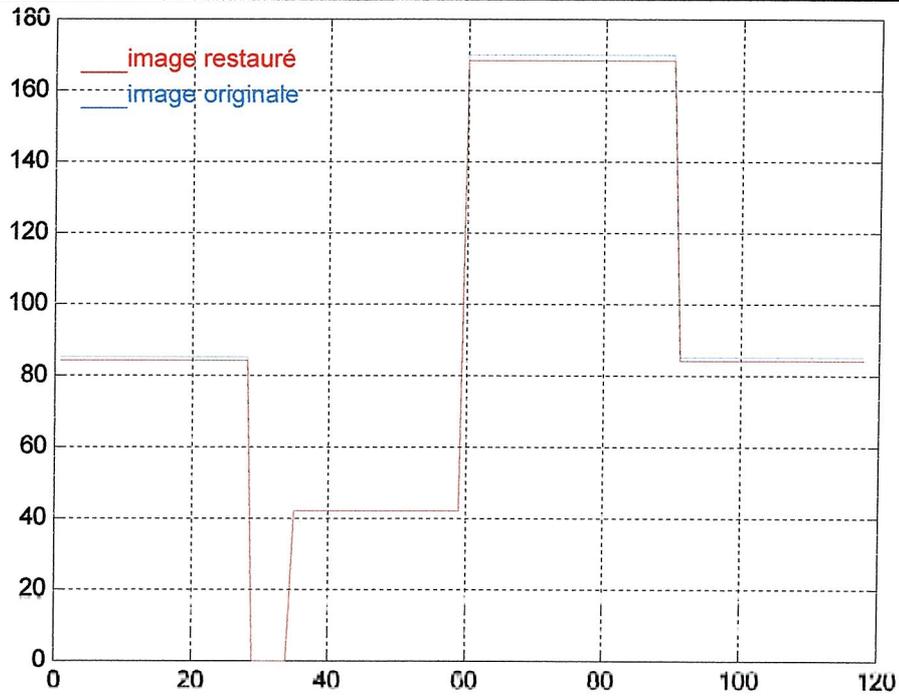


Figure 3.9 : Profil a la ligne 65

On prend comme pondération du terme de régularisation  $\beta = 1$ , avec  $T=70$ ,  $T_0=10$  et nombre d'itération  $nb\_it = 250$  on trouve  $SNR=39.7980$

L'algorithme de recuit simulé met plus longtemps (280 secondes pour notre exemple) et pour remédier ce problème on utilise l'analyse multirésolution pour accélérer l'algorithme et donc vers un meilleur résultat

### 3.10 Analyse du rapport signal a bruit (SNR) en fonction de la résolution

L'image bruitée a été étudiée après décomposition à différentes résolutions. Les images sont considérées à un niveau de résolution donné. La figure 10 représente l'évolution de SNR énergie ( étudié est défini par le rapport de l'énergie de l'image originale (non bruitée) sur l'énergie de l'image de bruit (erreur entre l'image originale et l'image bruitée) ) en fonction de la résolution , On constate sur ce graphique que le SNR de l'image à une résolution donnée est inférieur à celui de l'image à la résolution plus grossière : le SNR augmente lorsque la résolution diminue . Cette évolution se justifie par l'étude de la répartition de l'énergie dans les sous-images décomposées, d'une part sur une image type (exemple de Léna. . .), et d'autre part sur une image de bruit blanc.

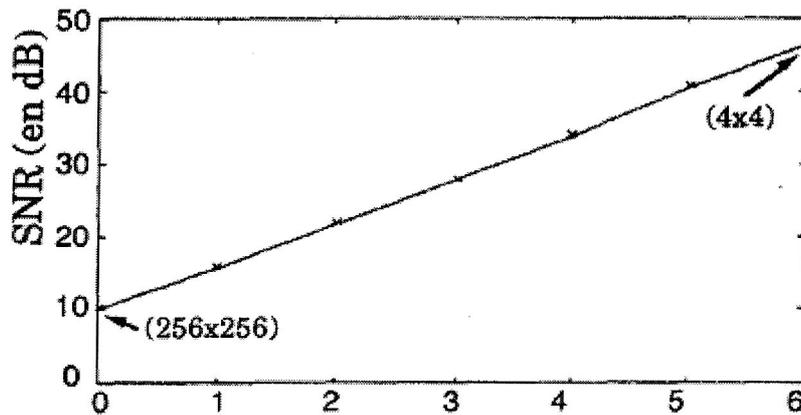


Figure 3.10. Évolution du SNR énergie (rapport des énergies) en fonction de l'indice

Dans le cas dyadique, l'image de résolution  $2^m$  (résolution indice  $m$ ) est décomposée en quatre sous-images : la sous-image de résolution inférieure  $2^{(m+1)}$ , qui correspond au filtrage passe-bas de la première image, et trois sous-images de hautes fréquences, qui correspondent aux détails perdus entre la résolution  $2^m$  et la résolution  $2^{(m+1)}$  dans le sens horizontal, vertical, diagonal. Pour une image visible du type Léna ou d'autre image, l'énergie est principalement concentrée dans les sous-images de basses fréquences : à chaque décomposition, l'énergie reste concentrée dans la sous-image de basses fréquences, chaque sous-image de coefficients en ondelettes ne comportant comparativement que très peu d'énergie .

Pour une image de bruit blanc, on constate au contraire une diminution constante de l'énergie dans les sous-images de basses fréquences. Les ondelettes biorthogonales utilisées étant proches d'une base orthogonale, la décomposition sépare le spectre de l'image en 4 zones d'égale énergie. Donc l'énergie dans la sous-image de basses fréquences diminue d'un facteur 4. Ceci est reproduit à chaque niveau de résolution d'après [53] la vérification expérimentale des propriétés du bruit blanc sont conservé à chaque niveau de résolution, après filtrage et sous échantillonnage. Ainsi la restauration est plus simple aux faibles résolutions puisque l'image est moins bruitée Elle est aussi plus rapide car il y a moins de points que sur l'image d'origine. La connaissance de l'évolution du SNR à travers les résolutions donne des indications pour adapter les paramètres du modèle Markovien aux différentes échelles.

### 3.11. Restauration multirésolution

#### 3.11 .1. L'algorithme multirésolution

L'algorithme de restauration multirésolution est le suivant :

L'image  $Y$  observée est décomposée jusqu'à une résolution  $2^M$  par une transformée en ondelettes biorthogonales. On note  $Y_m$  les suites de sous-images de basses fréquences et  $C_m^Y$  les suites de sous-images de hautes fréquences ou sous-images de coefficients en ondelettes à chaque résolution  $m$  en utilisent ici une ondelette de Haar,. Dans un second temps, on effectue les itérations de restauration et reconstruction pour tous les niveaux  $m$  de résolutions, de  $m = M$  à  $m = 0$ .

Une itération  $m$  est schématisée en figure 11.

