

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

présenté par

Ben rekia Othmane

pour l'obtention du diplôme master 2 en Électronique option système vision et robotique

Thème

Segmentation d'IRM Cérébrales par algorithme EM

Proposé par : Mr Kabir yacine

Année Universitaire 2012-2013

Remerciements

Je remercie ALAH de nous avoir donné la volonté et le courage qui nous ont permis de réaliser ce travail. Veuillez m'en garder dans le droit chemin.

Mes chaleureux remerciements vont tout d'abord à ma mère que nous avons soutenue par son amour, ses encouragements et son intérêt au long de mes études.

J'aime spécialement remercier mon promoteur Mr Kabir Yacine qui propose ce sujet et pour ses précieux conseils.

Je remercie vivement Mr Hezil Mohamed, Mrs Benali Khoudja Redha et Bellel, Mr Benyoucef Abdelatif, Mr Toufahi Mohamed, Mr Ahmed Hemi et Fodhala Okayl.

Mes derniers remerciements vont aux membres du jury pour avoir accepté d'examiner mon travail.

Depuis longtemps, le médecin n'a eu à sa disposition que des signes extérieurs et indirects pour se représenter les atteintes d'une maladie. L'expérience indispensable du praticien permettait alors de relier l'état du patient à une maladie particulière. La représentation qu'il pouvait avoir des organes atteints n'émanait que de timides suppositions fondées sur de précédentes autopsies. Mais tout ceci ne permettait pas au médecin de voir l'intérieur du corps de son patient, bien vivant, afin de se représenter l'état réel des organes malades et de suivre l'évolution de la maladie. L'invention du stéthoscope au XVIII^{ème} siècle par R. Laennec fut une véritable révolution : il était désormais possible d'entendre avec précision les bruits de certains organes et associer à un état sain ou pathologique de l'organe. Le stéthoscope a été le premier véritable instrument de diagnostic médical et reste encore le plus utilisé à l'heure actuelle.

Au fil des années, d'autres techniques furent mises au point dans le but d'observer cette fois ci les organes malades relatif à telle ou telle maladie. La découverte des rayons X en 1895 par W.C. Roentgen a permis d'examiner les parties (dures) du corps. La découverte de la radioactivité naturelle en 1898 par A. Becquerel et M. Curie, de la radioactivité artificielle par F. et I. Joliot-Curie en 1934, de la Résonance Magnétique Nucléaire (RMN) en 1946 par F. Bloch, l'amélioration des techniques de la physique puis finalement, dans ces dernières années, l'apparition de méthodes de traitement et de reconstruction d'images ont contribué à la création de scanners permettant de visualiser tout organe. Le médecin de cette décennie ne se contente plus de déduire l'état d'un organe, il peut le voir en image, localiser ses défauts, observer son comportement au cours du temps et apprécier l'effet d'une thérapeutique. L'ensemble de ces techniques récentes a ainsi donné naissance à *l'imagerie médicale*. L'une de ces techniques est l'Imagerie par Résonance Magnétique (IRM) que nous présenterons brièvement dans le premier chapitre Il s'agit d'une technique d'imagerie médicale non invasive et indolore qui permet d'observer l'anatomie d'un organe par une représentation 2D ou 3D. En 1971 aux Etats-Unis, R. Damadian [1] découvrait que des tissus sains et des tissus tumoraux donnaient des spectres de RMN différents. Cette différence est corrélée à la teneur en eau des tissus. C'est en 1973, toujours aux Etats-Unis, que P. Lauterbur obtînt les premières images du corps humain en IRM. Mais il fallut attendre 1980 pour pouvoir observer les premières images RMN in vivo de l'encéphale humain, fournies par Hawkes, Holland *et al.* [2] de l'université de Nottingham en Grande-Bretagne.

A l'heure actuelle, certaines techniques comme l'IRM fonctionnelle ou la TEP permettent d'étudier les mécanismes cérébraux mais celles-ci souffrent de l'absence d'informations relatives à l'anatomie du cerveau. Le lien entre l'étude fonctionnelle et l'information anatomique est cependant d'importance. En effet, les médecins envisagent par exemple d'utiliser la localisation des sillons corticaux, pour améliorer la qualité des résultats obtenus pour une tâche particulière. Ils espèrent également pouvoir profiter de cette information anatomique pour localiser et détecter des lésions observées en IRM chez des Scléroses En Plaques, et aussi détecter des changements de taille d'une tumeur cérébrale.

La segmentation des structures anatomiques cérébrales est un problème ardu et reste un sujet de recherche de forte actualité dans le domaine médical. La difficulté à segmenter une image est due à la complexité structurelle des images IRM et du contraste souvent insuffisant pour extraire la structure d'intérêt, sans aucune connaissance a priori sur sa forme et sa localisation.

L'objectif de ce mémoire est la segmentation des images cérébrales à partir des séquences d'images IRM. La segmentation proposée est basée sur l'utilisation d'un algorithme EM initialisé par un algorithme Kmeans, et afin de prendre en considération le contexte spatial nous introduisons les champs de Markov qui permettent de régulariser cette segmentation. Ce manuscrit est organisé en quatre chapitres

Chapitre 1 :

Dans ce chapitre, nous étudierons d'une manière succincte l'anatomie du cerveau humain et puis nous expliquerons le principe de l'Imagerie par Résonance Magnétique.

Chapitre 2 :

Ce chapitre débute par un rappel sur la théorie bayésienne ensuite présente une méthode classique pour l'estimation des paramètres maximum de vraisemblance et définit notre approche algorithme EM qui permet de compléter une série de données manquantes. Dans cette dernière partie on donne un exemple de l'application d'un algorithme EM sur mélange gaussien.

Chapitre 3 :

Décrit le processus de régularisation par champ de Markov

Chapitre 4 :

Il Présente les deux méthodes pour initialiser les paramètre puis l'implemontation des algorithme étudié .

La bibliographie : elle présente les différentes références utilisées, elle est décomposée en 2 parties : la bibliographie et la webliographie.

L'imagerie médicale est un domaine en plein développement ,le nombre croissant de travaux dans la littérature l'IRM s'accapare la plus grande part dans ces travaux en particulier l'IRM cérébrale qui est devenue un outil indispensable dans la pratique médicale et le nombre d'images toutes modalités confondues devient très important. Le traitement de ces images et leur interprétation par le praticien est un travail pénible et la subjectivité et l'imprécision pouvant affecter les résultats des analyses, c'est dans ce contexte que nous proposons de faire une segmentation par algorithme EM et champs de Markov.

Nous avons utilisé une segmentation basée sur la classification par algorithme EM, cette méthode permet de trouver les données manquantes pour obtenir une segmentation presque parfaite.

Nous avons donc appliqué notre méthode de régularisation par champ de Markov qui introduit la notion de voisinage, cette méthode robuste aux bruits, nous constatons que la segmentation par l'algorithme EM avec le champ de Markov donne un meilleur résultat parce que nous trouvons les données cachées et éliminons les bruits.

Résumé

Le but de ce projet est de combiner l'utilisation de l'algorithme EM et des champs de Markov pour segmenter en région distinctes une image avec initialisation des paramètres par K means .

La classification par algorithme EM permet de compléter une série de données manquantes, en se basant sur la vraisemblance maximale de l'ensemble des données.

Le formalisme des champs de Markov nous permet d'obtenir des classifications plus régulières en favorisant une régularité entre les pixels voisins .

Mots clés : algorithme EM, champs de Markov, segmentation, kmeans, classification.

Abstract

The goal of this project is to combine the use the EM algorithm and the fields of Markov to segment in distinct areas in image , with initialization of the parameter by kmeans.

The classification by EM algorithm allows to supplement a series of missing (hidden) data basing on Maximum-likelihood estimation of the whole of the data .

The formalism of the fields of Markov allows us to obtain more regular classification by supporting a regularity between the neighbors pixels .

Key word : EM algorithm ,fields of Markov ,segmentation ,k means ,classification .

ملخص

الهدف من هذا المشروع يتمثل في جمع بين الخوارزمية EM و حقل ماركوف لتجزئة الصورة الى مناطق واضحة مع تهيئة المقاييس ب K متوسط

تصنيف بخوارزمية EM يسمح بملا سلسلة المعطيات الناقصة و نركز على التوقعات العظمى لمجموعة المعطيات. عبارة حقل ماركوف يسمح لنا بإعطاء الترتيبات منتظمة وخاصة بين يكسل مجاورة .

الكلمة الرئيسية تجزئة حقل ماركوف ترتيبات التوقعات العظمى K متوسط

Sommaire

Liste des figures.....	1
Liste des tableaux.....	2
Notations.....	3
Chapitre1 : anatomie du cerveau et Imagerie par résonance magnétique	
1.1 L'encéphale, centre du système nerveux central	9
1.1.1 Le cerveau	10
1.1.2 Le cervelet	11
1.1.3 Le tronc cérébral.....	11
1.2 Les principaux tissus cérébraux	12
1.3 Couches externes au cerveau.....	15
1.4 Présentation des pathologies cérébrales	15
1.4.1 Sclérose en plaque (SEP).....	16
1.4.2 Accident vasculaire cérébrale (AVC):	16
1.4.3 Tumeurs cérébrales	17
1.5 Différents plans de coupes.....	19
1.6 L'imagerie par résonance magnétique (IRM).....	20
1.6.1 Principes physiques élémentaires de l'IRM.....	20
chapitre 2 : ALGORITHME EM	
2.1 Théorie bayésienne.....	32
2.1.1 L'estimateur du MAP :.....	33
2.1.2 L'estimateur du MPM :.....	34
2.1.3 L'estimateur du champ moyen :.....	34
2.2 Estimation par maximum de vraisemblance.....	34
2.3 L'algorithme EM.....	36
2.3.1 Rappels sur Inégalité de Jensen	36
2.3.2 Itérations EM.....	37
2.3.3 Preuve de la croissance de la vraisemblance d'une itération à l'autre.....	38

2.4	Modèle de mélange :	42
-----	---------------------	----

Chapitre 3 : CHAMP DE MARKOV

3.1	Voisinage et clique.....	48
3.1.1	Notion de voisinage	48
3.1.2	Notion de clique	49
3.2	Champs de Markov et de Gibbs.....	49
3.2.1	Champ de Markov	49
3.2.2	Champ de Gibbs	51
3.3	Modélisation markovienne de la segmentation.....	51
3.3.1	Modèle de Potts sans champ externe.....	52
3.3.2	Modèle de Potts avec champ externe	52
3.4	Champs de Markov cachés.....	52
3.5	Champs de Markov cachés à bruit indépendant.....	54
3.6	Processus de segmentation.....	55
3.7	Estimation des paramètres.....	55
3.7.1	Algorithme ICM.....	55
3.7.2	Algorithme EM	56

Chapitre 4 : IMPLEMENTATION ET RESULTAT

4.1	Implémentation	60
4.1.1	Environnement de travail.....	60
4.1.2	Présentation de logiciel Matlab	60
4.1.3	Logiciel MRICRO.....	61
4.1.4	Implémentation de l'algorithme EM.....	62
4.2	Résultat.....	64
4.2.1	Segmentation par algorithme EM sur image 2D.....	64
4.2.2	Segmentation par algorithme EM avec champ de Markov sur image 2D.....	65
4.2.3	la différence entre segmentation par algorithme EM et segmentation par algorithme EM avec modélisation Markovienne	66
4.2.4	Segmentation par algorithme EM avec champ de Markov sur image 3D.....	67
	Conclusion	70

Annexe A

Annexe B

Bibliographie

Ce chapitre est une introduction aux disciplines qui constituent la base et le contexte de L'étude sur l'extraction des tissus cérébraux.

La première partie concerne l'anatomie du système nerveux central humain. Elle présente les notions élémentaires de l'anatomie cérébrale et définit les termes et concepts essentiels qui permettent de mieux comprendre ce que l'on observe avec l'imagerie cérébrale.

La seconde partie est consacrée à l'Imagerie par Résonance Magnétique (IRM). Elle présente d'abord les principes physiques élémentaires de l'IRM et ensuite les paramètres de séquence ainsi que les différentes séquences de l'IRM, et enfin les défauts des images acquises par l'IRM.

1.1 L'encéphale, centre du système nerveux central

L'encéphale (du grec en, dans, et képhale, tête ; qui est placé dans la tête) est la partie du système nerveux contenue dans la boîte crânienne et reposant sur la base du crâne. Il comprend le cerveau, le cervelet et le tronc cérébral (voir Figure 1.1) [3]:

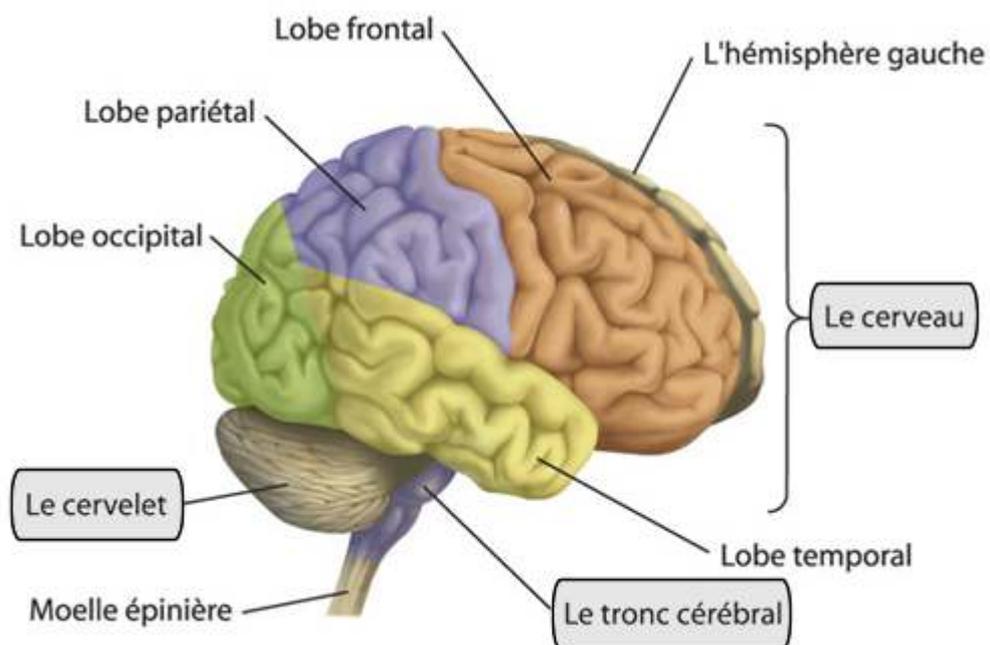


Figure 1.1 : Système nerveux central (Encéphale)

1.1.1 Le cerveau (du latin cerebrum)

Bien que représentant seulement 2% du poids total du corps humain (soit environ 1,4 Kilogrammes), le cerveau gère directement ou indirectement 98 % de ses fonctions. Il est Responsable des fonctions humaines les plus complexes comme la pensée, la résolution de Problèmes, les émotions, la conscience et les comportements sociaux, et régit les fonctions Essentielles du corps comme la respiration, le processus d'alimentation, le sommeil, les mouvements et les cinq sens. Le cerveau est la partie la plus volumineuse de l'encéphale. Il est composé du télencéphale et du diencephale. Le télencéphale est constitué de deux hémisphères séparés par la scissure inter-hémisphérique où pénètrent les méninges. Il est formé de substance grise externe appelée cortex, de substance blanche sous-jacente et de structures profondes appelées noyaux gris centraux. Le diencephale réunit les deux hémisphères cérébraux par le *corps calleux* et par le *trigone cérébral*. Il fait la jonction entre le télencéphale et le tronc cérébral. Ses principaux constituants sont le thalamus et l'hypothalamus.

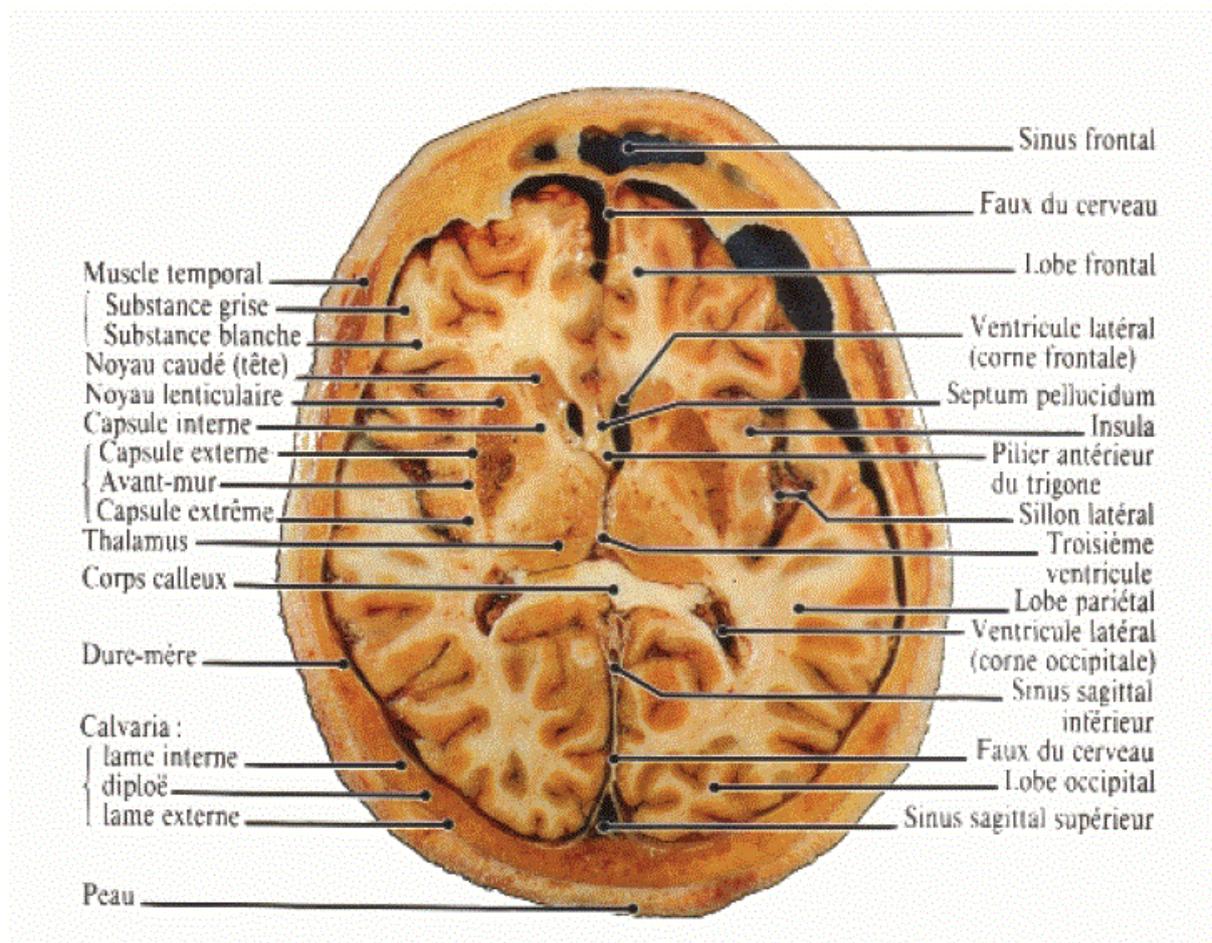


Figure 1.2: Photographie d'un cerveau post-mortem, visualisation des principaux éléments

1.1.2 Le cervelet (du latin cerebellum, diminutif de cerebrum)

est situé en parallèle sur deux grandes voies nerveuses : celles qui amènent les messages sensoriels vers les zones du cortex qui en font l'analyse ; et celles qui partent du cortex et descendent vers les muscles pour les faire se contracter. Le cervelet est en partie impliqué dans la coordination des mouvements du corps.

1.1.3 Le tronc cérébral

Le tronc cérébral (figure 1.3) se compose de trois parties : le *myélocéphale* (ou moelle allongée ou bulbe rachidien), le *mésencéphale* (ou isthme) et le *pont* (ou protubérance). Il est formé de substance blanche externe et de substance grise interne morcelée en de multiples noyaux de substance grise ou noyaux gris. La substance blanche du tronc cérébral est constituée de fibres sensibles dont la plupart vont au cerveau, et de fibres motrices en provenance du cerveau. Le tronc cérébral a donc un rôle conducteur mais il est également responsable de certaines fonctions inconscientes liées à la déglutition et à la respiration. Le cervelet, également constitué de substance blanche et de substance grise, est relié à chacune de ses trois parties par des pédoncules. Il possède trois lobes, comportant de nombreux sillons de plus petites tailles que ceux du cerveau, et a la charge du maintien général de la coordination (contractions musculaires et équilibre du corps). Il dispose, à sa surface, d'un cortex d'environ 1 mm d'épaisseur. [4]

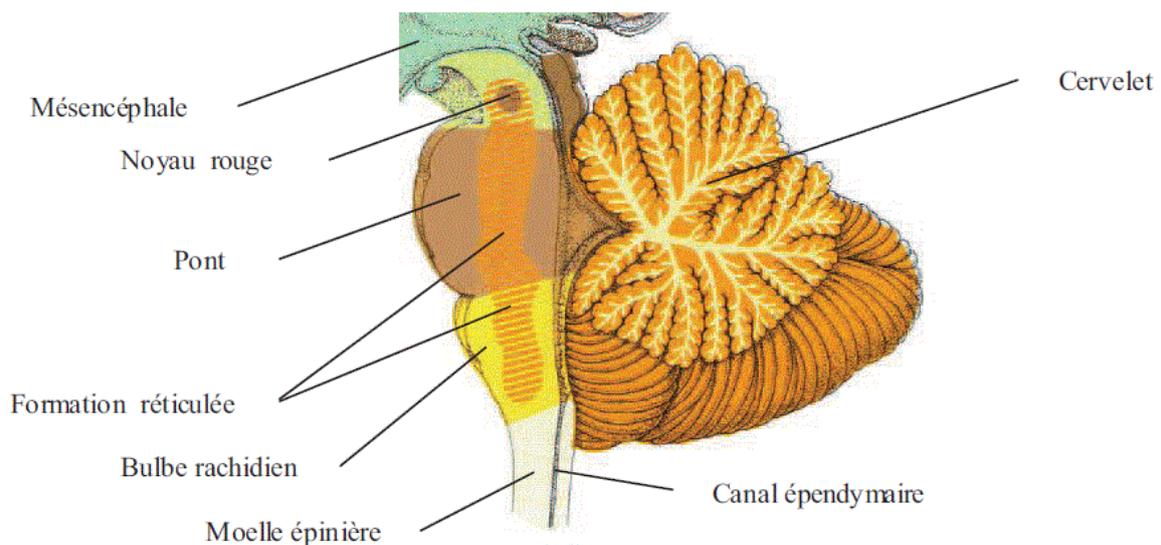


Figure 1.3 : Le tronc cérébral.

1.2 Les principaux tissus cérébraux

Outre la présence d'artères et de veines cérébrales, de tissus servant de parois, ou de nombreuses petites structures telles des glandes, l'encéphale contient principalement trois substances le liquide céphalo-rachidien(LCR), la substance grise et la substance blanche (voir Figure 1.4)

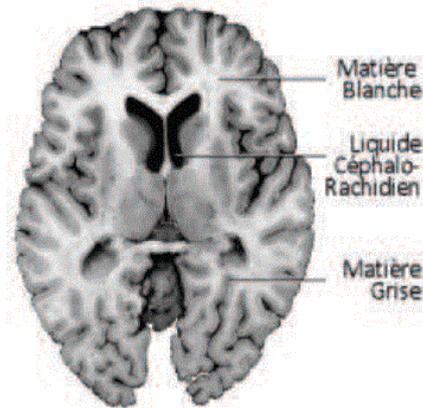


Figure 1.4: Une coupe histologique d'un cerveau mettant en évidence les trois matières principales de l'encéphale.

❖ Le liquide céphalo-rachidien (LCR)

Le liquide céphalo-rachidien (ou liquide cérébro-spinal, LCS) est le liquide dans lequel baignent le cerveau et le cervelet. D'un volume moyen de 150 ml, il est absorbé par le système veineux cérébral et continuellement renouvelé. Ses principaux rôles sont :

- protection mécanique du système nerveux central contre les chocs par amortissement des mouvements,
- protection contre les infections, car il contient les médiateurs de l'immunité Humorale et cellulaire,
- transport des hormones entre les différentes régions éloignées du cerveau.

❖ La substance grise

La substance grise (ou Matière Grise MG), correspond au corps cellulaire des neurones avec leur dense réseau de dendrites. Les principales structures de substance grise sont le *cortex*, à la périphérie du cerveau, les *noyaux gris centraux* et les *noyaux du diencéphale*

- **Le cortex**

Le cortex recouvre la totalité du télencéphale en une couche de 3 à 5 mm d'épaisseur et contient 70% de l'ensemble des neurones du cerveau. Du fait de cette grande densité de neurones, le cortex apparaît gris terne à l'oeil dans les images pondérées en *T1* obtenues par IRM. Il est le centre de la conscience. Le cortex est essentiellement formé de corps cellulaires de neurones et de prolongements dendritiques, dépourvus de myéline (Il contient 10 milliards de cellules nerveuses). Sa surface est particulièrement étendue. plissée en de nombreux sillons qui délimitent des circonvolutions. Deux tiers de la surface du cortex est enfouis dans les sillons. Les sillons les plus importants sont appelés scissures (de *Sylvius*, de *Rolando*) et divisent les hémisphères en quatre lobes : le lobe *frontal*, le lobe *pariétal*, le lobe *occipital* et le lobe *temporal* (voir Figure 1.1). Puisque les sillons corticaux constituent la base et le contexte de notre travail, nous allons consacrer une section pour décrire ces structures anatomiques

- **Les noyaux gris centraux**

Les noyaux gris centraux sont également formés de corps cellulaires neuroniques mais avec une densité moins importante que dans le cortex. Ils sont composés des noyaux du télencéphale, ces noyaux sont appelés les noyaux de base (ou corps strié), parmi eux, on distingue les *noyaux caudés* et les *noyaux lenticulaires*.

Le noyau caudé : En forme de virgule à grosse extrémité il est presque complètement enroulé autour du thalamus. Il longe à peu près dans toute son étendue le ventricule latérale. On lui distingue trois parties : la tête, le corps et la queue.

Noyaux lenticulaires : Le noyau lenticulaire se situe en dehors du noyau caudé et thalamus. Sa forme est celle d'une lentille biconvexe, triangulaire sur les coupes axiales et coronales. Le noyau lenticulaire se compose du putamen (externe), et du pallidum (interne).

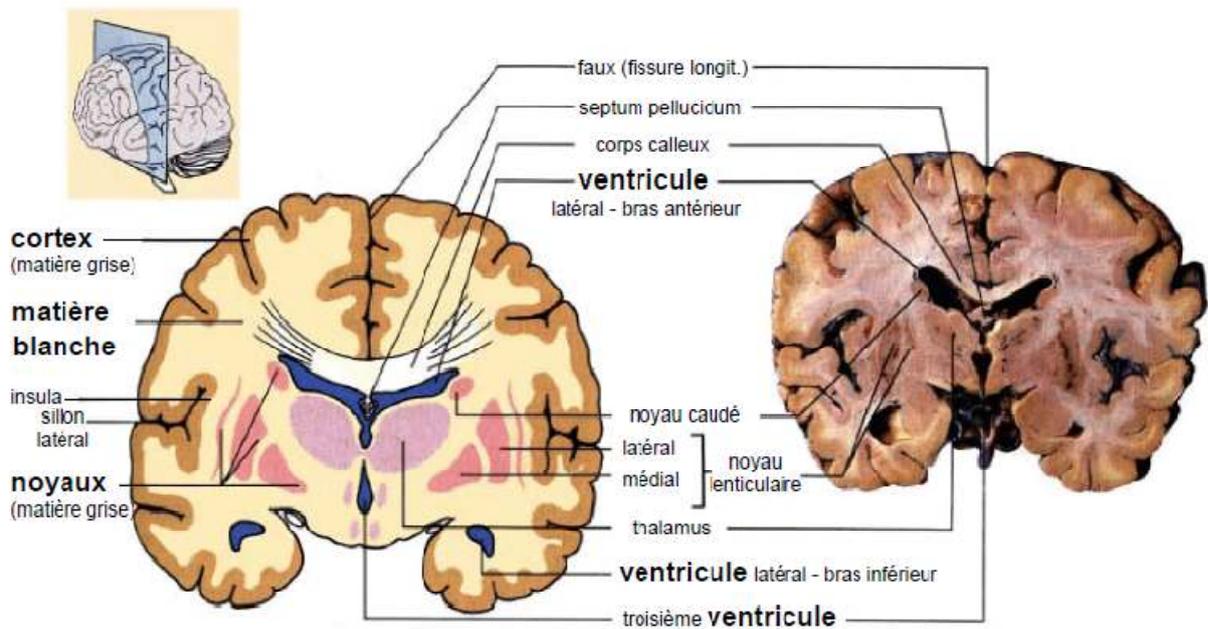


Figure 1.5 : quelques structures cérébrales

- **Les noyaux du diencéphale**

Parmi ces noyaux :

- **Le thalamus** : le thalamus est une masse grise qui flanque la cavité du troisième ventricule, aboutissent toutes les sensations avant qu'elles soient projetées vers la conscience. Le thalamus est connecté aux centres moteurs et coordinateurs.
- **L'hypothalamus** : l'hypothalamus est la paroi inférieure du troisième ventricule, se prolonge jusqu'à l'hypophyse. De petite dimension, il a la charge des équilibres, physiologiques du corps. Les nouveaux atlas associent dorénavant le pallidum au groupe de noyaux gris du diencéphale [5]

- ❖ **La substance blanche**

La matière blanche (MB), est constituée de fibres à myéline, organisées en fibres d'association reliant deux régions d'un même hémisphère, en fibres commissurales inter-hémisphériques passant par le corps calleux et le trigone cérébral et en fibres de projection entre les différentes parties du corps et le télencéphale. Toutes ces fibres, sensibles ou motrices, aboutissent aux centres nerveux sous-jacents ou en viennent. Ces

centres sont les corps striés, les noyaux gris du diencephale, le cervelet, le tronc cérébral et la moelle épinière. La substance blanche sous-corticale est majoritairement constituée d'axones à myéline provenant des neurones du cortex. [4]

1.3 Couches externes au cerveau

Le cerveau est entouré d'une série de couches successives de tissus protecteurs (**fig1.6**) : tout d'abord, de l'intérieur vers l'extérieur, **les méninges**, puis l'os et enfin le scalp (peau). Les méninges sont constituées de trois membranes successives (**fig1.6**) : **la pie mère**, qui transporte les vaisseaux sanguins, et est en contact direct au cerveau. Ensuite se trouve la **membrane arachnoïdienne** ressemblant à une toile d'araignée. Le LCR circule entre celle-ci et la pie mère. La dernière membrane extérieure, plus épaisse et résistante est **la dure mère**. Elle protège les couches précédentes, plus fines, de l'os. **L'os** sombre en IRM comporte en son centre des portions contenant de la moelle osseuse, claire en IRM. Une fine membrane, appelé **le périoste** est située immédiatement à l'extérieur de l'os.

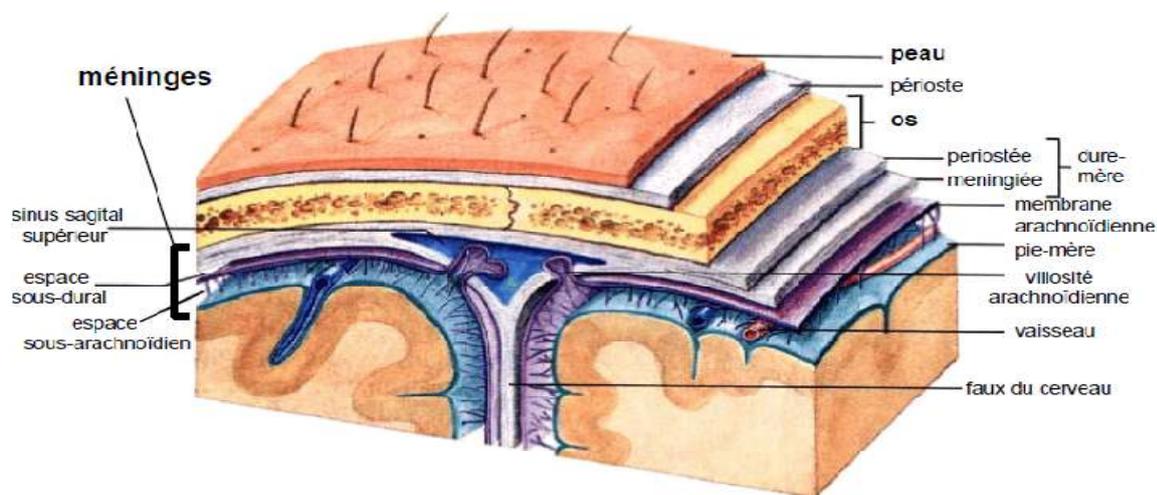


Figure 1.6 Couche externe au cerveau

1.4 Présentation des pathologies cérébrales

Les pathologies cérébrales peuvent être classées en six catégories :

Les pathologies expansives, les pathologies vasculaires cérébrales, les pathologies due à un traumatisme (accident), les pathologies démyélinisâtes et inflammatoires, les pathologies infectieuses et les autres pathologies telles que les pathologies dégénératives (maladie d'Alzheimer, maladie de parkinson).