A la première itération, l'image bruitée  $Y_M$  est restaurée par l'algorithme recuit simulé.

Notons  $\hat{X}_m$  l'image restaurée au niveau  $m$ . La nouvelle image observée  $Z_{M-1}$  au niveau  $M-1$  est la synthèse de  $\hat{X}_M$  avec les coefficients en ondelettes bruités  $C_M^Y$  du niveau  $M$ , L'algorithme est ensuite itéré sur  $m=M$  à  $0$ .

On remarque qu'à la première itération (niveau  $M$ ), on a  $Z_M = Y_M$ . Ensuite  $Z_m \neq Y_m$  pour  $0 \leq m \leq M-1$ . Les résultats que nous présentons concernent une image bruitée. Étant donné le niveau de bruit très important, il apparaît que lors de la dernière synthèse  $\hat{X}_1$  avec  $C_1^Y$ , l'image construite  $Z_0$  est très bruitée bien que les résolutions d'indices  $m$  avec  $M \leq m \leq 1$  aient déjà été restaurées. Cette image est presque aussi difficile à traiter que l'image brute d'origine, et les résultats de restauration mono et multirésolution sont semblables. Le problème est que le SNR de la sous-image de coefficients en ondelettes  $C_1^Y$  est très faible, Ainsi à la synthèse de  $\hat{X}_1$  avec  $C_1^Y$ , la partie signal introduite est très faible et le bruit domine largement.

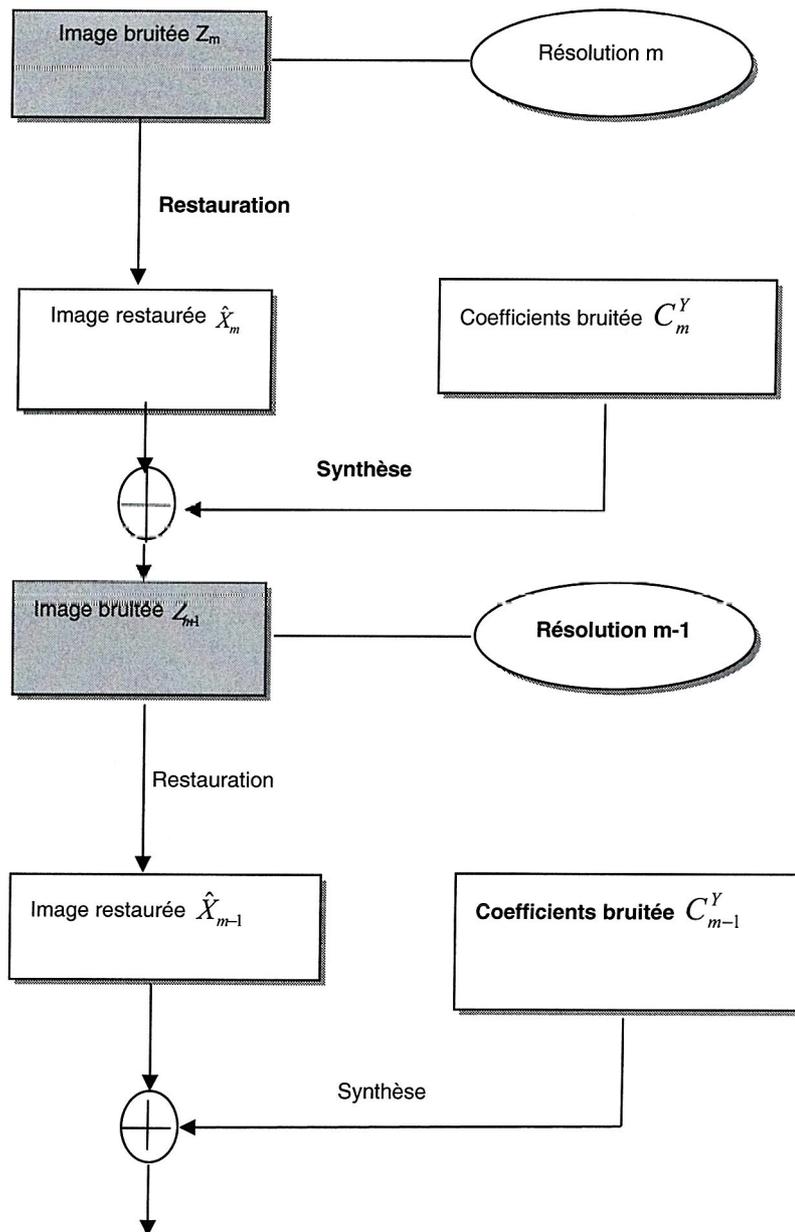


Figure 3.11. Schéma d'une itération de l'algorithme de restauration multirésolution.

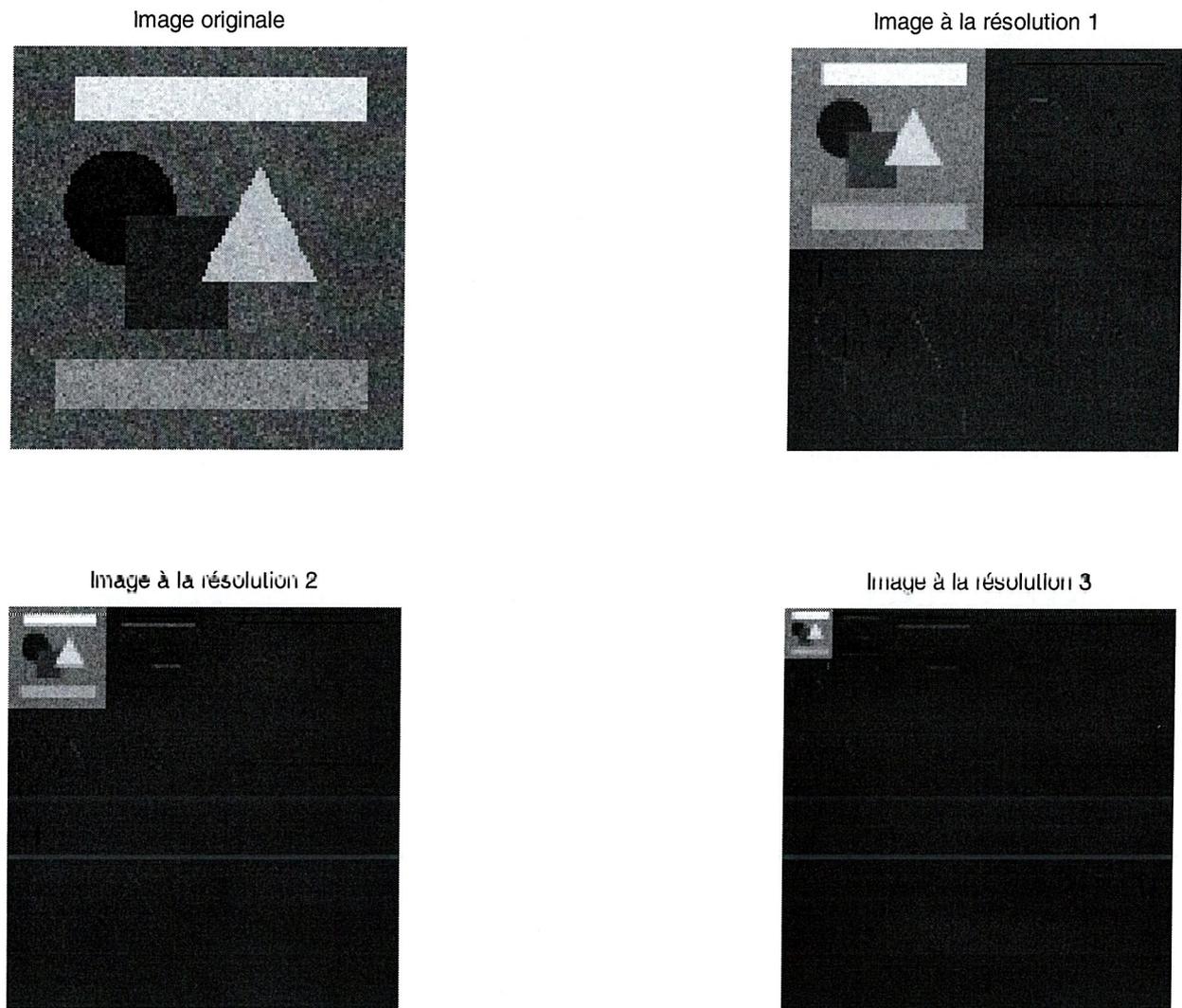


Figure 3.12. La décomposition d'image par l'analyse multirésolution.

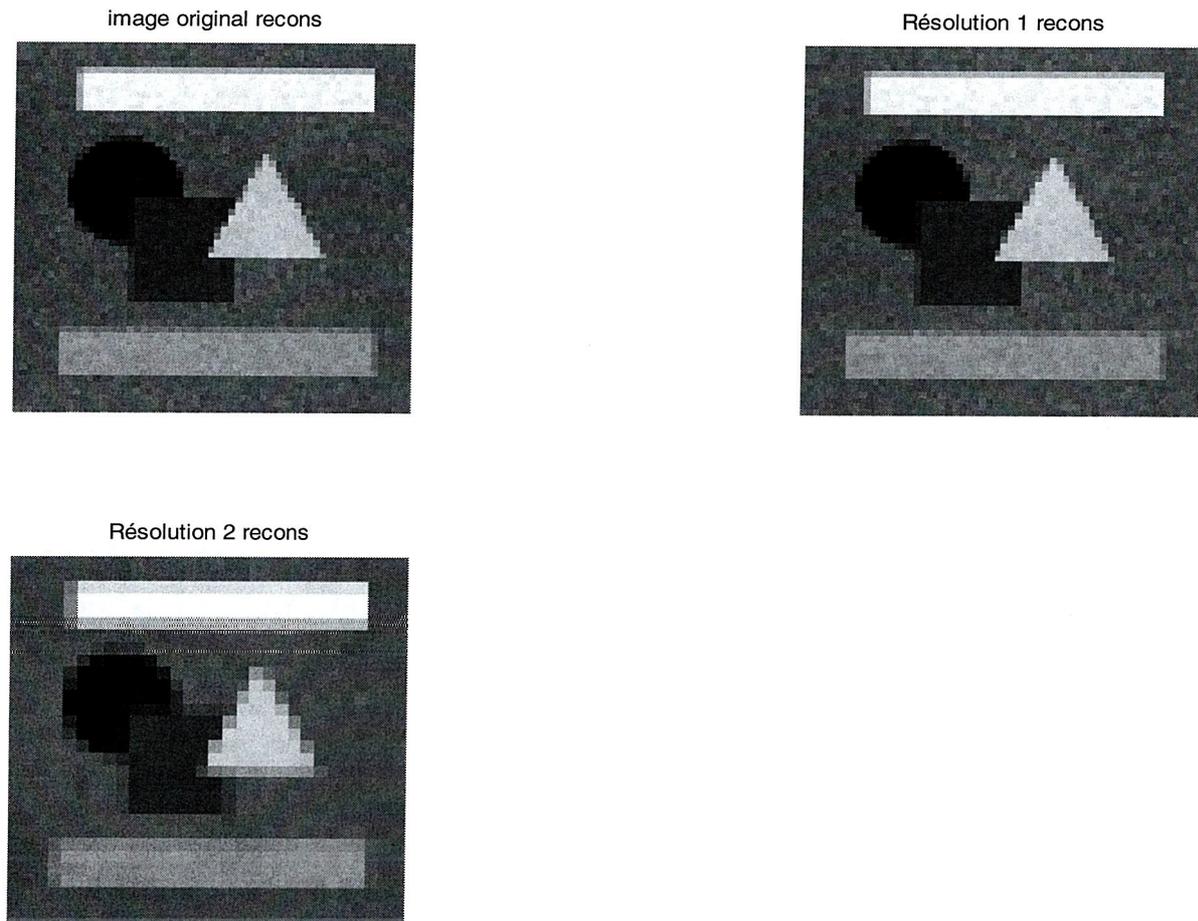


Figure 3.13. La restauration d'image par la méthode de recuit simulé et l'analyse multirésolution.

### 3.11.2. Interprétation des Résultats

Le choix des paramètres est fixé de manière empirique actuellement. Quelques considérations permettent d'orienter ce choix : les paramètres doivent être adaptés en fonction de l'image cherchée et du bruit. Le temps CPU est diminué par rapport à l'algorithme en monorésolution. Cependant, la qualité de la restauration reste médiocre. Le seul réglage des paramètres  $\lambda$  permet pas de diminuer le bruit sans dégrader certains éléments importants de l'image,

### 3.12. Restauration des coefficients d'ondelettes

Revenons au problème évoqué en partie précédente, dû au niveau de bruit important de la sous-image de coefficients  $C_1^Y$ , Dans la restauration multirésolution par recuit simulé, on perd à la dernière synthèse ( $Z_0 = \hat{X}_1$  synthétisé avec  $C_1^Y$ ), les bénéfices de la restauration de  $\hat{X}_1$ , car l'image  $Z_0$  est presque aussi difficile à traiter que l'image bruitée  $Y_0$  elle-même.