1.4.1 Sclérose en plaque (SEP) :

La **sclérose en plaques** (SEP) est une maladie Inflammatoire chronique du **système nerveux central**. Elle se caractérise par la destruction de la myéline (gaine protectrice du nerf) et par un dépôt anormal de tissu conjonctif et durcissement de ce dernier (sclérose), visible sous forme de plaques. La myéline est une substance lipidique (graisseuse) entourant les fibres nerveuses. Cette destruction rend la transmission nerveuse difficile chez le malade et provoque des troubles de la vision, moteurs et autres.

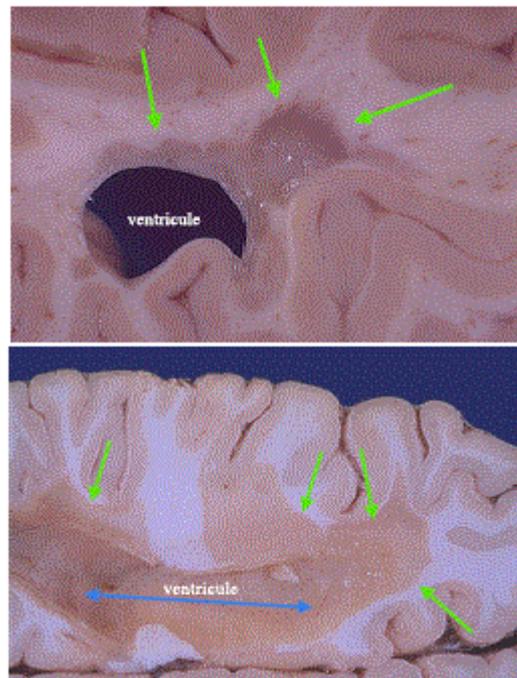


Figure 1.7. Sclérose en plaque

1.4.2 Accident vasculaire cérébral (AVC):

Un accident vasculaire cérébral est un arrêt brutal de l'irrigation d'une partie du cerveau il survient lors de l'obstruction ou de la rupture d'une artère du cerveau appelée également artère cérébrale. L'AVC est classé par localisation anatomique, distribution vasculaire, étiologie, âge de l'individu affecté et **nature hémorragique ou ischémique de l'accident**. (L'obstruction est un accident ischémique, et la rupture est un accident hémorragique). Les tissus qui ne reçoivent plus d'irrigation vont mourir progressivement (les spécialistes en médecine cardio-vasculaire disent qu'ils se lysent). Ce phénomène fait suite à un arrêt de l'oxygénation de la partie du cerveau concerné.

Les conséquences sont plus ou moins graves : malaise passager, paralysie, coma ; avec des **séquelles plus ou moins importantes.**

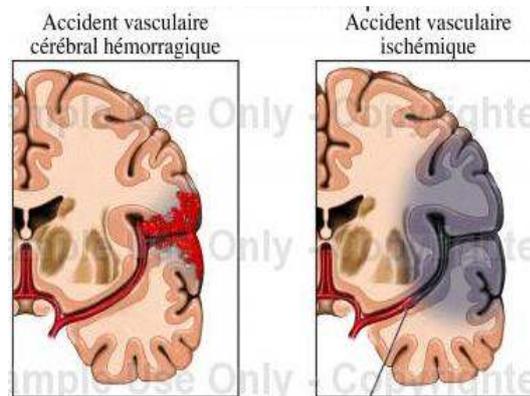


Figure 1.8- AVC importantes.

1.4.3 Tumeurs cérébrales :

A- Définition d'une tumeur :

Une tumeur est une prolifération de cellules nouvelles composant un tissu pathologique (malade). Ceci est le résultat d'une activité anormale de ces cellules, qui se poursuit sans raison apparente, les cellules ayant tendance à persister ou à augmenter de volume. Ces cellules n'ont pas obligatoirement de rapport avec un éventuel processus inflammatoire. Cette prolifération aboutit à une augmentation du volume d'un organe ou d'une partie de l'organe ce qui traduit la présence du processus pathologique

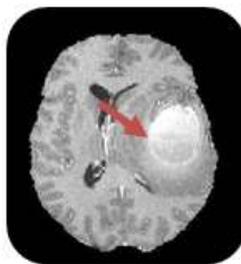


Figure 1.9- Présence d'une tumeur

B -Tumeurs cérébrales bénignes et tumeurs cérébrales malignes?

Les tumeurs cérébrales bénignes sont constituées de cellules qui croissent lentement et qui, à l'analyse microscopique n'ont pas de caractère d'activité importante. Une tumeur cérébrale bénigne peut être soit simplement surveillée radiologiquement soit être extirpée chirurgicalement complètement ou de façon partielle. Si une tumeur bénigne n'est pas complètement enlevée, elle peut récidiver.

Les tumeurs cérébrales malignes, classées « cancéreuses », définissent certaines tumeurs primaires ainsi que toutes les lésions cérébrales métastatiques. Elles sont constituées de cellules qui se divisent relativement vite. Ces tumeurs croissent donc rapidement et peuvent envahir et endommager les structures cérébrales importantes. Elles peuvent être traitées par la chirurgie, la radiothérapie, la chimiothérapie ou la combinaison de celles-ci.

C -Classification des tumeurs cérébrales :

Cette question s'avère difficile, cette classification faisant intervenir à la fois la localisation de la tumeur, sa nature (nature de la cellule devenue cancéreuse), et son grade (degré de malignité). De plus, plusieurs normes existent. Ainsi, dans le cerveau, on trouve :

- **Les cellules du système nerveux, dites cellules gliales :**

Qui dégénèrent en gliomes. On peut citer :

- Les astrocytes (type I et II), qui atteints par le cancer donnent des astrocytomes,
- Les oligodendrocytes, qui produisent des oligodendrogliomes,
- Les épendymocytes, qui provoquent des épendymogliomes.

- **Les cellules non gliales :**

- les neurones dégénèrent en neurocytomes,
- les cellules méningées forment des méningiomes
- cette catégorie de cellules provoque également les lymphomes, les tumeurs neuro ectodermes primitives, les gangliomes, les médulloblastomes...

- **Les métastases :**

Sont des tumeurs secondaires provenant de cellules tumorales d'un autre organe et ayant essaimé . Cette liste donne un aperçu non exhaustif de la nature des différentes tumeurs ; à cela, il faut ajouter son grade (de I à IV selon la norme de l'OMS), qui indique son degré de malignité.

Cette liste contient quatre classes de tumeurs :

- Les gliomes de Haut Grade (HG). Ce sont les plus fréquents.
- Les oligodendrogliomes (OD).
- Les méningiomes (MN).
- Les métastases (MT).

A- Quels sont les signes et symptômes associés à la présence de tumeurs cérébrales?

Les manifestations cliniques d'une tumeur cérébrale sont variables. En effet, elles dépendent de la localisation, du volume de la lésion et de l'hypertension intracrânienne et l'hydrocéphalie et/ou de l'œdème associées ou non.

<ul style="list-style-type: none"> ✓ céphalées ✓ crises d'épilepsie ✓ vertiges/ étourdissements 	<ul style="list-style-type: none"> ✓ déficits moteurs ou sensitifs ✓ nausées matinales/vomissements ✓ troubles auditifs 	<ul style="list-style-type: none"> ✓ troubles visuels ✓ changement de la personnalité
--	--	---

Tab 1.1- signes et symptômes de la présence des tumeurs

1.5 Différents plans de coupes

L'observation par des coupes en deux dimensions du cerveau peut être effectuée selon plusieurs angles de vue. Ainsi, on distingue trois axes anatomiques pour réaliser les coupes (fig. 1.10).[6]

1. Coupes axiales : Ces coupes correspondent quasiment à un plan horizontal. En imagerie de résonance magnétique, elles correspondent à un plan perpendiculaire à l'axe du champ magnétique principal.
2. Coupes sagittales : Ces coupes sont prises dans des plans parallèles au plan inter-hémisphérique. Il s'agit de vues latérales du cerveau.
3. Coupes coronales : Ce sont des coupes perpendiculaires aux coupes axiales et sagittales Elles correspondent à des vues de face du cerveau.

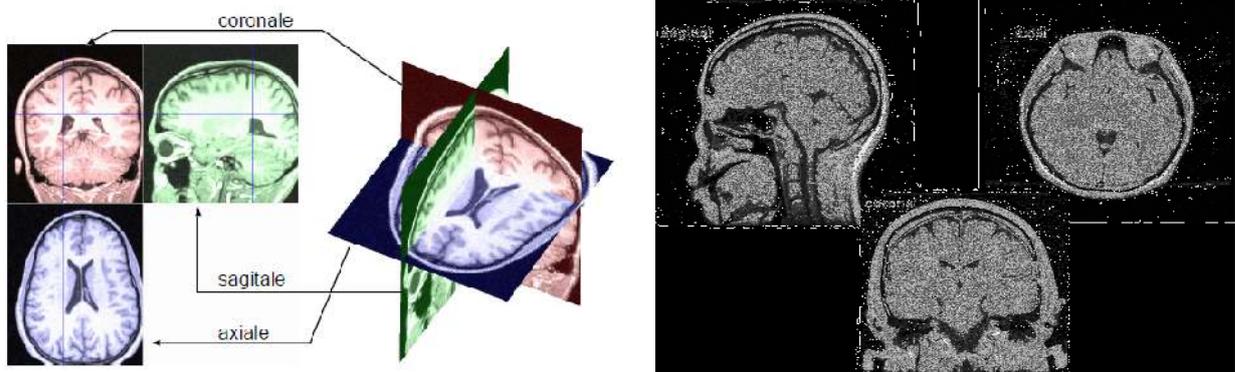


Figure 1.10 Différentes coupes du cerveau

1.6 L'imagerie par résonance magnétique (IRM)

La première image obtenue par résonance magnétique nucléaire, ou IRM, a été obtenue en 1973 par Lauterbur à la suite des travaux de Bloch et Purcell sur la résonance magnétique nucléaire (RMN). Le principe consiste à mesurer l'aimantation nucléaire des tissus biologiques et à reconstruire une image à partir de ces aimantations. En pratique, le noyau d'hydrogène (proton) est surtout étudié, en raison de sa grande abondance et de sa grande spécificité.

Intérêt de l'IRM :

L'intérêt de l'IRM est multiple :

- a) L'IRM *anatomique* fournit des détails anatomiques avec une résolution de l'ordre de millimètre et permet de différencier les tissus mous du cerveau, ce qui justifie son choix dans notre étude .
- b) L'IRM *fonctionnelle* (IRMf) permet de visualiser des information sur l'activité cérébrale à partir de séries temporelles d'images (plusieurs centaines d'images !5 s par image) de basse résolution (!4 mm)
- c) L'IRM *angiographique* (ARM) , en saturant les tissus fixe et en rehaussant le sang en mouvement , permet de visualiser l'arbre vasculaire cérébrale sans l'injection d'un produit de contraste
- d) L'IRM de *diffusion* détermine la direction de diffusion des molécules d'eau et par conséquent celles des fibres tissulaire.
- e) L'IRM par *transfert d'aimantation* fournit d'autres informations physiologiques liées aux échanges d'aimantation entre les tissus et l'eau. Un appareil IRM permet également d'analyser les propriétés chimiques d'une région donnée (spectroscopie) avec de nombreuses applications médicales.

1.6.1 Principes physiques élémentaires de l'IRM

Cette section résume les principes élémentaires de l'imagerie par résonance magnétique.

Moment magnétique des noyaux

L'existence du moment magnétique des noyaux a été montrée simultanément par Bloch et Purcell. Le corps humain étant constitué en moyenne de 70 % d'eau, on s'intéresse en pratique à la molécule d'eau et en particulier au noyau d'hydrogène (proton). Le noyau

d'hydrogène se comporte comme une charge en rotation autour de son axe : c'est le mouvement de spin (**fig 1.11**). Les protons peuvent alors être assimilés à des dipôles magnétiques.

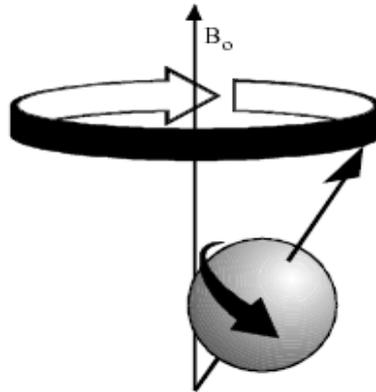


Fig.1.11- Le mouvement de spin

En l'absence de tout champ magnétique, ceux-ci vont s'orienter dans l'espace de façon aléatoire. Ce mouvement confère au noyau un moment cinétique qui dépend de sa masse et un moment magnétique qui dépend de sa charge. Lorsqu'un proton est placé dans un champ magnétique B_0 , son moment magnétique est animé d'un mouvement de précession (**Fig 1.12**) autour de l'axe de B_0 et sa fréquence de rotation est exprimée par la relation de Larmor suivante :

$$f_0 = \frac{\gamma}{2\pi} B_0 \quad (1.1)$$

Avec : γ : le rapport gyromagnétique spécifique au noyau étudié

Dans le cas de proton, la mécanique quantique n'autorise que deux états possibles de spin : $(-1/2, +1/2)$.

Les protons s'orientent alors par rapport à B_0 et décrivent autour de ce champ un mouvement de précession, de vitesse angulaire constante. L'ensemble des protons s'orientant dans le sens de B_0 forme alors une sous-population de spins $+1/2$; le reste s'orientant dans le sens inverse forme la sous-population de spins $-1/2$. La sous population de spins $+1/2$ étant la plus importante, il existe une aimantation résultante M_0 proportionnelle au nombre de protons présents. M_0 ne possède qu'une composante longitudinale, dans le sens de B_0 .

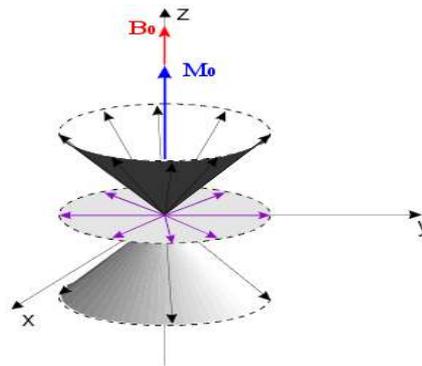
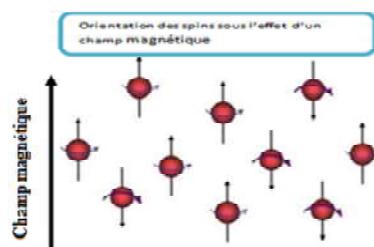


Fig1.12- Mouvement de précession d'un ensemble de moments magnétiques

Le moment magnétique résultant M_0 sera orienté dans le sens de B_0 , sa composante en Z, M_0 est petite devant B_0 . Pour mesurer ce moment magnétique, il faut donc le basculer dans le plan transversal (x, y) . Pour cela on applique une onde radiofréquence (RF) égale à f_0 et perpendiculaire à B_0 .



1.6.2 Phénomène de résonance et de relaxation :

Les protons qui étaient alignés selon B_0 , reçoivent alors un apport d'énergie sous forme d'onde radiofréquence(RF) créée par un champ magnétique B_1 , non colinéaire à B_0 , et tournant à la fréquence f_0 , ils résonnent donc. Ils passent d'un niveau bas à un niveau haut de l'énergie, cette transition correspond au **phénomène de la résonance magnétique** et le champ macroscopique est basculé de sa position d'équilibre M_0 vers une position M tant que B_1 dure. On décompose alors M en un moment magnétique longitudinal ML (ou Mz) et en un moment magnétique transversal MT (ou Mxy). A l'arrêt de B_1 , un signal dit de précession libre est enregistré. Il accompagne le retour à la position d'équilibre (en spirale) du vecteur M .

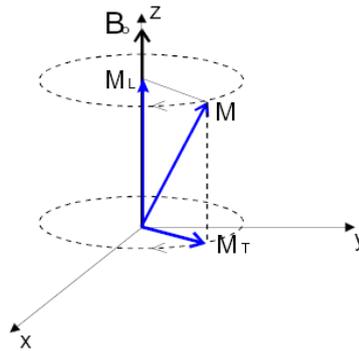


Fig 1.13. Excitation par l'onde RF

Si on supprime le champ B_1 , le vecteur M a tendance à revenir à sa position initiale c'est-à-dire colinéaire à l'axe Oz . Sa composante M_y se déphase et tend alors vers 0, de même pour M_x , tandis que M_z croit. Cette décroissance de M_y se fait de façon exponentielle et engendre un courant induit dans une bobine située sur l'axe Oy . Bloch a montré que l'évolution de ce mouvement est liée au temps de relaxation longitudinal $T1$ et au temps de relaxation transversal $T2$. Les valeurs de ces derniers dépendent des tissus biologiques rencontrés. (Tableau 1.2) :

	$T1$	$T2$
<i>Liquide Céphalo-rachidien</i>	2500 ms	2000 ms
<i>Matière grise</i>	900 ms	90 ms
<i>Matière blanche</i>	750 ms	80 ms
<i>Graisse</i>	300 ms	40 ms

Tab 1.2- Ordre de grandeur des temps de relaxation à 1.5 Tesla

Les images résultantes de la mesure de l'aimantation M en tout point d'un volume dépendent principalement de trois **séquences** intrinsèques :

- ▶ $T1$: le temps de relaxation longitudinal
- ▶ $T2$: le temps de relaxation transversal
- ▶ P : la densité de protons qui définit la quantité de signal considéré.