C'est pourquoi nous avons tenté de traiter l'image  $C_1^Y$  pour la synthétiser ensuite avec  $\hat{X}_1$  et donner directement l'image finale restaurée  $\hat{X}_0$ . Nous utilisons dans cette partie une décomposition de l'image de coefficients  $C_1^Y$

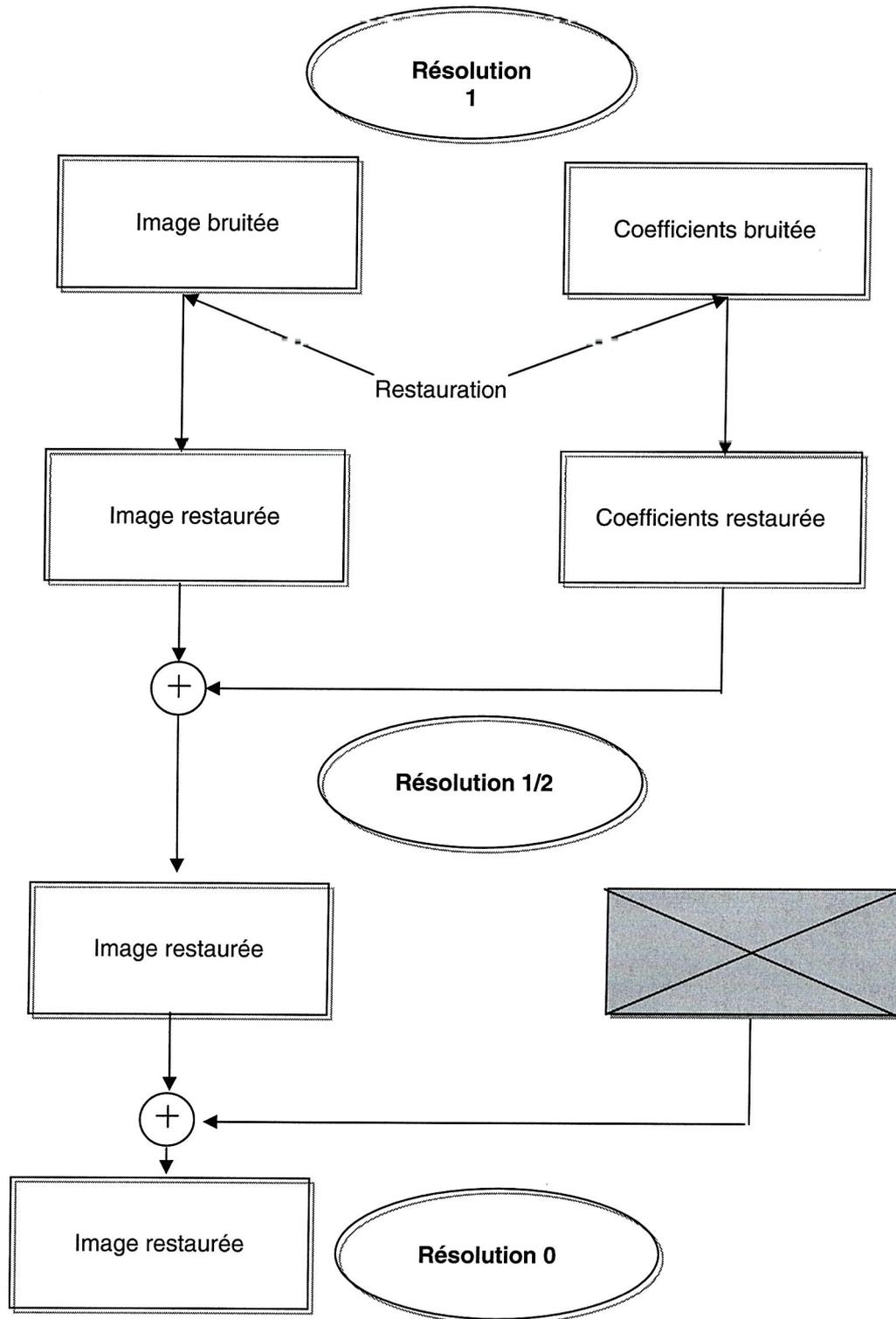


Figure 3.14 . Schéma de la dernière itération de l'algorithme multirésolution avec traitement de la sous-image de coefficients en ondelettes  $C_1^Y$