La relaxation :

On parle de relaxation quand, après absorption de l'énergie électromagnétique fournie par B_1 , les noyaux tendent à retrouver la distribution de Boltzmann (c'est à dire quand \vec{M} revient à sa position d'équilibre). On peut la décomposer en deux phénomènes, la relaxation longitudinale et la relaxation transversale.

La relaxation longitudinale :

A l'équilibre $M_z = M_0$, Après le basculement $M_z = 0$

Le retour de M_z à sa valeur de départ M_0 est exponentielle :

$$M_z(t) = M_0 (1 - e^{-t/T_1}) \quad (1.2)$$

Cette relaxation longitudinale, dite relaxation T1 ou encore relaxation "spin-réseau", correspond au retour à l'équilibre énergétique du système après l'excitation. La constante de temps T1 est le temps nécessaire pour que les protons atteignent les deux tiers de leur aimantation. Elle dépend en fait de la mobilité des atomes d'hydrogène ou de celle des molécules où ils sont engagés. T1 sera d'autant plus court que ces hydrogènes seront liés à de grosses molécules.

La relaxation transversale :

A l'équilibre $M_{xy} = 0$

Après le basculement de 90 degrés, $M_{xy} = M_0$. Le retour de M_{xy} vers 0 est exponentiel

$$M_{xy}(t) = M_0 e^{-t/T_2} \quad (1.3)$$

Cette décroissance de la composante transversale se fait en général plus vite que ne le veut le simple retour à l'équilibre de la composante longitudinale. Elle se caractérise par le temps de relaxation T2 (encore appelé temps de relaxation "spin-spin"). T2 est en réalité le temps pendant lequel l'intensité décroît de deux tiers de sa valeur initiale. Ce temps de relaxation T2 est toujours inférieur au temps de relaxation T1. Il dépend lui aussi de la mobilité des atomes ou des molécules sur lesquelles ces protons sont engagés. Ces temps de relaxation vont varier pour un tissu donné selon l'organisation physico-chimique de l'eau dans ce tissu, et c'est sur ces variations que l'on s'appuie pour détecter au sein d'un tissu les modifications liées à la présence d'une lésion.

Paramètres de séquence

Les paramètres de séquence sont les paramètres que le manipulateur fixe sur la console pour définir la séquence IRM. Ils permettent de contrôler l'influence des différents paramètres tissulaires **T1**, **T2** et densité de proton dans le signal et de moduler ainsi le contraste dans l'image.

➤ **Temps d'écho**

Le signal de précession libre ne peut être enregistré directement après l'excitation (Déphasage parasite induit par les gradients). C'est pourquoi il est acquis sous la forme d'un Écho de spin ou de gradient. Par définition, le délai entre le milieu de l'impulsion d'excitation et le sommet de l'écho est appelé temps d'écho, il est noté TE .

➤ **Temps de répétition**

L'image est constituée à partir de la répétition de la même séquence avec un gradient de Phase G_p d'amplitude différente. Le temps qui sépare deux répétitions est appelé temps de Répétition est noté TR . Le TR , comme le TE , est un facteur de contraste. S'il est suffisamment Long, l'aimantation repousse tout le signal qui ne dépend pas de la vitesse d'aimantation (donc de $T1$), mais essentiellement de la densité protonique. S'il est court. le système atteint après quelques répétitions un régime stationnaire et l'aimantation tend vers une valeur d'équilibre dépendant de la vitesse d'aimantation des tissus, et donc de leur $T1$. L'image révèle ainsi les différences de $T1$ entre les tissus.

Observation des tissus dans des images pondérées en $T1$ et $T2$:

Dans les images pondérées en $T1$, on observe du plus foncé au plus clair :

- l'air, les yeux, l'os et le liquide céphalo-rachidien
- les muscles, la peau et la substance grise,
- la substance blanche,
- la graisse et le sang

En $T2$, toujours du plus foncé au plus clair, on observe :

- l'air, l'os et le sang,
- les muscles et la peau,
- la substance blanche,
- la substance grise,
- la graisse, les yeux et le liquide céphalo-rachidien

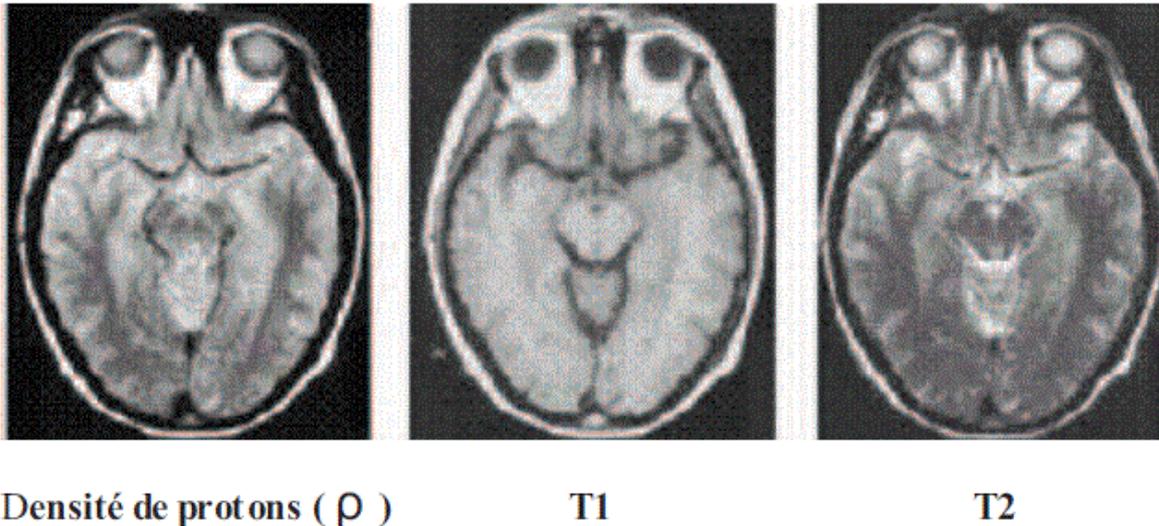


Figure 1.14 : Les différentes séquences d'une image IRM.

*Défauts des images acquises par l'IRM

Outre les trois paramètres qui entrent en jeu dans la formation d'une image : la densité Protonique ρ , le temps de relaxation $T1$ et le temps de relaxation $T2$, un certain nombre de Facteurs que l'on ne maîtrise pas toujours affectent la qualité des images et provoquent des Artefacts, compliquant la segmentation des tissus cérébraux ou la reconnaissance de structures cérébrales. De plus amples informations sont données dans [6] et [7]. Les artefacts d'acquisition en IRM sont de natures différentes de celles observées dans D'autres domaines du traitement d'images. On distingue essentiellement quatre effets: le *bruit*, le *mouvement*, les *inhomogénéités d'intensité* et les *effets de volume partiel*

○ **Bruit**

Comme tout dispositif de mesure physique, les données sont entachées de bruit. Il provient à La fois du patient (agitation thermique des protons à l'origine d'émissions parasites) et de la Chaîne de mesure (convertisseurs analogique- numérique, antenne, ...). La perturbation par le Bruit est généralement quantifiée par le rapport signal sur bruit (RSB), fonction de l'amplitude du signal observé par rapport à l'importance de la variation du bruit. Ce rapport, et donc la qualité de l'image, peut être amélioré en considérant différentes antennes, un champ magnétique plus intense, une matrice de résolution moins fine, ou encore une multiplication des mesures. Néanmoins, il restera toujours un bruit dans l'image reconstruite finale.

○ **Mouvement**

Tout mouvement de protons lors de la formation de l'image a des conséquences sur la Qualité de l'image acquise. De même, comme en photographie, les mouvements du sujet Nuisent à la qualité de l'image. Dans les vaisseaux, les protons possèdent un mouvement plus ou moins rapide qui va Perturber les moments magnétiques longitudinaux et transversaux des régions analysées et Donc modifier des signaux de résonance magnétique.

○ **L'inhomogénéité RF (Radiofréquence)**

Des variations spatiales lentes (inhomogénéité) de l'intensité des images peuvent Apparaître dans des images IRM (Voir Figure 1.15). Elles sont provoquées par les Inhomogénéités dans le champ magnétique principal, des inhomogénéités dans le champ Généré par les impulsions RF et dans le gain de l'antenne RF. Cet artéfact n'est pas très Gênant pour le clinicien, sauf dans des cas extrêmes, où le contraste entre tissus disparaît. En Revanche, il peut poser d'importants problèmes à un système de traitement d'images automatique.

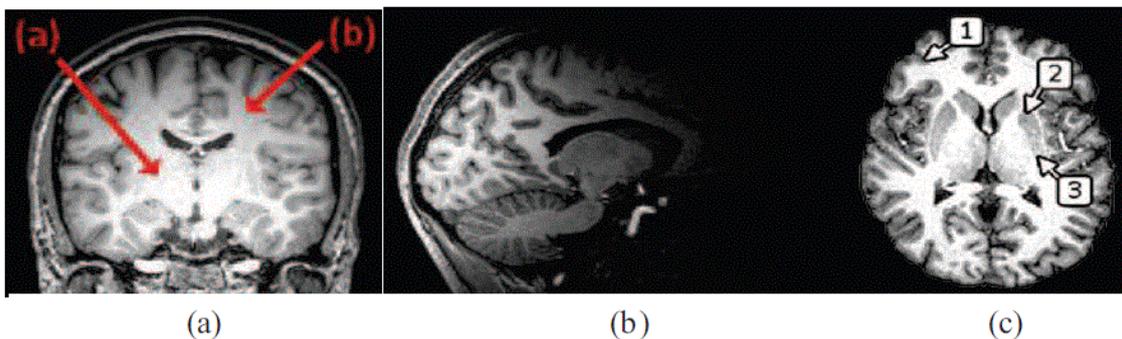


Figure. 1.15 – Artéfact d'inhomogénéité d'intensité : l'image (a) montre des variations d'intensité de la matière blanche dans le volume. L'image (b) est une acquisition réalisée avec une antenne de surface, conduisant à de très fortes inhomogénéités. L'image (c) illustre les inhomogénéités dues aux propriétés biologiques des tissus : les marques (1), (2) et (3) pointent toutes de la matière grise, mais qui présente une intensité variable à cause de compositions histologiques différentes selon les régions.

○ **L'effet de volume partiel**

L'effet de volume partiel est lié à la discrétisation de l'espace : lorsque la surface entre Plusieurs objets se trouve dans un même élément de volume discret (le voxel), la mesure dans Ce voxel résulte d'un mélange des contributions des différents objets (voir Figure 1.16). L'effet de volume partiel se produit dans les cas suivants :

- a) Le voxel imagé chevauche la frontière entre deux ou plusieurs tissus différents. Ce cas est prépondérant en IRM.
- b) La résolution spatiale dépasse la taille des structures présentes dans l'image.

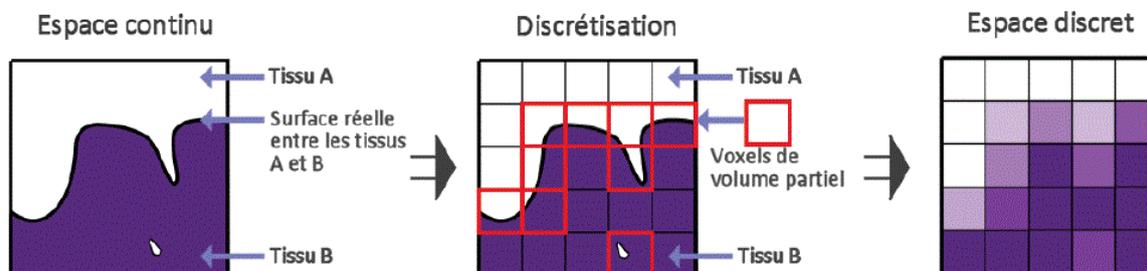


Figure 1.16 : Artéfact de volume partiel dû à la discrétisation de l'espace

Cet effet se manifeste principalement à l'interface entre les matières (MG, MB, LCR, graisse, os) ou lors de la présence de structures trop fines pour être visibles à la résolution de l'image : vaisseaux sanguins (diamètre de 40micromètres à quelques millimètres), structures grises fines, ... L'effet de volume partiel (Partial Volume Effect, ou PVE en anglais) est Particulièrement présent à l'interface LCR-MG dans les replis du cortex, car l'épaisseur des sillons corticaux est généralement inférieure à la résolution des images. [3]

1.6.3 Avantages et inconvénients de l'IRM :

- **Avantage :**
 - Pour commencer, l'IRM a un faible coût en électricité par rapport aux autres technologies de la médecine : en effet lorsque l'appareil se met en marche il faut l'alimenter, mais ensuite il n'y a quasiment aucune perte d'énergie. Il existe plusieurs types d'IRM (en fonction de leur puissance en Tesla) qui permettent d'observer différentes parties du corps : on n'est pas obligé de s'équiper d'un énorme imageur pour effectuer une IRM du genou par exemple.
 - Elle permet l'acquisition d'images bi ou tridimensionnelles
 - Les images issues d'un examen d'IRM sont de très hautes résolutions : il sera donc possible d'observer certains détails invisibles lors des scanners ou des radiographies, puisqu'elle a l'avantage de présenter différents contrastes (T1, T2, Rho (densité de proton), Flair)

- Les IRM sont inoffensives et on pourrait passer un examen d'IRM tous les jours.
- De plus c'est un examen non invasif.
- Enfin, cette technique permet en outre la réalisation d'étude du cerveau, aussi bien anatomique que fonctionnelle grâce aux techniques d'IRM et IRMf (IRM fonctionnelles) :

-Technique d'imagerie par résonance magnétique nucléaire (IRM) permettant de cartographier les activités *fonctionnelles* du cerveau. Le principe consiste à mesurer l'oxygénation (rapport oxyhémoglobine/désoxyhémoglobine), qui augmente localement dans les aires activées suite à un apport accru en sang frais.

- **Inconvénients :**

- Les coûts des consultations pour les patients sont assez élevés : entre 70 € et 250 € pour effectuer une IRM.
- Lors d'une IRM le sujet est très à l'étroit et subit d'importants bruits.
- Les contre-indications limitent les utilisations de l'IRM sur certains patients.

L'inconvénient majeur réside dans le fait que le patient doit respecter une stricte immobilité pendant toute la durée de l'examen (environ une quinzaine de minutes) et que par ailleurs le fait d'être enfermé dans un tunnel est parfois difficile à supporter pour certaines personnes (en particulier les claustrophobes). De plus du fait du champ magnétique puissant pouvant déplacer les objets dits ferromagnétiques, il existe certaines contre-indications : les valves cardiaques artificielles métalliques, les stimulateurs cardiaques, certains clips cérébraux ou certaines prothèses ou objets métalliques. Il va de soi que tous les objets métalliques ou magnétiques doivent être déposés à l'extérieur de la salle d'examen (clefs, pièces de monnaie, carte bancaire...etc). La salle d'examen est par ailleurs protégée (blindée) par une protection spécifique (cage de Faraday) qui la met à l'abri des perturbations ou pollutions magnétiques extérieurs (émission d'ondes CB... etc.).

L'IRM est une technique de visualisation du cerveau par excellence, vu la grande information qu'elle fournit sur le corps humain particulièrement sur le cerveau tout en assurant le bien être du patient puisqu'elle ne présente pas de danger contre sa santé, mais la question qui se pose est comment peut-on distinguer les différentes régions du cerveau sur cette image ? Pour un expert et vu son expérience et l'importante quantité de connaissances qu'il a acquis, il peut répondre à cette question, mais il a du faible parfois à limiter les différentes régions présentes avec exactitude surtout lors de la présence d'une pathologie.

et cela peut produire des séquelles indésirables pour le patient surtout après un geste chirurgical. Pour éviter d'éventuels effets insouhaitables une solution est la segmentation. Suite à cela la littérature a connu un nombre énorme de méthodes de segmentation, nous allons donc fournir un état de l'art sur ces méthodes dans les chapitres suivant.