### 3.13 Résultats

- Application a l'image de Souk Ahras

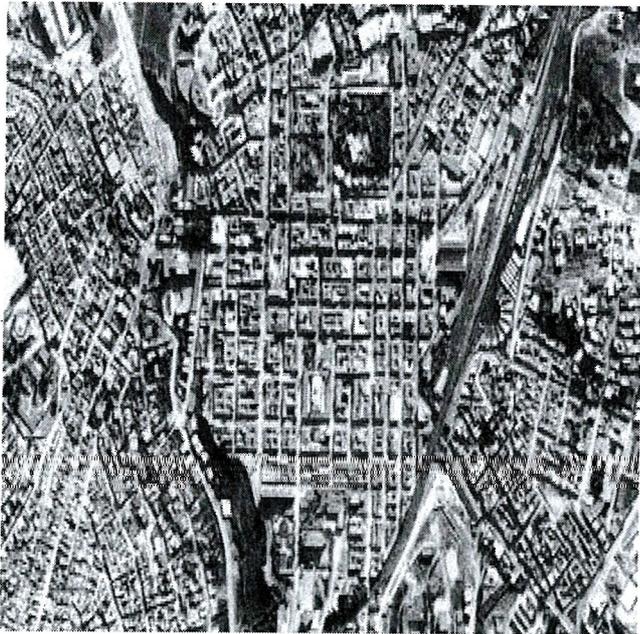


Image originale

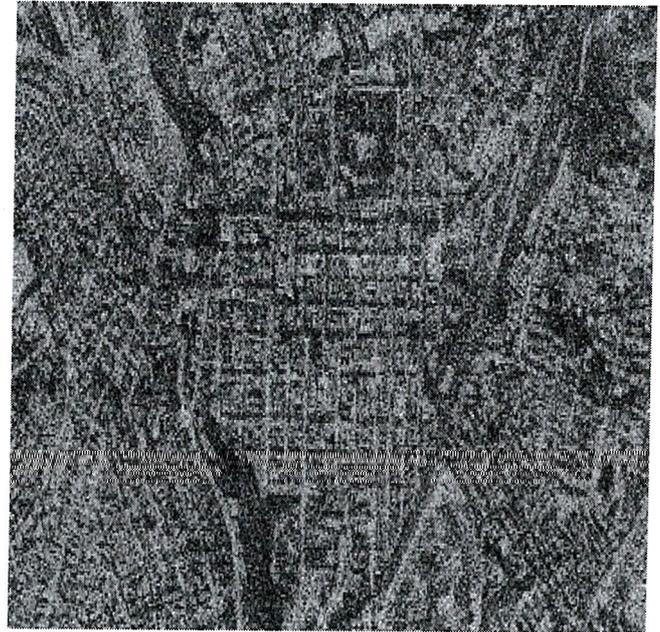


Image Bruitée avec variance=12

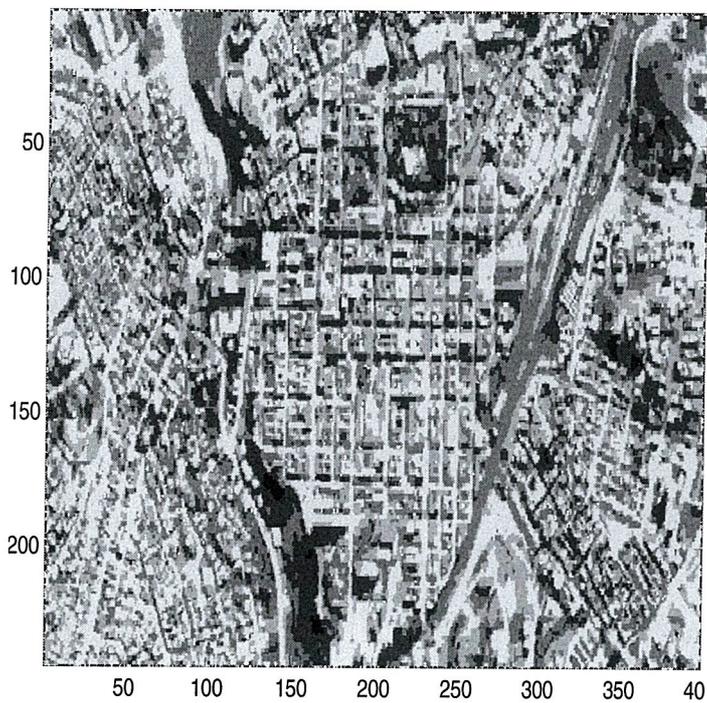


Image reconstituée

Figure 3.15. Exemple d'application sur l'image de Souk Ahras

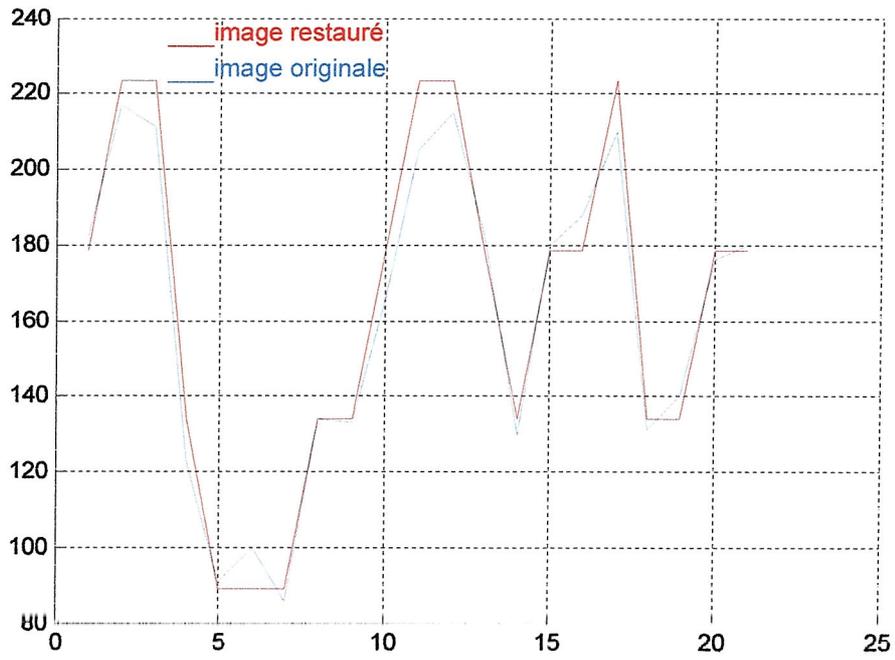


Figure 3.16. Profil l'image de Souk Ahras a la ligne 100

- Application a l'image de Hôtel Emarate

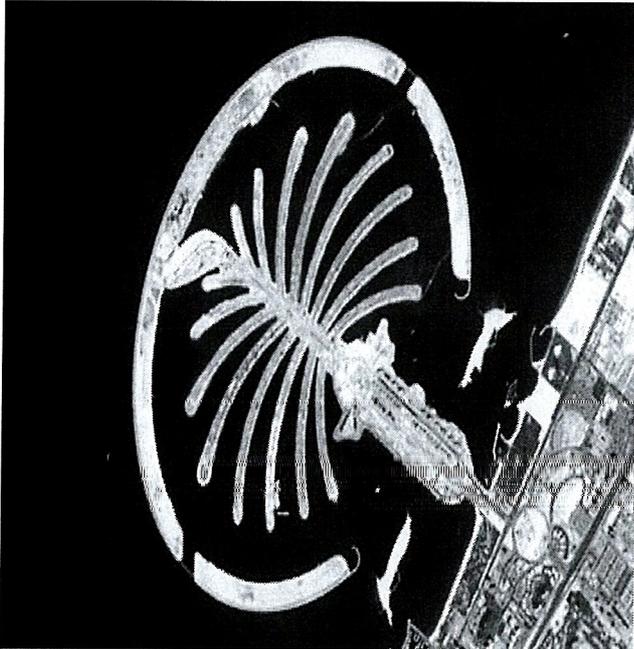


Image originale

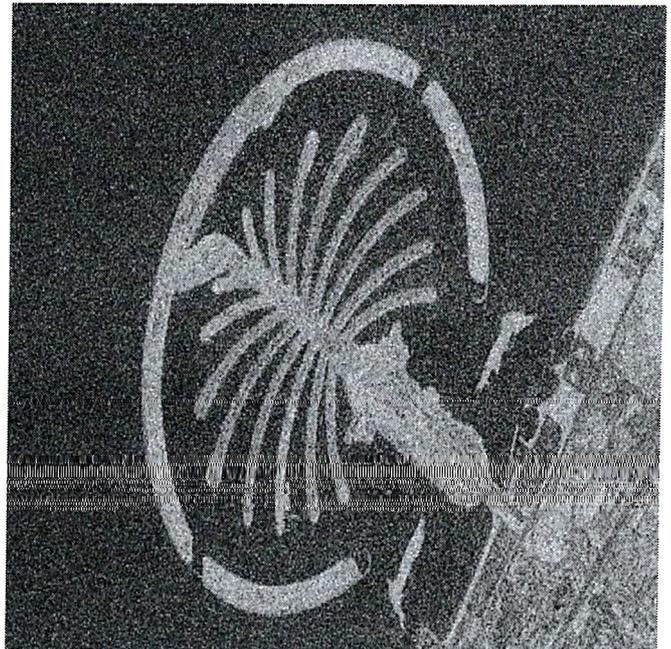


Image bruitée avec variance=12

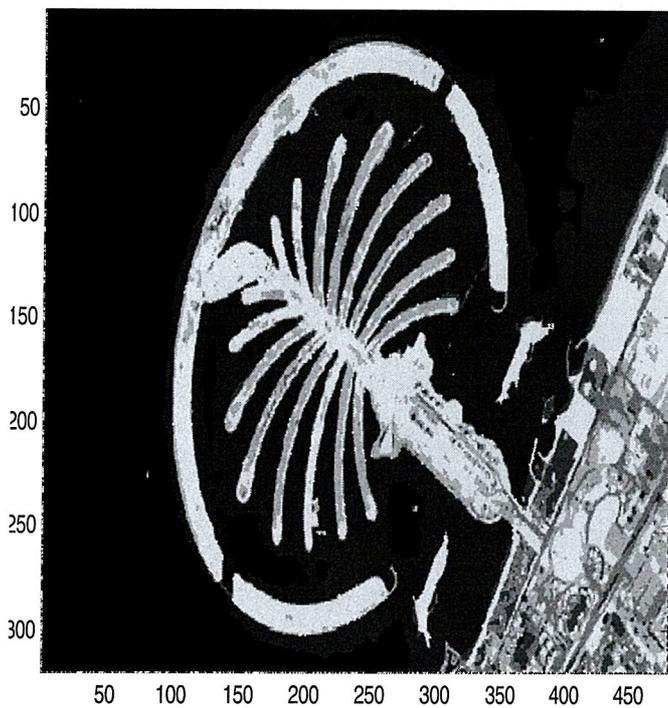


Image reconstituée

Figure 3.17. Exemple d'application sur l'image d'hôtel Emarate

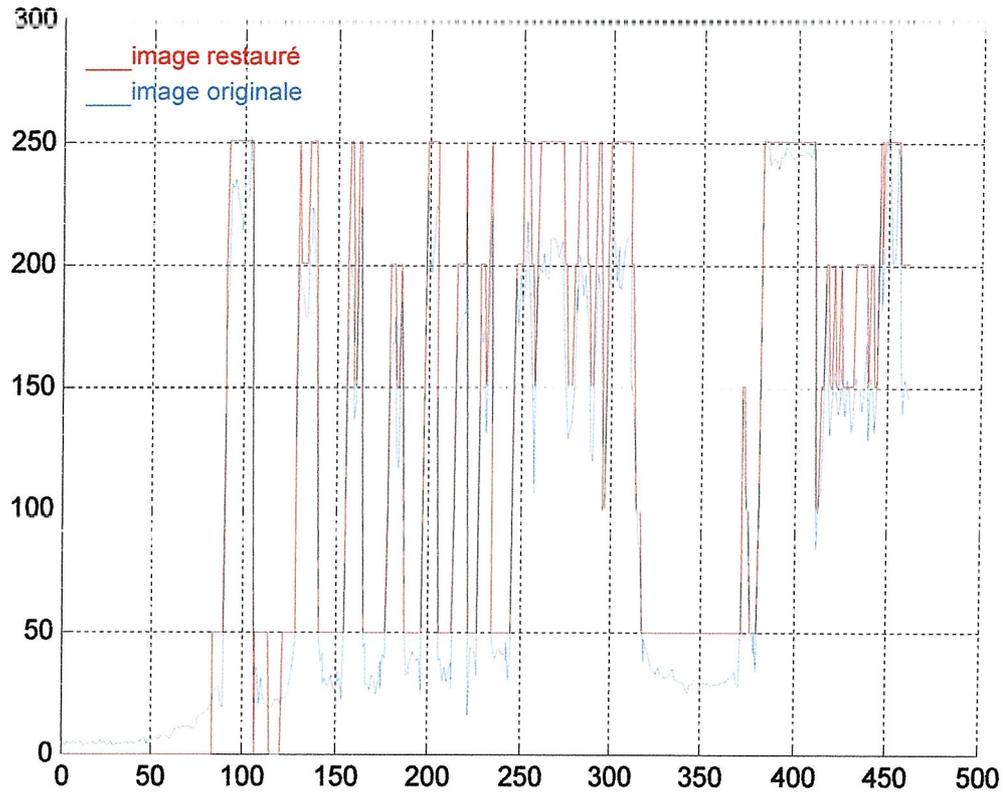


Figure 3.18. Profil l'image d'hôtel Emarate a la ligne 175

- Application a l'image de Cote méditerranéée

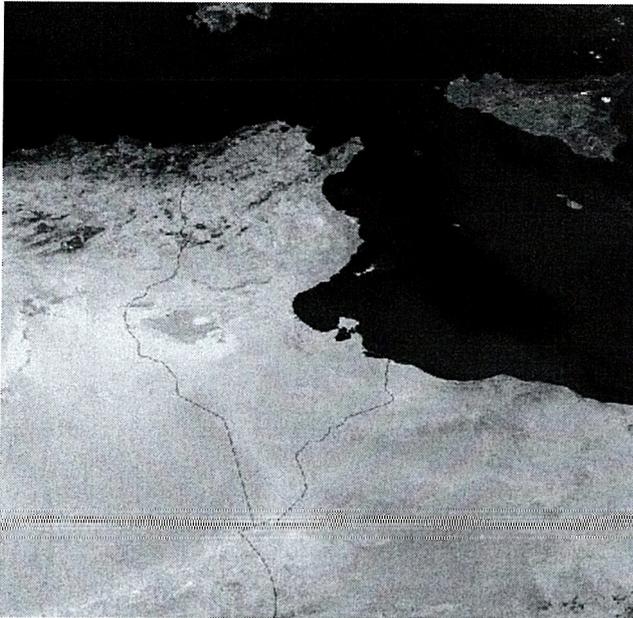


Image Originale

Image bruitée avec variance =12

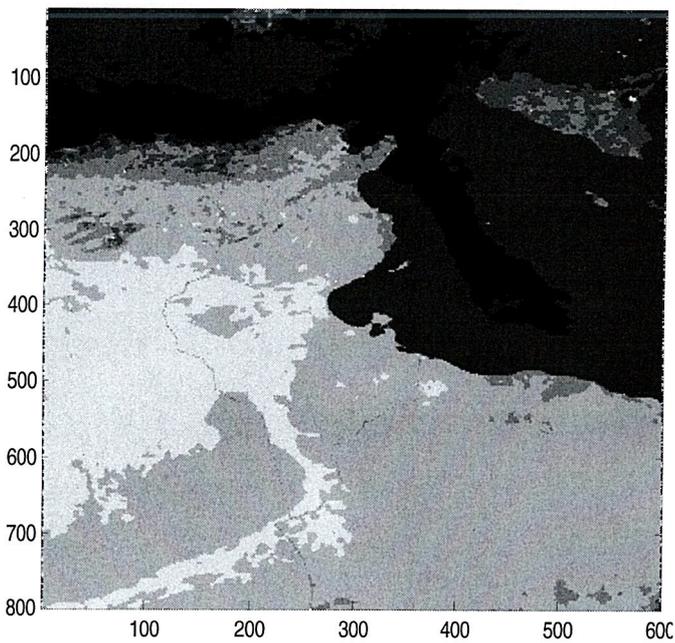


Image reconstituée

Figure 3.19. Exemple d'application sur l'image de cote méditerranéée

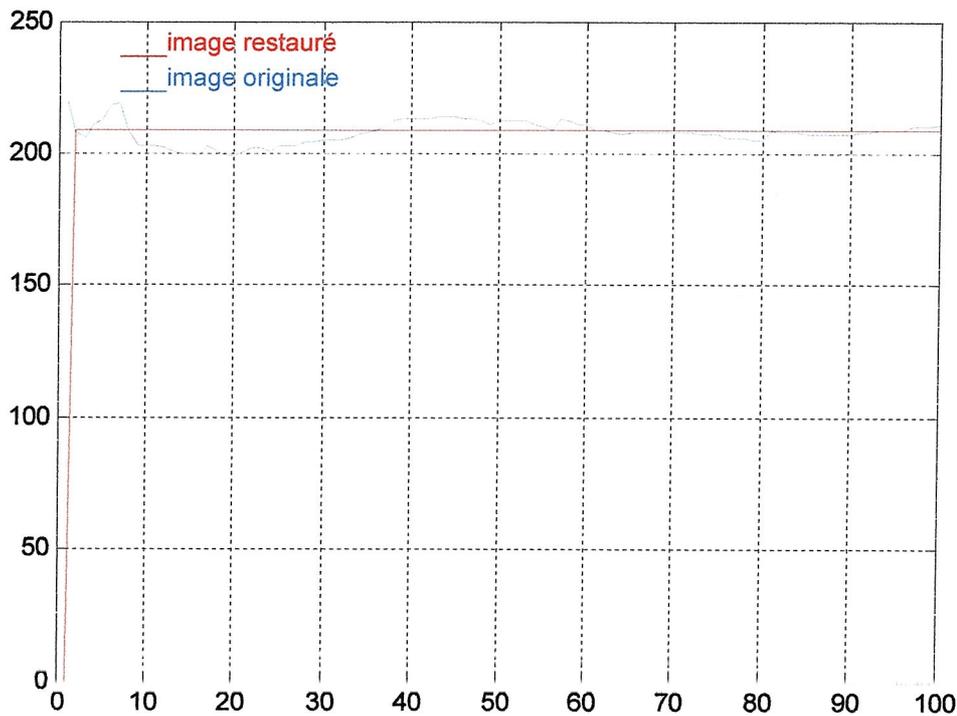


Figure 3.20. Profil l'image méditerranée a la ligne 475

### 3.14. Interprétation des résultats

La restauration s'obtient donc en minimisant l'énergie sur la sous image de coefficients, par l'algorithme recuit simulé. la décomposition faite par ondelette de Haar a 3 niveau ( $M=3$ ), L'image restaurée ou l'image finale est obtenue par synthèse avec l'image des coefficients restaurés (les suites de sous-images de basses fréquences et les suites de sous-images de hautes fréquences)

On prend comme pondération du terme de régularisation  $\beta=1$  et  $T=50, T_0=1$  et voisinage d'ordre 8.

Le temps CPU pour traiter ces images est réduite.

les résultats obtenus dans le tableau suivent :

Le nom d'image	Le temps CPU (sec)	SNR
1- Souk Ahras (250x400)	200	38.830