Ce chapitre présente l'utilisation de méthodes statistiques pour dans la segmentation algorithmique EM (Expectation Maximisation), c'est une méthode paramétrique non supervisée. L'algorithme EM étudié dans ce chapitre est un classifieur, il affecte une étiquette à chaque pixel. En un mot le principe de ce classifieur consiste à trouver les paramètres d'un modèle statistique donné correspondant le mieux possible au résultat de l'expérience. L'analogie avec la physique statistique réside dans la manière d'appréhender l'espace des solutions.

La partie 2.1 rappelle la théorie bayésienne et les différents principaux d'estimateur. La partie 2.2 présente une méthode classique pour l'estimation des paramètres (Estimation par maximum de vraisemblance) la partie 2.3 détaille notre approche pour estimer et calculer les paramètres (algorithme EM) qui permet de compléter une série de données manquantes en se basant sur la vraisemblance maximale et dans la dernière partie 2.4 on donne un exemple de l'application d'un algorithme EM sur un mélange gaussien.

2.1- Théorie bayésienne

Dans le cadre de la segmentation statistique, une image est représentée par un couple de variables aléatoires $(Y, X) = ((Y_i)_{i \in N}, (X_i)_{i \in N})$ où N est l'ensemble des pixels (ou voxels dans le cas de données tridimensionnelles), Y le champ des observations et X celui des étiquettes. Chaque X_i prend ses valeurs dans un ensemble $\Omega = \{\omega_1, \dots, \omega_k\}$ appelé ensemble des classes et chaque Y_i prend ses valeurs dans \mathbb{R}^m , où m est par exemple le **nombre de** modalités dans le cas de la segmentation d'IRM (ou le nombre de canaux d'observation de manière plus générale). Les observations données par le champ Y sont considérées comme une version dégradée (filtrée, bruitée,...) de X , le problème de la segmentation d'images consiste alors à estimer la réalisation de X ayant observé $(Y_i)_{i \in N}$.

On suppose dans ce cas l'unicité de X associée à des observations à caractère multimodal. La recherche des informations cachées X à partir des observations Y est généralement basée sur l'établissement d'un modèle direct qui intègre toute la connaissance disponible sur la formation des Y à partir des X . Retrouver les informations cachées à partir des observations consiste alors à inverser ce modèle direct. Les problèmes inverses en traitement d'images étant généralement des problèmes mal-posés,

cette tâche s'avère complexe. Pour tenter de contourner cette difficulté, on peut favoriser certains types de solution en imposant une connaissance a priori sur les étiquettes et leur agencement spatial.

Dans la suite du manuscrit la notation P désignera une loi de probabilité dans le cas où l'ensemble des configurations est discret. Dans le cas où cet ensemble est continu, la même notation P doit être comprise comme une densité de probabilité.

La théorie bayésienne est basée sur la spécification de la distribution a posteriori $P(X=x|Y=y)$. La règle de Bayes permet de déterminer la probabilité a posteriori de étiquettes connaissant le champ des observations.

$$p(X=x|Y=y) = \frac{p(Y=y|X=x)P(X=x)}{P(Y=y)} \quad (2.1)$$

où $P(Y=y|X=x)$ est la vraisemblance des observations conditionnellement aux informations cachées, $P(X=x)$ est l'a priori sur les étiquettes et $p(Y=y)$ une constante de normalisation. Nous allons maintenant présenter différents estimateurs permettant d'associer à chaque estimation \tilde{x} de x une fonction de coût $L(x, \tilde{x})$ mesurant la "distance" entre le champ des étiquettes et son estimation. L'estimateur optimal au sens de Bayes est celui qui minimise le coût moyen, le risque de Bayes :

$$\tilde{x}_{opt}(y) = \underset{\tilde{x}}{\operatorname{argmin}} E[L(X, \tilde{x})|Y=y]$$

Les principaux estimateurs bayésiens sont présentés dans les trois paragraphes suivants.

2.1.1-L'estimateur du MAP :

L'estimateur du Maximum a Posteriori (MAP) pénalise de la même façon toutes les estimations différentes de la vraie solution x . La fonction de coût est la suivante :

$$L(x, \tilde{x}) = 1 - \delta(x, \tilde{x}) \text{ avec } \delta(a, b) = \begin{cases} 1 & \text{si } a = b \\ 0 & \text{sinon} \end{cases}$$

le symbole de Kronecker. L'estimateur basé sur cette fonction de coût est :

$$\tilde{x}_{MAP}(y) = \underset{x}{\operatorname{argmax}} p_x|Y(y)$$

2.1.2-L'estimateur du MPM :

L'estimateur du Mode des Marginales a Posteriori (MPM) pénalise une solution Proportionnellement au nombre de sites erronés. La fonction de coût qui lui est associée est la suivante :

$$L(x, \tilde{x}) = \sum_{s \in S} \mathbf{1} - \delta(x_s, \tilde{x}_s)$$

L'estimateur du MPM minimise en chaque site la distribution marginale locale a posteriori :

$$\forall s \in S \quad \tilde{x}_{sMPM}(y) = \underset{x_s}{\operatorname{argmax}} p_{x_s | Y^{(x_s \setminus y)}}$$

2.1.3-L'estimateur du champ moyen :

La fonction de coût de l'estimateur du champ moyen (CM) est quadratique

$$L(x, \tilde{x}) = \sum_{s \in S} (x_s - \tilde{x}_s)^2$$

L'estimateur basé sur cette fonction de coût est :

$$\forall s \in S \quad \tilde{x}_{sCM}(y) = \mathbb{E}[X_s | Y = y]$$

On peut noter que dans le cas où les étiquettes X prennent leurs valeurs dans l'ensemble discret.

L'estimateur du champ moyen ne sera pas nécessairement dans cet ensemble.

2.2-Estimation par maximum de vraisemblance

a) Principe

La méthode du maximum de vraisemblance (MV) est celle qui est le plus souvent utilisée. Nous en présentons ici le principe général. Dans sa forme paramétrique, on définit une fonction de densité $f(Y/\theta)$ qui dépend de l'ensemble des paramètres θ , Par exemple, cette fonction peut être une gaussienne où θ est composé de la moyenne et de la variance (μ, σ^2) . On suppose que notre ensemble de N données est extrait de cette distribution

c'est-à-dire les observations $Y = \{y_1, \dots, y_N\}$. Sont des réalisations indépendantes et identiquement distribuées d'une variable aléatoire La densité de l'échantillon Y peut donc s'écrire :

$$f(y \setminus \theta) = \prod_{i=1}^N f(y_i \setminus \theta) \quad (2.2)$$

Cette fonction exprime la vraisemblance des paramètres selon les données. Soit X la variable cachée, Y sont les données observées. On suppose que les $y_i, i \in \{1, \dots, N\}$ sont des données i.i.d, l'équation (2.2) devient :

$$f(Y \setminus \theta) = \prod_{i=1}^N f(y_i \setminus \theta) = \prod_{i=1}^N \sum_x f(y_i, x \setminus \theta) \quad (2.3)$$

C'est la fonction de vraisemblance. Il est plus facile de calculer le logarithme de cette fonction, on utilise donc la log-vraisemblance (log-likelihood) :

$$\log f(Y \setminus \theta) = \sum_{i=1}^N \log f(y_i \setminus \theta) \quad (2.4)$$

Et qu'il est généralement bien plus simple de travailler sur une somme plutôt que sur un produit .Et avec les variables cachées x donc la log-vraisemblance complet devient :

$$\text{Log } f(Y \setminus \theta) = \sum_i \log \sum_x f(y_i, x \setminus \theta) \quad (2.5)$$

b) L'estimateur au sens maximum de vraisemblance (EMV) de θ :

$$\tilde{\theta} = \text{argmax}_{\theta} f(y \setminus \theta) \quad (2.6)$$

C'est donc, à échantillon fixé, la valeur du vecteur de paramètres θ qui rend aussi vraisemblables que possible les observations obtenues. Si l'on considère que $f(Y/\theta)$ a une forme simple, une distribution normale par exemple, on peut trouver le maximum en utilisant la première dérivée de $f(Y/\theta)$ par rapport à $\theta = (\mu, \sigma^2)$. Dans ce cas, la dérivée donne simplement la moyenne et la variance de l'échantillon. Pour des cas plus compliqués comme des échantillons avec données manquantes, il n'y a pas de solution analytique et il faut utiliser une méthode itérative comme l'algorithme EM.

2.3- L'algorithme EM

L'algorithme EM (pour Expectation-Maximisation) est un algorithme itératif du à Dempster Laird et Rubin (1977). Il s'agit d'une méthode d'estimation paramétrique s'inscrivant dans le cadre général du maximum de vraisemblance.[8]

Lorsque les seules données dont on dispose ne permettent pas l'estimation des paramètres. et/ou que l'expression de la vraisemblance est analytiquement impossible à maximiser l'algorithme EM peut être une solution. De manière grossière et vague, il vise à fournir un estimateur lorsque cette impossibilité provient de la présence de données cachées ou manquantes, lorsque la connaissance de ces données rendrait possible l'estimation des paramètres. L'algorithme EM tire son nom du fait qu'à chaque itération il opère deux étapes distinctes :

- A) la phase « Expectation », souvent désignée comme « l'étape E », procède comme son nom le laisse supposer à l'estimation des données inconnues, sachant les données observées et la valeur des paramètres déterminée à l'itération précédente.
- B) la phase « Maximisation », ou « étape M », procède donc à la maximisation de la Vraisemblance, rendue désormais possible en utilisant l'estimation des données inconnues effectuée à l'étape précédente, et met à jour la valeur du ou des paramètre(s) pour la prochaine itération.

En bref, l'algorithme EM procède selon un mécanisme extrêmement naturel : s'il existe un obstacle pour appliquer la méthode MV, on fait simplement sauter cet obstacle puis on applique effectivement cette méthode. Le côté itératif de l'algorithme pourra peut-être paraître un peu mystérieux pour l'instant, mais comme nous le verrons, l'algorithme garantit que la vraisemblance augmente à chaque itération. Ce qui conduit donc à des estimateurs de plus en plus corrects.

2.3.1- Rappels sur Inégalité de Jensen[10]

* **Fonction convexe.**

Soit Une application $f:[a, b] \rightarrow \mathbb{R}$ dite convexe sur $[a, b]$ si pour tous y_1, y_2 de cet intervalle et tout $\lambda \in [0, 1]$

$$f(\lambda y_1 + (1-\lambda)y_2) \leq \lambda f(y_1) + (1-\lambda)f(y_2) \quad (2.7)$$

f est dite strictement convexe si l'inégalité (2.7) est stricte, et f est dite concave si $-f$ est convexe.

***Théorème Inégalité de Jensen [9]**

Soit f une fonction convexe définie sur un intervalle I Si $y_1, \dots, y_n \in I$ et $\lambda_1, \dots, \lambda_n \geq 0$ tels que $\sum \lambda_i = 1$ alors :

$$f(\sum_{i=1}^n \lambda_i \cdot y_i) \leq \sum_{i=1}^n \lambda_i \cdot f(y_i) \tag{2.8}$$

2.3.2-Itérations EM

On suppose des données complètes X partiellement observées sous la forme des données incomplètes y . On note z la donnée additionnelle ou variable cachée, $x = [y; z]$. On pose $p(x|\theta)$ la fonction de vraisemblance de x , où θ est un vecteur de paramètres prenant ses valeurs dans $\Theta \in \mathbb{R}^r$. à partir de la formule de Bayes :

$$p(z|y, \theta) = \frac{p(z, y, \theta)}{p(y, \theta)} = \frac{p(x|\theta)}{p(y|\theta)} \tag{2.9}$$

Par simple substitution dans l'expression ci-dessus, on obtient

$$P(y|\theta) = \frac{p(x|\theta)}{p(z|y, \theta)}$$

En passant au logarithme des deux côtés de l'équation, on a alors

$$\log p(y|\theta) = \log p(x|\theta) - \log p(z|y, \theta) \tag{2.10}$$

On a besoin de fournir une première estimation a priori pour les paramètres, que l'on notera θ' , et on rappelle que y est connu et fixé. La distribution de z (inconnu) étant donné les observations disponibles y est $p(z | y, \theta')$. On moyenne (2.10) sur la distribution des données inconnues, ou en d'autres termes, on calcule l'espérance des deux membres de l'équation (2.10) par rapport à $p(z | y, \theta')$. Puisque $\mathbb{E} (h(X) / X) = h(X)$ quelle que soit $h(X)$, on peut écrire :

$$\log p(\mathbf{y} \setminus \theta) = \mathbb{E}(\log p(\mathbf{x} \setminus \theta) / \mathbf{y}, \theta') - \mathbb{E}(\log p(\mathbf{z} \setminus \mathbf{y}, \theta) / \mathbf{y}, \theta') \quad (2.11)$$

Si l'on introduit la notation suivante :

$$Q(\theta, \theta') = \mathbb{E}(\log p(\mathbf{x} \setminus \theta) / \mathbf{y}, \theta') \quad (2.12)$$

Cette fonctionnelle est appelée **Q**-fonction. Soit $\theta^{(0)}$, la valeur initiale de θ , à la $s^{\text{ème}}$ itération de l'algorithme EM, les étapes E et M peuvent alors s'exprimer formellement de la façon suivante :

L'étape E : calcul de la **Q**-fonction $Q(\theta, \theta^{(s)})$

$$Q(\theta, \theta^{(s)}) = \mathbb{E}(\log p(\mathbf{x} \setminus \theta) / \mathbf{y}, \theta^{(s)}) \quad (2.13)$$

L'étape M : choisir $\theta^{(s+1)}$ qui maximise $Q(\theta, \theta^{(s)})$

$$\theta^{(s+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}} Q(\theta, \theta^{(s)}) \quad (2.14)$$

Où $\theta^{(s)}$ est la valeur du paramètre à la $s^{\text{ème}}$ itération. On itère les étapes E et M jusqu'à ce que le critère de convergence soit atteint. Généralement ce critère est un critère sur la variation de la log-vraisemblance.

2.3.3-Preuve de la croissance de la vraisemblance d'une itération à l'autre

À l'itération (s) , nous disposons d'une valeur $\theta^{(s)} \in \mathbf{R}^m$ du vecteur de paramètres. Le but est de la mettre à jour avec une « meilleure » valeur θ , augmentant la vraisemblance. Donc telle que $\mathbb{E}(\log p(\mathbf{y} \setminus \theta) - \log p(\mathbf{y} \setminus \theta^{(s)})) \geq 0$. On souhaite bien sûr que cette différence soit la plus grande possible. on ne sait pas maximiser $p(\mathbf{y} \setminus \theta)$ donc on ne sait pas non plus maximiser $\mathbb{E}(\log p(\mathbf{y} \setminus \theta))$ Un moyen d'optimiser malgré tout, dans une certaine mesure. cette différence, peut consister à chercher une fonction $\theta \rightarrow Q(\theta \setminus \theta^{(s)})$ que l'on sait maximiser, et qui est telle que :

$$\begin{cases} \mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)}) \geq \mathbf{Q}(\boldsymbol{\theta} \setminus \boldsymbol{\theta}^{(s)}) \forall \boldsymbol{\theta} \in \mathbf{R}^m \\ \mathbf{Q}(\boldsymbol{\theta}^{(s)} \setminus \boldsymbol{\theta}^{(s)}) = \mathbf{0} \end{cases} \quad (2.15)$$

Ainsi, $\mathbf{Q}(\boldsymbol{\theta} \setminus \boldsymbol{\theta}^{(s)})$ borne inférieurement $\mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)})$ et son maximum est au moins égal à 0. Trouver un $\boldsymbol{\theta}'$ qui maximise $\boldsymbol{\theta} \rightarrow \mathbf{Q}(\boldsymbol{\theta} \setminus \boldsymbol{\theta}^{(s)})$ conduit donc mécaniquement à obtenir $\mathbf{f}(\boldsymbol{\theta}', \boldsymbol{\theta}^{(s)}) \geq \mathbf{0}$ C'est à dire une nouvelle valeur $\boldsymbol{\theta}'$ plus vraisemblable de paramètres. Afin de trouver une telle fonction \mathbf{Q} , nous utilisons une représentation marginale de la vraisemblance selon les données cachées $\mathbf{z}=(z_1, \dots, z_n)$:

$$p(\mathbf{y} \setminus \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{y}, \mathbf{z} \setminus \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{y} \setminus \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} \setminus \boldsymbol{\theta}) \quad (2.16)$$

Il vient alors :

$$\begin{aligned} \mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)}) &= \log p(\mathbf{y} \setminus \boldsymbol{\theta}) - \log p(\mathbf{y} \setminus \boldsymbol{\theta}^{(s)}) \\ &= \log(\sum_{\mathbf{z}} p(\mathbf{y} \setminus \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} \setminus \boldsymbol{\theta})) - \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log p(\mathbf{y} \setminus \boldsymbol{\theta}^{(s)}) \end{aligned} \quad (2.17)$$