3- Hôtel Emarate (320x470)	210	39.920
3- méditerranée (800x600)	361	38.945

# **Conclusions & Perspectives**

## Conclusions & Perspectives

La restauration est une tâche importante dans le traitement d'images consiste à construire ou récupérer une image, qui a été dégradée en ayant a priori connaissance du phénomène de dégradation. Ce problème est comme la plupart des applications au traitement d'images qualifié de problème inverse mal posé. Il est alors nécessaire de régulariser le problème en imposant des contraintes sur la solution, en essayant de formaliser les connaissances a priori que l'on a de l'image à reconstruire. Cela permet non seulement d'obtenir une solution plus stable, mais aussi de construire une image acceptable

Nous avons présenté différents schémas de restauration utilisant des champs de Markov en analyse multirésolution.

La multirésolution permet d'améliorer la qualité des résultats et de diminuer les temps de calculs. L'analyse multirésolution est obtenue par décomposition en ondelettes biorthogonales successives, qui fournissent à chaque niveau de résolution des composantes basses et hautes fréquences de l'image . Dans un premier temps, seule la restauration des sous-images de basses fréquences est considérée . Il est ensuite montré qu'un traitement complémentaire de la composante haute fréquence augmente la qualité des résultats .

En perspectives, de nouvelles voies d'études champs de Markov et l'analyse multirésolution ont été ouvertes par ce mémoire, et pourraient être envisagées comme futurs axes de recherche :

- Combiner l'analyse multirésolution avec le champs de Markov pour la déconvolution (fou+bruit) d'images 2D et aussi dans le cas des images 3D.
- Restauration des images couleurs
- Développer des algorithmes de restauration des séquences d'images.
- Traiter l'effet de Bord