Cette expression utilise le logarithme d'une somme : en se souvenant de l'inégalité de Jensen (2.8). On commence à voir apparaître clairement une façon de minorer $\mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)})$ Nous réécrivons (2.17) en introduisant dans la somme de gauche les $p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)})$ Présents dans la somme de droite.

$$\mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)}) = \log \left(\sum_{\mathbf{z}} \frac{p(\mathbf{y} \setminus \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} \setminus \boldsymbol{\theta})}{p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)})} \right) p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) - \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log p(\mathbf{y} \setminus \boldsymbol{\theta}^{(s)}) \quad (2.18)$$

Et enfin, en remarquant que $\sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) = 1$ nous appliquons l'inégalité de Jensen :

$$\begin{aligned} \mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)}) &\geq \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log \left(\frac{p(\mathbf{y} \setminus \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} \setminus \boldsymbol{\theta})}{p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)})} \right) - \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log p(\mathbf{y} \setminus \boldsymbol{\theta}^{(s)}) \\ &= \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log \left(\frac{p(\mathbf{y} \setminus \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} \setminus \boldsymbol{\theta})}{p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)})} \right) - \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log p(\mathbf{y} \setminus \boldsymbol{\theta}^{(s)}) \\ &= \sum_{\mathbf{z}} p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) \log \left(\frac{p(\mathbf{y} \setminus \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} \setminus \boldsymbol{\theta})}{p(\mathbf{z} \setminus \mathbf{y}, \boldsymbol{\theta}^{(s)}) p(\mathbf{y} \setminus \boldsymbol{\theta}^{(s)})} \right) \end{aligned}$$

$$= \sum_z p(z|y, \theta^{(s)}) \log \left(\frac{p(x|\theta)}{p(x|\theta^{(s)})} \right) \quad (2.19)$$

$$= Q(\theta|\theta^{(s)})$$

Nous avons donc obtenu une fonction $\theta \rightarrow Q(\theta|\theta^{(s)})$ vérifiant les conditions (2.15), il est évident avec (2.19) que $Q(\theta^{(s)}|\theta^{(s)}) = 0$. Finalement, nous posons :

$$\begin{aligned} \theta^{(s+1)} &= \operatorname{argmax}_{\theta} Q(\theta|\theta^{(s)}) \\ &= \operatorname{argmax}_{\theta} \left\{ \sum_z p(z|y, \theta^{(s)}) \log \left(\frac{p(x|\theta)}{p(x|\theta^{(s)})} \right) \right\} \\ &= \operatorname{argmax}_{\theta} \left\{ \sum_z p(z|y, \theta^{(s)}) \log p(x|\theta) \right\} \\ &= \operatorname{argmax}_{\theta} E(\log p(x|\theta) | y, \theta^{(s)}) \end{aligned}$$

On détermine bien ainsi une valeur $\theta^{(s+1)}$ plus vraisemblable que $\theta^{(s)}$ puisque :

$$\log p(x|\theta^{(s+1)}) - \log p(x|\theta^{(s)}) = \mathbf{f}(\theta^{(s+1)}, \theta^{(s)}) \geq Q(\theta^{(s+1)}, \theta^{(s)}) \geq Q(\theta^{(s)}|\theta^{(s)}) \geq 0$$

Chaque itération fait augmenter la log-vraisemblance, donc l'algorithme converge directement vers un maximum local de la fonction de log-vraisemblance. La figure 2.1 schématise le fonctionnement de l'algorithme EM

*Organigramme

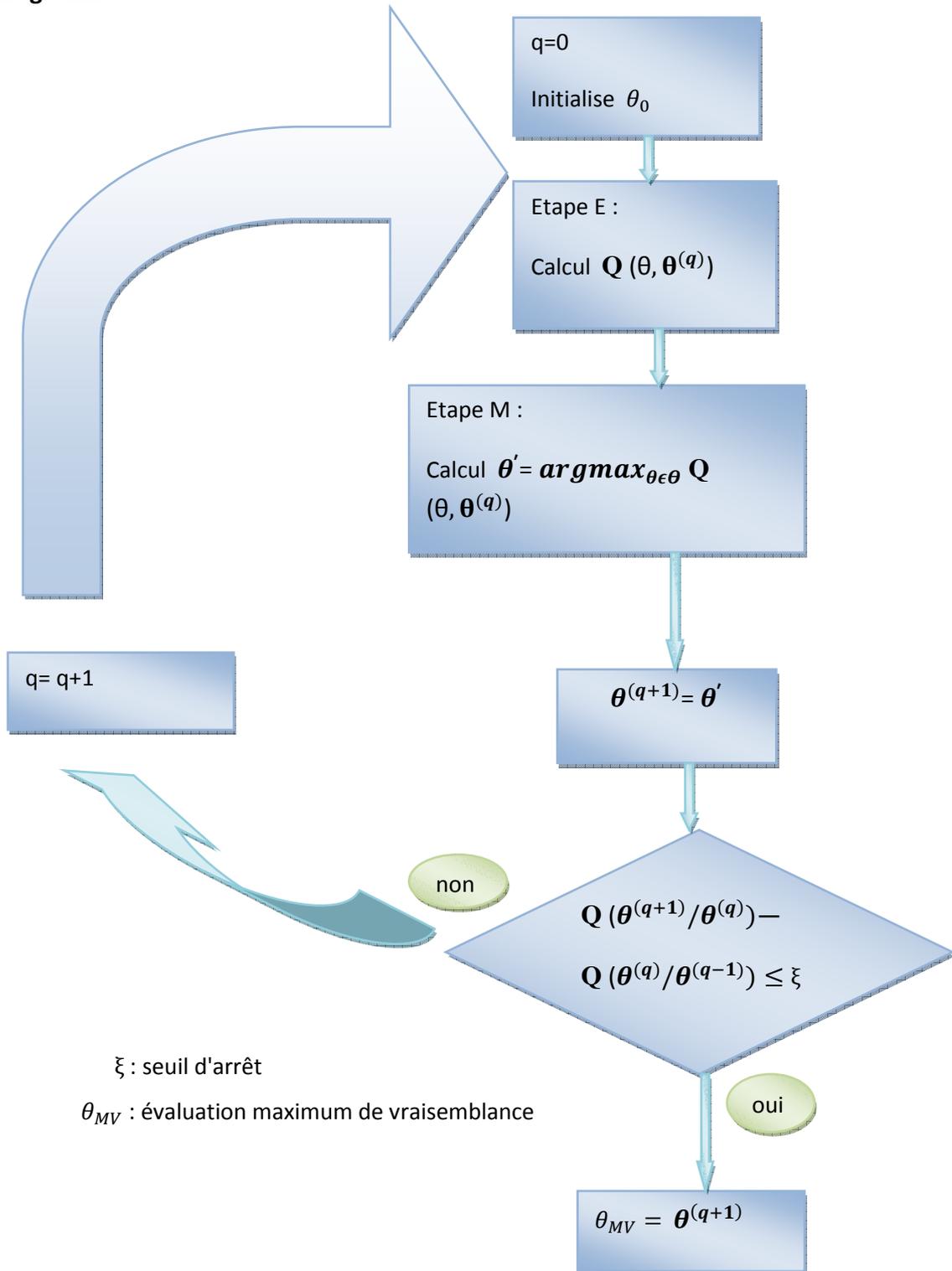


Figure 2.1 : le fonctionnement de l'algorithme EM

2.4-Modèle de mélange :

Dans un modèle de mélange la fonction de densité est définie comme une somme pondérée de k fonctions paramétriques ayant pour ensemble de paramètres θ . Une famille de mélange fini de densité est donc de la forme suivante :

$$F(y|\theta) = \sum_{j=1}^k \pi_j f_j(y|\theta_j) \tag{2.20}$$

Avec $y \in R^m$ et Les π_j sont les proportions de mélange et doivent respecter les propriétés suivantes:

$\pi_j \geq 0, \sum_{j=1}^k \pi_j = 1. f_j(y|\theta_j)$ est une distribution paramétrée par θ_j et θ

$= (\pi_1, \dots, \pi_k, \theta_1, \dots, \theta_k)$. Soit $y = (y_1, y_2, \dots, y_n)$ un échantillon de n observations indépendantes et identiquement distribuées (i.i.d.) et tel que chacun des y_i suit la distribution $f(y_i, \theta)$. Sous cette hypothèse. la log-vraisemblance $\log f(y|\theta)$ s'écrit :

$$\log f(y|\theta) = \log \prod_{i=1}^n f(y_i|\theta) = \sum_{i=1}^n \log \sum_{j=1}^k \pi_j f_j(y_i|\theta_j) \tag{2.21}$$

Dans le cadre d'une application à la classification, on considère la donnée incomplète z comme la classe d'appartenance de chacun des éléments à classifier. On écrit pour l'élément y_i :

$$z_{ij} = \begin{cases} 1 & \text{si } y_i \text{ appartient à la classe } c_j \\ 0 & \text{sinon} \end{cases}$$

Généralement, les $z_i = (z_{i1}, \dots, z_{ik})$ sont pris comme des réalisations indépendantes suivant une distribution multinomiale. En d'autres termes, dans le cas d'un modèle de mélange, la variable cachée z devient alors l'appartenance de chaque échantillon à l'une des k classes. Les données complètes sont donc $x = (x_1, \dots, x_n)$ avec $x_i = (y_i, z_{ij})$ et la vraisemblance associée s'écrit :

$$P(y_i, z_i) = P(y_i|z_i)P(z_i) = (f_1(y_i|\theta_1))^{z_{i1}} \pi_1^{z_{i1}} \dots \dots (f_k(y_i|\theta_k))^{z_{ik}} \pi_k^{z_{ik}} \tag{2.22}$$

La log-vraisemblance complète associée est donc :

$$\log p(\mathbf{x} \setminus \boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^k z_{ij} \log \pi_j f_j(y_i \setminus \boldsymbol{\theta}_j) \quad (2.23)$$

L'algorithme EM répète deux phases

Etape E

On calcule l'espérance conditionnelle de la log-vraisemblance complète

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(s)}) = E(\log p(\mathbf{x} \setminus \boldsymbol{\theta}) / y, \boldsymbol{\theta}^{(s)}) \quad (2.24)$$

Comme la log-vraisemblance complète (2.23) est une fonction linéaire des z_{ij} , ce calcul se limite à remplacer les z_{ij} par leur espérance conditionnelle, elle s'appelle probabilités a posteriori :

$$E[z_{ij} \setminus y_i, \boldsymbol{\theta}^{(s)}] = \frac{\pi_j^{(s)} f_j(y_i \setminus \boldsymbol{\theta}_j^{(s)})}{\sum_{j=1}^k \pi_j^{(s)} f_j(y_i \setminus \boldsymbol{\theta}_j^{(s)})} \quad (2.25)$$

Etape M

On recherche la valeur $\boldsymbol{\theta}$ qui maximise (2.24). Cette étape dépend complètement du modèle de mélange choisi. Dans le paragraphe suivant, nous allons expliciter le calcul dans le cas d'un mélange de gaussiennes.

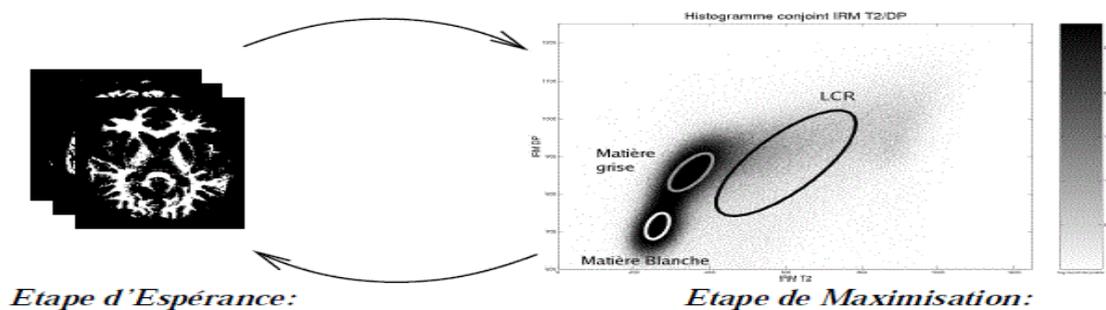


Figure 2.2 Schéma simplifié de l'algorithme EM : alternance entre l'étape d'espérance Calcul des probabilités a posteriori et l'étape de

maximisation estima des paramètres de classes.

***Application aux mélanges de gaussiennes :**

Parmi les représentations les plus courantes de la radiométrie des tissus cérébraux. le modèle gaussien est certainement celui qui remporte l'adhésion la plus large. Ainsi, à chaque tissu est associée une loi radiométrique gaussienne de dimension m . Elle se caractérise alors par sa moyenne μ et sa matrice de variance σ^2 . Ce modèle gaussien est celui que nous adoptons dans cette phase d'apprentissage.

Dans ce contexte d'apprentissage statistique, on considère chaque voxel $i, i \in \{1, \dots, n\}$, est décrit par un vecteur d'intensité m dimensionnel $y_i = \{y_i^1, \dots, y_i^m\}$. Ces vecteurs sont considérés comme les réalisations d'une variable aléatoire suivant un modèle de mélange de gaussiennes m -dimensionnelles défini comme :

$$f(y_i, \theta) = \sum_{j=1}^k \pi_j (2\pi)^{-\frac{m}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (y_i - \mu_j)^T \Sigma_j^{-1} (y_i - \mu_j)\right) \quad (2.26)$$

où k est le nombre de composantes du modèle (pour notre application $k = 3$) et $\theta = (\pi, \mu, \sigma^2)$ est l'hyper-paramètre du modèle. L'ensemble des données peut être noté $Y = \{y_1, \dots, y_n\}$ Un vecteur d'intensité m dimensionnel y_i . Les vecteurs d'intensité y_i sont indépendants identiquement distribués et suivent

un modèle de mélange de gaussiennes de paramètres $\theta = (\pi, \mu, \sigma^2)$

Les (π_j) sont les proportions du mélange, $j \in \{1, \dots, k\}$.

Les (μ_j) sont les vecteurs de moyennes, $j \in \{1, \dots, k\}$.

Les (σ_j^2) sont les matrices de variances, $j \in \{1, \dots, k\}$.

k est le nombre de composantes du modèle et correspond au nombre de tissus cérébraux, c'est-à-dire trois. c_j est une des trois composantes du mélange, c'est-à-dire une des trois classes de tissus. Y est l'ensemble des données observées. x_i est l'étiquette attribuée au voxel $i, x_i \in \{c_1, \dots, c_k\}$. L'image $X = (x_1, \dots, x_n)$ est l'image de segmentation que l'on cherche. Le but est de donner une étiquette x_i à chaque pixel i , avec $x_i \in \{c_1, \dots, c_k\}$. c_j étant une des trois composantes. D'une image de segmentation est calculée. L'image résultante $X = (x_1, \dots, x_n)$ est l'image de segmentation d'ensemble des données Y Pour calculer les

images de segmentation, il faut, une fois le modèle paramétrique défini, estimer et optimiser le calcul des paramètres

$\theta = (\pi, \mu, \sigma^2)$, suivante présente l'estimation de ces paramètres. Où μ_j et σ^2_j sont respectivement le vecteur moyenne et la matrice de covariance de la $j^{\text{ème}}$ gaussienne. Le calcul de l'espérance conditionnelle équation. (2.25) revient à remplacer les z_{ij} par leur Estimation :

$$\tilde{z}_{ij} = p(x_i = c_j | y_i, \theta^{(s)}) = \frac{\pi_j^{(s)} f_j(y_i | \theta_j^{(s)})}{\sum_{j=1}^k \pi_j^{(s)} f_j(y_i | \theta_j^{(s)})} \quad (2.27)$$

Dans ce cas particulier, la phase de maximisation peut se réduire à trois formules de ré-estimation des paramètres :

Estimation de la probabilité a priori de la $j^{\text{ème}}$ classe $\tilde{\beta}_j$:

$$\tilde{\beta}_j^{(s)} = \frac{\sum_{i=1}^n \tilde{z}_{ij}}{n} \quad (2.28)$$

Estimation du vecteur moyenne de la $j^{\text{ème}}$, classe $\tilde{\mu}_j$:

$$\tilde{\mu}_j^{(s)} = \frac{\sum_{i=1}^n \tilde{z}_{ij} y_i}{\sum_{i=1}^n \tilde{z}_{ij}} \quad (2.29)$$

Estimation de la matrice de covariance de la $j^{\text{ème}}$, classe $\tilde{\sigma}^2_j$:

$$\tilde{\sigma}^2_j^{(s)} = \frac{\sum_{i=1}^n \tilde{z}_{ij} (y_i - \tilde{\mu}_j^{(s)}) (y_i - \tilde{\mu}_j^{(s)})^T}{\sum_{i=1}^n \tilde{z}_{ij}} \quad (2.30)$$

❖ Discussion

- L'algorithme EM jouit d'une excellente stabilité numérique. Il mène à une augmentation systématique de la vraisemblance des observations, ce qui lui confère une propriété importante du point de vue de l'optimisation. Il

permet aussi d'incorporer élégamment des contraintes sur les hyperparamètres à estimer.