## BIBLIOGRAPHIE

- [1] E.H.L. Aarts and P.J.M. Van Laarhoven, "A parallel statistical cooling algorithm", *Proceed. of Symp. on The Theoretical Aspects of Computer Science*, Springer Lecture Notes in Computer Science, pp. 87-97, Jan. 1986.
- [2] R.Azencott, "Markov Fields and Image Analysis", 6ème Congrès de Reconn. des Formes et Intell. Artificielle, Antibes, France, Nov.1987, pp.1183-1191.
- [3] R.Azencott, "Image Analysis and Markov Random fields", *Proc. International Conf. on Indus. And Appl. Mathematics*, SIAM, Philadelphia, pp.53-61 (1988).
- [4] M.Bertero, T.Poggio, V.Torre, "Ill-posed problems in early vision", *Proc.IEEE* 76, pp.869-889 (1988).
- [5] J.Besag, "Spatial interaction and the Statistical Analysis of Lattice Systems", *J. Royal Stat. Soc. B* 34 75-83 (1972).
- [6] J.Besag, "On the Statistical Analysis of Dirty Pictures", *Journal Royal Stat. Soc.*, Vol.B-48, N.3, pp.259-302 (1986).
- [7] A.Blake and A.Zisserman, "Visual reconstruction", The MIT Press, Cambridge (1987).
- [8] H.Derin et al., "Modeling and Segmentation of Noisy and Textured Images Using MRF", *IEEE. Trans. Pattern Anal. and Machine Intell.* Vol.PAMI-9, N.1, pp.35-55, June (1989).
- [9] B.Diu, C.Guthmann, D. Lederer, B.Roulet, "Le ferromagnétisme", Complément III.J, §IV dans "Physique Statistique", Hermann Ed., (1989).
- [10] S.Geman and D.Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE transaction on Pattern Analysis and Machine Intelligence* 6(6), pp. 721-741 (1984).
- [11] B.Gidas, "Metropolis-type Monte Carlo simulation algorithms and simulated annealing", Technical report, Division of Applied Mathematics, Brown University, (1991).
- [12] W.K.Hastings, "Biometrika", 57, 97, (1970).
- [13] P.Lalande, "Détection du mouvement dans les séquences d'images selon une approche markovienne ; application à la robotique sous-marine", Thèse de l'Univ. de Rennes I, (1990).
- [14] L.Landau, E. Lifchitz, "Physique Statistique", §141, Ed. Mir, Moscou (1967).

- [15] S.Kirkpatrick et al., "Optimization by Simulated Annealing", *Science*, V.220, 671-80, (1983).
- [16] J.L.Marroquin, "Probabilistic solution of inverse problems", PhD thesis, MIT, Artificial Intell. Lab., (1985).
- [17] D.W.Murray et al., "Scene Segmentation from Visual Motion Using Global Optimization", *IEEE. Trans. Pattern. Anal. and Mach. Intell.* Vol. PAMI-9, No.2, pp.220-228, Jan. (1987).
- [18] N.Metropolis et al. "Equation of state calculations by fast computing machines", *J.Chemical. Phys.* Vol.21, 1087-1091, (1953).
- [19] G. Prémont, "Photothyristors PnpN et recuit simulé optoélectronique pour la vision bas-niveau", Thèse de l'université de Paris-Sud, Orsay, (1999).
- [20] D.Terzopoulos, "Image Analysis using Multi-grid relaxation methods", *IEEE, Trans. Pattern. Anal. and Machine Intell.*, Vol.PAMI-8, N.2, pp.129-139, March (1986).
- [21] J.Zerubia et al., "MFA using compound Gauss-MRF for edge detection and image restoration". Rapport interne de l'INRIA, Sophia-Antipolis, N.1295, Oct. 1990.
- [22] K.L.Chung, "Markov Chains with Stationary Transition Probabilities", Springer, N.Y., 1967.
- [23] A.D.Sokal, "Monte Carlo Methods in Statistical Mechanics : Foundations and New Algorithms", Troisième cycle de la physique en Suisse Romande, Semestre d'été (1989).
- [24] D.Geman, S.Geman, C.Geffine and P.Dong, "Boundary Detection by Constrained Optimization", *IEEE Trans. Pattern Anal. & Mach. Intell.*, PAMI 12(7), pp. 609-628 (1990).
- [25] B.B. Hubbard, *Ondes et ondelettes*, Pour la science, Belin, Paris, 1995.
- [26] Y. Meyer, *Wavelets : Algorithms et Applications*, Colin, Paris, 1992.
- [27] J. Neiryck and M. Kunt, *Traité d'électricité : traitement numérique des signaux*, Presses polytechniques et universitaires romandes, Lausanne, 2nd éd., 1989.
- [28] D. Gabor, "Theory of communication", *J. Inst. Elect. Eng.*, 93, pp. 429-457, 1946.
- [29] C.J. Goswami and A.K. Chan, *Fundamentals of Wavelets. Theory, Algorithms, and Applications*, John Wiley & Sons, 1999.

- [30] J. Morlet and al., "Wave propagation and sampling theory : Part II, Sampling theory and complex waves", *Geophysics*, vol. 47, n° 1, pp. 222-236, 1982.
- [31] I. Daubechies and J.C. Lagarias, "Two-scale difference equations I. Existence and global regularity of solutions", *SIAM J. Math. Anal.*, vol. 22, n° 5, pp. 1388-1410, 1991.
- [32] A. Antoniadis and R. Carmona, "Multiresolution analyses and wavelets for density estimation", Preprint, 1991.
- [33] J.F. Muzy, E. Bacry, and A. Arneodo, "The multifractal formalism revisited with wavelets", *Int. J. of Bifurcation and chaos*, vol. 4, pp. 245, 1994.
- [34] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges", *IEEE, Trans. Pattern. Anal. Machine Intell.*, vol. 14, pp. 710-732, 1992.
- [35] A. Bijaoui, "Algorithmes de la transformation en ondelettes : Application à l'imagerie astronomique", *Colloque Int. ondelettes et paquets d'ondes*, Rocquencourt (France), 1991.
- [36] E. Bournay Bouchereau, "Analyse d'images par transformées en ondelettes. Application aux images sismiques", Thèse de Doctorat de l'Université de Grenoble I, 1997.
- [37] P.J. Burt and H. Adelson, "Unitary representations of the affine groupe", *IEEE Trans. Comp.*, vol. 31, Apr. 1983.
- [38] M. Farge, N. Kevlahan, V. Perrier, and E. Goirand, "Wavelets and turbulence", *Proc. Of the IEEE* vol. 84, n° 4, pp. 639-669, Apr. 1996.
- [39] S. Mallat and W.L. Hwang, "Singularity detection and processing with wavelets", *IEEE, Trans. Inform. Theory*, vol. 38, n° 2, pp. 617-643, 1992.
- [40] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, PA, 1992.
- [41] Y. Meyer, *Ondelettes et opérateurs I - Ondelettes*, Hermann, Paris, 1990.
- [42] S. Mallat, "A theory of multiresolution signal decomposition : the wavelet representation", *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 11, pp. 674-693, 1989.
- [43] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.
- [44] M. Bourges-sévenier, "Réalisation d'une bibliothèque C de fonctions ondelettes", *Rapport interne IRISA*, 864, Sept. 1994.
- [45] C.K. Chui, *An introduction to wavelets*, Academic Press, Boston, 1992.

- [46] M.J. Shenzen, "The discrete wavelet transform : Wedding the à trous and Mallat algorithms", IEEE Trans. On Sig. Proc., vol. 40, pp. 2579-2590, oct. 1996.
- [47] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of  $L^2(\mathbb{R})$ ", Trans. Amer. Math. Soc., vol. 315, pp. 69-87, sep. 1989.
- [48] A. Cohen, "Biorthogonal wavelets", in [37], pp. 123-152.1989
- [49] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets", Comm. Pure and Appl. Math., vol. 45, pp. 485-560, 1992.
- [50] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure and Appl. Math., vol. 41, pp. 909-996, 1988.
- [51] J. Kovacevic and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for  $\mathbb{R}^n$ ", IEEE Trans. Info. Theory, vol. 38, pp. 533-555, mar. 1992.
- [52] E.P. Simoncelli and E.H. Adelson, "Non-separable extensions of quadrature mirror filters to multiple dimensions", Proceedings of the IEEE, vol. 78, n° 4, pp. 652-664,1990.
- [53] L. Blanc-Féraud, M. Barlaud ." Noisy Image Restoration Using Markov Random Field in a Multiresolution Scheme" Traitement du Signal volume 10 - n° 2.1991

## **Résumé**

La modélisation de l'image et de ses discontinuités par champ markovien a pour objectif dans la restauration d'image de préserver les contours.

Dans cette mémoire l'image est d'abord décomposée à différents niveaux de résolutions par transformée en ondelettes. Un champ markovien est défini pour chaque résolution et l'image est restaurée itérativement en fonction des résolutions croissantes. Cet algorithme est analysé comparativement à un algorithme monorésolution et montre son efficacité en terme de réduction de temps de calcul et qualité de l'image restaurée.