- L'inconvénient opposé à ces avantages est la lenteur de convergence au voisinage de l'optimum, ce qui est directement lié au taux de données manquantes. Sous peu de conditions, l'algorithme EM garantit la convergence vers un point stationnaire de la fonction de vraisemblance. Cependant, cet optimum n'est pas global mais local, la plupart du temps dépendant du point de départ.
- La propriété importante de l'algorithme EM est que la maximisation de $Q(\theta, \theta^{(s)})$ entraîne un accroissement de la fonction de vraisemblance $\log p(\mathbf{y}|\theta)$ des observations.

La théorie des champs de Markov est une branche de la théorie des probabilités pour l'analyse des dépendances spatiales de phénomènes physiques.

Dans cette chapitre nous commençons par rappeler quelques notion de base, tels le système de voisinage et cliques. Après nous présentons la définition mathématique des champs de Markov et distribution de Gibbs .Ensuit quelque sections important sur champ de Markov comme modélisation Markovienne et champ de Markov cache, dans la dernière section nous Présentons l'estimation des paramètres et donne un exemple d'algorithme EM

3.1 Voisinage et clique

3.1.1 Notion de voisinage :

En analyse d'images, l'image est représentée par une grille $S = \{s = (i, j); 0 \leq i \leq n - 1; 0 \leq j \leq m - 1\}$ rectangulaire finie bidimensionnelle de taille $n \times m$, contenant un ensemble de "sites" $s \in S$, appelés également "pixels". On appelle un système de voisinages toute famille $N = (V_s)_{s \in S}$ vérifiant les propriétés suivantes :

- Un site n'est pas son propre voisin : $s \notin V_s$.
- La relation de voisinage est symétrique : $s \in V_t \Leftrightarrow t \in V_s$.

A titre d'exemple, l'ensemble des "voisins" d'un site s peut être défini comme un ensemble de sites situés dans un rayon r : $V_s = \{t \in S / d(s, t) < r, t \neq s\}$ où $d(s, t)$ est la distance euclidienne entre le pixel s et le pixel t , et r est à valeur dans l'ensemble des entiers [11].

Notons que les systèmes de voisinage les plus usuels, représentés en figure 3.1, sont :

- Système "4 - connexité", également appelé voisinage du premier ordre;
- Système "8 - connexité", également appelé voisinage du second ordre

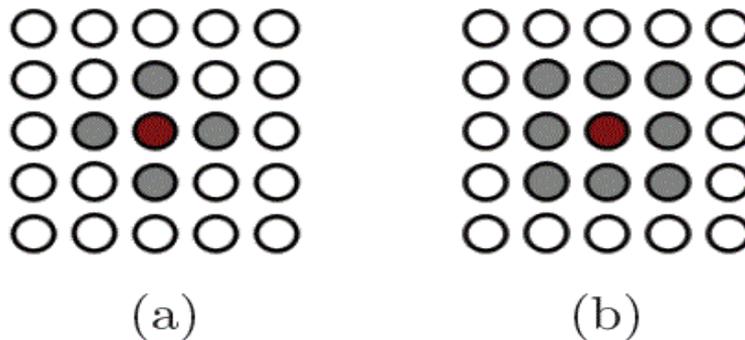


Figure 3.1 : Représentation sur une grille régulière des voisinages d'ordre 1 (4-voisinage, image a) et d'ordre 2 (8-voisinage, image b).

A partir d'un système de voisinage, un système de clique est déduit.

3.1.2 Notion de clique :

Un sous ensemble c de S est appelé clique relative au système de voisinage N , si c est un singleton ou si tous les sites distincts de c sont voisins. Une clique est dite d'ordre P . Si elle contient P éléments, en d'autres termes, l'ordre d'une clique est par définition son Cardinal. Figure 3.2 illustre la notion de cliques dans le cas de voisinage en 4-connexité et en 8-connexité. On notera C l'ensemble des cliques associées à un système de voisinage.

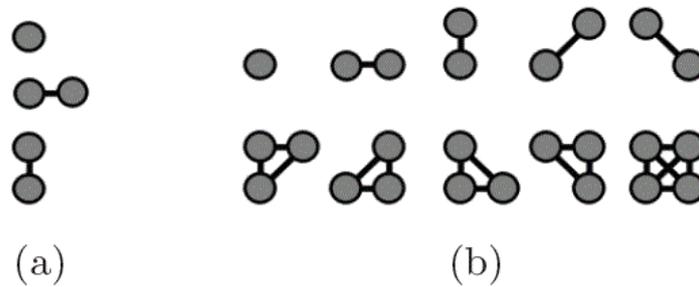


Figure 3.2 : Représentation des cliques qui interviennent dans les voisinages d'ordre 1 (image a) et d'ordre 2 (image b).

3.2 Champs de Markov et de Gibbs

3.2.1 Champ de Markov :

Soit S un ensemble de pixels et $z = (z_s)_{s \in S}$ une famille de variables aléatoires, définie sur S , où chaque z_s est à valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$. Une telle famille sera dans la suite appelée "champ aléatoire". Soit un système de voisinage N on notera par la même lettre p les diverses lois de probabilité liées à z .

On dit qu'un champ z est un champ de Markov relativement à N , si et seulement si les deux propriétés suivantes sont vérifiées [11]:

- Propriété Markovienne : $\forall z, p(z_s/z_t, t \neq s) = p(z_s/z_t, t \in V_s)$ (3.1)
- Positivité : $\forall z, p(z) > 0$

La propriété (3.1) signifie que la probabilité en un site s conditionnelle à tous les autres

Site n'est fonction que de la configuration du voisinage du site considéré. En d'autres termes, la probabilité d'apparition d'une classe donnée en un site s conditionnellement au reste du champ, est égale à sa probabilité d'apparition connaissant simplement ses voisins. Notons que tous les champs positifs aléatoires peuvent être considérés comme Markoviens à condition de choisir un système de voisinage suffisamment grand [11]. Cependant l'intérêt de restreindre le voisinage est de permettre des temps de calculs exploitables.

3.2.2 Champ de Gibbs :

Définition 1: La mesure de Gibbs d'hamiltonien $u : \Omega \rightarrow \mathbb{R}$ est la probabilité P définie sur Ω par :

$$P(z) = W^{-1}(\exp -u(z)) = \frac{\exp -u(z)}{\sum_{z \in \Omega} \exp -u(z)} \quad (3.2)$$

Avec

$$u(z) = \sum_{c \in C} u_c(z)$$

Où $W = \sum_{z \in \Omega} \exp -u(z)$ est une fonction de normalisation, aussi appelée fonction de partition de Gibbs, Il est cependant quasi impossible à calculer, même pour des configurations simples : $\text{Card}(\Omega) = 2^{262144}$ par exemple pour une image simple 512×512 à deux niveaux de gris !

et u est une énergie définie à une constante près qui se décompose en une somme de fonctions potentielles u_c associées aux cliques $c \in C$

Définition 2 : Le champ de Gibbs de potentiel associé au système de voisinage N est le processus aléatoire z dont la probabilité est une mesure de Gibbs associée à N , L'énergie totale d'un champ de Gibbs se décompose donc sous forme d'une somme d'énergies locales. Le lien entre champ de Gibbs et processus Markovien est effectué par le théorème fondamental suivant :

Théorème 1. : Théorème d'Hammersley-Clifford

Z est un champ de Markov si et seulement si $P(z)$ est une distribution de Gibbs

L'équivalence champ de Markov/distribution de Gibbs permet d'avoir une formule explicite de la distribution jointe. Grâce aux fonctions potentiels il est possible de spécifier les dépendances spatiales locales du champ de Markov. L'expression des probabilités conditionnelles est alors

$$\forall z, p(z_s/z_{V_s}) = \frac{\exp - \sum_{c \in C} u_c(z)}{\sum_{z'} \exp - \sum_{c \in C} u_c(z)} \quad (3.3)$$

3.3 Modélisation markovienne de la segmentation

Les données cachées \mathbf{Z} (la segmentation cherchée) sont décrites par un modèle de champ de Markov discret, de distribution de probabilité $p(z/\theta_z)$ définie par une énergie u et dépendant d'un paramètre : θ_z

$$p(z/\theta_z) = W^{-1}(\exp -u(z/\theta_z)) \quad (3.4)$$

Où W est la constante de normalisation :

$$W = \sum_{z'} (\exp -u(z'/\theta_z))$$

Par ailleurs on définit la dépendance des intensités observées par rapport aux étiquettes en précisant la probabilité conditionnelle suivante :

$$p(y/z, \theta_y) = \prod_{i \in S} p(y_i/z_i, \theta_y) \quad (3.5)$$

Où θ_y sont des paramètres à préciser.

On utilise classiquement une modélisation gaussienne :

$$p(y_i/z_i, \theta_y) = g_k(y_i) \text{ si } z_i = e_k$$

Où est une densité de probabilité gaussienne de moyenne et de variance dépendant de k . représente alors l'ensemble des moyennes et variances pour les K classes :

$$\theta_y = \{(\mu_k, \sigma_k^2), 1 < k < K\}$$

En utilisant la règle de Bayes on voit facilement que le champ conditionnel \mathbf{Z} sachant $\mathbf{Y}=\mathbf{y}$ est également un champ de Markov de fonction énergie u (définie à une constante près) :

$$u(z/y, \theta) = u(z/\theta_z) - \sum_{i \in S} \log p(y_i/z_i, \theta_y) \quad (3.6)$$

avec $\theta = (\theta_y, \theta_z)$. Le second terme représente le *terme d'attache aux données* (les intensités y_i), alors que $u(z/\theta_z)$ intervient lui comme un *terme de régularisation* traduisant la corrélation spatiale.

3.3.1 Modèle de Potts sans champ externe

Le terme de régularisation spatiale est souvent modélisé par un modèle de Potts simple sans champ externe :

$$u(z/\theta_z) = \sum_{i=1}^N \left[-\frac{\beta}{2} \sum_{j \in N(i)} z_i^t, z_j \right] \quad (3.7)$$

(z_i^t Note la transposée de z) de sorte que θ_z se réduit ici à un paramètre $\beta \geq 0$. Cette expression traduit le fait que l'énergie sera alors d'autant plus petite qu'un site est dans la même configuration que ses voisins. Nous avons alors :

$$u(z/y, \theta) = \sum_{i=1}^N \left[-\frac{\beta}{2} \sum_{j \in N(i)} z_i^t, z_j - \log p(y_i/z_i, \theta_y) \right] \quad (3.8)$$

3.3.2 Modèle de Potts avec champ externe :

Un modèle plus général consiste à ajouter au modèle de Potts (3.7) un terme jouant le rôle de champ externe et permettant de prendre en compte des informations *a priori* disponibles sur les étiquettes. Ainsi nous considérons :

$$u(z/\theta_z) = \sum_{i=1}^N \left[z_i^t \alpha_i - \frac{\beta}{2} \sum_{j \in N(i)} z_i^t, z_j \right] \quad (3.9)$$

Les α_i sont des paramètres supplémentaires de sorte que $\theta_z = \{\alpha_1, \dots, \alpha_N, \beta\}$. Les α_i sont des vecteurs de dimension K qui peuvent s'interpréter comme des poids associés aux K classes en chacun des sites i . Dans le cas particulier où $\beta = 0$, ce qui revient à supposer qu'il n'y a pas d'interaction spatiale entre les sites, on a :

$$p(z_i = e_k, \theta_z) = \frac{\exp -\alpha_i(k)}{\sum_{l=1}^K \exp -\alpha_i(l)} \quad (3.9)$$

Si bien que α_i pondère l'importance relative des différentes étiquettes au site i . Plus généralement, si l'importance relative des différentes classes est connue (*a priori*), elle peut être prise en compte à travers les α_i .

3.4 Champs de Markov cachés

Soit S un ensemble de pixels de cardinal N ($N = \text{Card}(S)$), et $Z = (Z_s)_{s \in S}$ $Y = (Y_s)_{s \in S}$ deux champs aléatoires, l'un caché et l'autre observé. Chaque variable aléatoire

Z_s est à valeurs dans l'ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$. et chaque variable aléatoire Y_s est à valeurs dans \mathbb{R} . En segmentation d'images, l'objectif est de retrouver le champ inobservé Z , à partir des données observées Y qui représentent la version bruitée du champ Z . à cet effet, nous devons définir la loi de (Z, Y) , ce qui est classiquement obtenu à partir de la loi de Z et celle de Y conditionnelle à Z . Dans le cas de la modélisation par champs de Markov cachés (CMC), Z est supposé markovien. Sa loi est alors donnée par :

$$p(Z) = W \exp[-U(Z)], \quad (3.10)$$

avec W une constante de normalisation, appelée également "*fonction de partition*", et U "*la fonction énergie*". Celle-ci se décompose en une somme de fonctions U_C , appelée « *fonctions de potentiel* », associées à l'ensemble des cliques C (rappelons qu'une clique est soit un singleton, soit un ensemble de pixels mutuellement voisins). Ainsi, $U(x)$ s'écrit :

$$U(z) = \sum_{C \in \mathcal{C}} u_C(z_C) \quad (3.11)$$

La fonction de partition W du fait du nombre trop important des réalisations possibles du champ Z ; cependant, il reste possible de simuler une réalisation de ce champ à partir de la loi conditionnelle de Z_s sachant Z_{V_s} tel que Z_{V_s} est la restriction du champ Z au voisinage V_s . L'échantillonneur de Gibbs que nous allons utiliser au cours de cette thèse et qui a été détaillé (voir annexe B), permet cette simulation lorsque la loi conditionnelle est connue. Cette loi est donnée par :

$$p(Z_s/Z_{V_s}) = \frac{\exp[-\sum_{C \in \mathcal{C}, S \in C} u_C(z_C, z_{C-\{s\}})]}{\sum_{\omega \in \Omega} \exp[-\sum_{C \in \mathcal{C}, S \in C} u_C(z_C = \omega, z_{C-\{s\}})]} \quad (3.12)$$

Ainsi, dans le cas des champs de Markov cachés, le champ inobservé Z est supposé markovien, ce qui justifie l'appellation CMC. La loi de Z étant définie, il reste à définir la loi conditionnelle $p(Y/Z)$. Trois CMC, de généralité croissante, peuvent être présentés selon les hypothèses posées sur cette loi conditionnelle et que nous décrivons ci-après.

3.5 Champs de Markov cachés à bruit indépendant

Les champs de Markov cachés à bruit indépendant (CMC-BI) sont la modélisation la plus simple des CMC. La probabilité conditionnelle $p(Y/Z)$ vérifie les deux hypothèses suivantes:

- 1) $p(y_s/z) = p(y_s/z_s)$ pour chaque $s \in S$
- 2) $p(y/z) = \prod_{s \in S} p(y_s/z)$

La deuxième hypothèse signifie que les y_s sont indépendants conditionnellement à Z . Du fait de 1 et 2, la loi de $p(y/z)$ s'écrit sous la forme :

$$\begin{aligned} p(y/z) &= \prod_{s \in S} p(y_s/z_s) \\ &= \exp \sum_{s \in S} \log(p(y_s/z_s)) \end{aligned} \quad (3.13)$$

La loi du couple (Z, Y) résulte des formules (3.10), (3.11) et (3.13) Elle s'écrit comme suit :

$$\begin{aligned} P(z, y) &= p(z) p(y/z) \\ &= W \exp [-(\sum_{c \in C} u_c(z_c) + \sum_{s \in S} \log(p(y_s/z_s)))] \end{aligned}$$

La probabilité *a posteriori* $p(z/y)$ du champ Z est alors donnée par :

$$\begin{aligned} P(z/y) &= W(y) \exp [-U(z, y)] \\ &= W(y) \exp [-(\sum_{c \in C} u_c(z_c) + \sum_{s \in S} \log(p(y_s/z_s)))] \end{aligned}$$

Avec :

$$W(y) = \sum_{z \in \mathcal{O}^N} \exp [-(\sum_{c \in C} u_c(z_c) + \sum_{s \in S} \log(p(y_s/z_s)))]$$

Cette formulation montre que la distribution *a posteriori* du champ Z reste une distribution de Gibbs, ce qui signifie que la loi du champ caché conditionnellement à l'observation est également une loi de Markov. Cela rend possible la simulation des réalisations de Z selon sa probabilité *a posteriori*, et permet ainsi d'utiliser différentes méthodes de segmentation bayésienne pour plus détaillé (voir section 2.1 de chapitre précédent), tels que le "Maximum a posteriori" MAP, "Iterated Conditional Mode" ICM (voir annexe A) ou le "Maximum des Marginales a Posteriori" MPM.

3.6 Processus de segmentation

Une fois le modèle précisé la segmentation peut être réalisée selon le principe de Maximum A Posteriori (MAP). Il s'agit de maximiser en \mathbf{z} la probabilité :

$$P(\mathbf{z}/y, \theta) = \frac{\exp -u(\mathbf{z}/y, \theta)}{\sum_{\mathbf{z}'} \exp -u(\mathbf{z}'/y, \theta)} \quad (3.14)$$

Le calcul nécessite d'évaluer la fonction de partition $\sum_{\mathbf{z}'} \exp -u(\mathbf{z}'/y, \theta)$. Dès que V est trop grand le nombre de configurations pour \mathbf{z} explose, rendant le calcul impossible de manière directe. Avec un volume $256 \times 256 \times 256$ et $K=3$ classes.

W est une somme de $3^{256^3} = 3^{16777216}$ termes ! (3^{256} est déjà un nombre à 123 chiffres...) De plus, tout ou partie des paramètres θ sont en général inconnus et donc à estimer comme précisé dans la section suivante.

3.7 . Estimation des paramètres

Nous nous plaçons dans un cadre de segmentation markovienne non supervisée : les paramètres $\theta = (\theta_y, \theta_z)$. du modèle de Markov ne sont pas connus et il faut les estimer.

3.7.1 Algorithme ICM :

L'algorithme ICM (Iterated Conditional Mode, [Besag, 1986]) est un algorithme qui alterne segmentation et estimation des paramètres et donc se ramène à chaque étape au cas avec données complètes. A l'itération q , l'étape de segmentation est réalisée en cherchant le mode conditionnel de la distribution du champ de Markov pour la valeur courante θ^{q-1} des paramètres du modèle et à l'aide de la segmentation courante \mathbf{z}^{q-1} , l'étape d'estimation, θ_y, θ_z sont réestimés pour la segmentation courante, respectivement par maximum de la log-vraisemblance conditionnelle et par maximisation de la pseudo-vraisemblance (voir annexe A pour plus détails) .

Cet algorithme donne des segmentations acceptables mais est connu pour induire un biais dans l'estimation des paramètres, principalement du à l'utilisation d'une segmentation « dure » (opposée à floue) à chaque étape de l'algorithme.

3.7.2 Algorithme EM :

Dans ce paragraphe nous allons présenter l'estimation des paramètres du modèle CMCBI que nous utiliserons dans la suite de ce manuscrit. L'estimation des paramètres peut se faire de manière stochastique (SEM (Stochastic Expectation-Maximisation)[14], ICE (Iterative Conditional Estimation) [13]) ou déterministe (algorithme EM (Expectation-Maximisation) [12]). Dans notre cas, les paramètres du modèle ont été estimés avec l'algorithme EM. Nous détaillerons donc seulement cette méthode d'estimation. L'EM est une méthode d'optimisation itérative pour estimer les paramètres au sens du maximum de vraisemblance. Au lieu d'utiliser seulement les données observées Y , les observations sont augmentées avec les variables cachées Z pour simplifier les calculs. Les données observées sont ainsi considérées comme des données incomplètes, auxquelles on ajoute les données manquantes Z pour aboutir aux données complètes (Z, Y) . Dempster et al. [15] proposent un algorithme en deux étapes :

– Etape E (Estimation)

$$Q(\theta, \theta^q) = E[\log p(z, y / \theta) / y, \theta^q]$$

– Etape M (Maximisation)

$$\theta^{(q+1)} = \operatorname{argmax}_{\theta} Q(\theta, \theta^q)$$

Dans le cas des Chaines de Markov cachées, nous avons la relation suivante :

$$P(z, y / \theta) = p(z_1 / \theta_z) \prod_{n=2}^N p(z_n / z_{n-1}, \theta_z) \prod_{n=1}^N p(y_n / z_n, \theta_y) \quad (3.12)$$

Nous pouvons donc écrire :

$$\begin{aligned} Q(\theta, \theta^q) &= E[\log p(z, y / \theta) / y, \theta^q] \\ &= \sum_{n=1}^N E[\log p(y_n / z_n, \theta_y) / y, \theta^q] + \sum_{n=2}^N E[\log p(z_n / z_{n-1}, \theta_z) / y, \theta^q] \\ &\quad + E[\log p(z_1 / \theta_z) / y, \theta^q] \\ &= \sum_i W_1^{(q)}(i) \log \pi_i + \sum_{n=1} \sum_i \sum_j \xi_n^{(q)}(i, j) \log t_{ij} + \sum_n \sum_i W_n^{(q)}(i) \log f_i(y_n) \end{aligned}$$

avec :

– les probabilités marginales a posteriori

$$\begin{aligned} W_1^{(q)} &= p(z_n = \omega_i / y, \theta^q) \\ &= \frac{\alpha_n(i) \beta_n(i)}{\sum_j \alpha_n(i) \beta_n(i)} \end{aligned} \quad (3.13)$$

– les probabilités conjointes a posteriori

$$\begin{aligned} \xi_n^{(q)}(i,j) &= p(z_{n-1}=\omega_j, z_n=\omega_i / y, \theta^q) \\ &= \alpha_{n-1}(j) t_{ij} f_i(y_n) \beta_n(i) \end{aligned} \quad (3.14)$$

les probabilités initiales

$$\pi_i = p(z_1 = \omega_i)$$

– la matrice de transition

$$t_{ij}^n = p(z_{n+1} = \omega_j / z_n = \omega_i)$$

L'algorithme EM sur une chaîne de Markov a été développé par Baum [Baum 72]. On maximise Q sous les contraintes suivantes : $\sum_i \pi_i = 1$ et $\sum_j t_{ij} = 1$. En utilisant les multiplicateurs de Lagrange, on obtient [16] :

$$\pi_i^{q+1} = w_i^q(i) \quad (3.15)$$

$$t_{ij}^{q+1} = \frac{\sum_{n=2}^N \xi_n^{(q)}}{\sum_{n=1}^{N-1} w_n^q(i)} \quad (3.16)$$

$$\sum_n \sum_i w_n^q(i) \log f_i(y_n) = \sum_n \sum_i w_n^q(i) \left[-\log \sigma_i - \frac{1}{2\sigma_i^2} (y_n - \mu_i)^2 \right]$$

En dérivant $\sum_n \sum_i w_n^q(i) \log f_i(y_n)$ par rapport à μ_i (resp. σ_i) et en égalant le résultat à 0, on obtient :

$$\mu_i^{q+1} = \frac{\sum_n w_n^q(i) y_n}{\sum_n w_n^q(i)} \quad (3.17)$$

$$\sigma_i^{2(q+1)} = \frac{\sum_n w_n^q(i) [y_n - \mu_i]^2}{\sum_n w_n^q(i)} \quad (3.18)$$

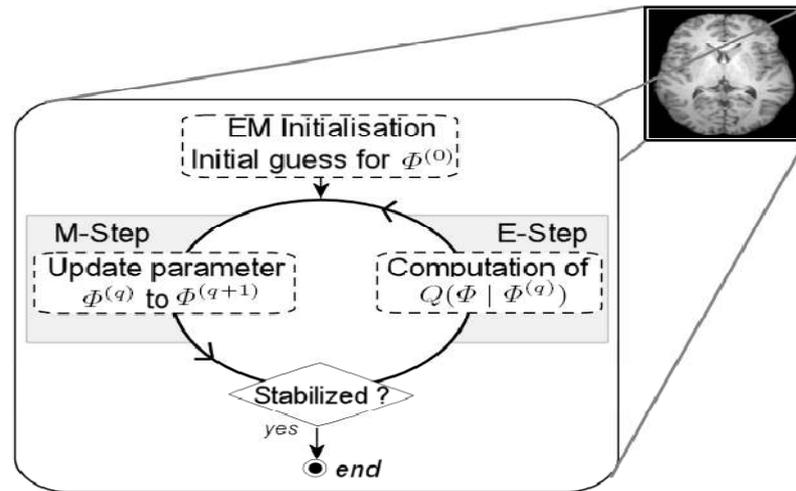


Figure. 3.3 – Schéma classique de l’estimation par EM : une seule procédure estime des Paramètres globaux sur tout le volume. Ces modèles globaux ne reflètent Cependant pas les propriétés locales de l’image.

Dans la majorité des approches une seule procédure EM globale estime les paramètres d’un champ de Markov sur toute l’image (voir **Figure. 3.3**). Les modèles estimés globalement ne reflètent cependant pas les propriétés locales de l’image. Dans le chapitre suivant on implémentées et présentée le résultat obtenu.

- X** Champ des étiquettes.
- Y** Champ des observations.
- x** Réalisation du champ des variables cachées.
- y** Réalisation du champ des observations.
- S** Ensemble de sites ou nœuds.
- E(x)** Espérance mathématique de **x**.
- L(.,.)** Fonction de coût.
- V_s** Ensemble des voisins du site **s**.
- N** Système de voisinage.
- c** Clique relative à un système de voisinage.
- C** Ensemble de cliques.
- U(x)** Energie associée au champ **x**.
- $\tilde{\theta}$** Estimateur de θ .
- s** Image.
- W** Partition associée à l'image.
- N** Nombre de voxels dans l'image.
- K** Nombre de segments.
- Ω** Ensemble des classes.
- Θ** Ensembles des paramètres.

Acronyme

CMC	Chains de Markov Cachée
CMC-BI	Chaîne de Markov cachée à bruit indépendant
CMC	Chaîne de Markov Couple
ddp	densité de probabilité
EM	Expectation-Maximisation
EMS	Expectation-Maximisation Segmentation
HMC	Hidden Markov Chain
IRM	Imagerie par Résonance Magnétique
IID	indépendantes et identiquement distribué
LCR	Liquide céphalo-rachidien
MAP	Maximum a Posteriori
MPM	l'estimateur du mode marginal
MV	maximum de vraisemblance
MB	Matière Blanche
MG	Matière Grise
MRF	Markov Random Field
SEP	Sclérose en Plaques
SPM	Statistical Parametric Mapping

Figure 1.1 : Système nerveux central (Encéphale).....9

Figure 1.2: Photographie d'un cerveau post-mortem, visualisation des principaux
Éléments10

Figure 1.3 : Le tronc cérébral.....11

Figure 1.4: Une coupe histologique d'un cerveau mettant en évidence les trois matières
principales de l'encéphale12

Figure 1.5 : quelques structures cérébrales.....14

Figure 1.6 Couche externe au cerveau.....15

Figure 1.7. Sclérose en plaque16

Figure 1.8- AVC importantes.....17

Figure 1.9- Présence d'une tumeur.....17

Figure 1.10 : Différentes coupes du cerveau.....19

Figure.1.11 : Le mouvement de spin.....21

Figure 1.12 : Mouvement de précession d'un ensemble de moments magnétiques.....22

Figure 1.13 : Excitation par l'onde RF.....23

Figure 1.14 : Les différentes séquences d'une image IRM.....26

Figure. 1.15 : Artéfact d'inhomogénéité d'intensité27

Figure 1.16 : Artéfact de volume partiel dû à la discrétisation de l'espace.....28

Figure 2.1 : le fonctionnement de l'algorithme EM.....41

Figure 2.2 :Schéma simplifié de l'algorithme EM43

Figure 3.1 : Représentation sur une grille régulière des voisinages d'ordre 1.....48

Figure 3.2 : Représentation des cliques qui interviennent dans les voisinages49

Figure. 3.3 : Schéma classique de l'estimation par EM58

Figure 4.1 : présentation du logiciel Matlab.....61

Figure 4.2 :La fenêtre du logiciel M RICRO61

Figure 4.3 : segmentation par algorithme EM avec différente initialisation
des paramètres.....64

Figure 4.4 :segmentation par algorithme EM65

Figure 4.5 :Segmentation d'une IRM cérébrale : a gauche image originale, a droite.
segmentation en utilisant un algorithme EM avec un a priori markovien.....66

Figure 4.6 :Segmentation d'une IRM cérébrale : (a). image originale,(b). segmentation en
utilisant un algorithme EM, (c). segmentation en utilisant un algorithme EM avec
un a priori markovien.....67

Figure 4.7 :(a) images original (b) segmentation des IRM par algorithme EM avec a priori
Markovienne en 3D.....68

Table 1.1 signes et symptôme de la présence des tumeurs

Table 1.2 ordre de grandeur des temps de relaxation à 1.5 tesla

- [1] R .DAMADIAN ,tumor detection dy nuclear magntic resonance science vol . 171 .pp 1151.1153 1971
 - [2] R.CHAWKES,GNHOLLAND,w,s mooret b .s Worthington nuclear magnetic resonance N M R tomography of the brain a preliminary chemical assessment with demonstration of pathology journal of computer assisted tomography vol 4 pp 577 586 1980
 - [3] Benoit Scherrer , « *Segmentation des tissus et structures sur les IRM cérébrales : agents markoviens locaux coopératifs et formulation bayésienne* », thèse de doctorat, Institut National Polytechnique de Grenoble, 2008.
 - [4] M.Cyril JAGGI , « *segmentation par méthode markovienne de l'encéphale humain en imagerie par résonance magnétique théorie, mise en oeuvre et évaluation* »,Thèse de doctorat,Université de CAEN/Basse-Normandi,1998
 - [5] A Boucher, "*Une approche décentralisée et adaptative de la gestion d'information en vision*". Thèse de doctorat, Université Joseph Fourier, Grenoble. 1999.
 - [6] D.I. HOULT ET C.N. CHEN, "*Biomedical magnetic resonance technology*" Medical Science Series, Bristol and New York, 1989.
 - [7] C. GUINET ET J. GRELLET « *Abrégés d'imagerie radiologique* », Introduction à l'IRM : de la théorie à la pratique, Paris, 1992.
 - [8] Sean Borman. The Expectation Maximization algorithm : A short tutorial. www.isi.edu/natural-language/teaching/cs562/.../B06.pdf, Juillet 2004.
 - [9] Michael Collins. The EM algorithm. www.cse.unr.edu/bebis/CS679/Readings/EM_Algorithm_Review.pdf, Septembre 1997.
 - [10] Stephan Morgenthaler. *Génétiq ue statistique*. Collection « Statistique et probabilités appliquées ». Springer, 2008.
 - [11] S. Z. Li, "Markov Random Field Modelling in Computer Vision", *Springer*,1995.
 - [12] M.A. Tanner. Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions. Springer Verlag, 1993.
-

- [13] W. Pieczynski. Statistical image segmentation. *Machine Graphics and Vision*, vol. 1, no. 2, pages 261–268, 1992.
- [14] P. Masson & W. Pieczynski. SEM Algorithm and Unsupervised Statistical Segmentation of satellite Images. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 31, no. 3, pages 618–633, 1993.
- [15] A.P. Dempster, N.M. Laird & D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, vol. 39, pages 1–38, 1977.
- [16] J .N. Provost. Classification Bathym´etrique en Imagerie Multispectrale Spot. PhD thesis, Universit´e de Bretagne Occidentale, 2001.



Annexe A

Echantillonnage des processus de Markov

Ainsi, il suffit de définir séquentiellement une topologie, un système de cliques. Des potentiels de clique u_c et une fonction d'énergie u pour préciser le processus de Markov attaché à l'image. L'ensemble de ces données permet, comme précisé précédemment, d'accéder à la probabilité d'une configuration (i.e. pour le traitement d'images, le fait d'observer l'image que l'on voit), et aux probabilités conditionnelles locales. Se pose donc maintenant le problème de la réalisation du tirage d'une configuration donnée suivant la loi de Gibbs caractéristique (via le théorème d'Hammersley-Clifford) du champ de Markov ainsi défini.

➤ Première approche : échantillonneur de Gibbs

L'échantillonneur de Gibbs est un algorithme de relaxation probabiliste, qui se propose de réaliser ce tirage par une méthode itérative, convergeant vers une configuration suivant la loi de Gibbs. A l'itération q , la construction de la réalisation courante se fait à l'aide des étapes suivantes :

- Choix d'un site s (par tirage selon une loi uniforme sur E)
- Calcul de la probabilité conditionnelle

$$p(z_s/z_t, t \in V_s) = \frac{\exp -u(z_s, z_s, t \in V_s)}{\sum_{e \in E} \exp -u(e, z_s, t \in V_s)}$$

Suivant la configuration des sites $t \in V_s$ à l'itération précédente

- mise à jour du site s par tirage aléatoire selon $p(z_s/z_t, t \neq s)$

Dans le cas où chaque site s est visité une infinité de fois, on peut montrer que les réalisations générées sont des réalisations de la loi de Gibbs, indépendamment de la configuration initiale utilisée pour l'algorithme. En pratique, on fera converger l'algorithme en itérant le procédé "un nombre suffisant de fois", ou en détectant le faible nombre de changement dans les mises à jour de sites d'une itération à l'autre.

➤ Seconde approche : algorithme de Métropolis

L'algorithme de Métropolis, issu de la physique statistique, est également un algorithme de relaxation probabiliste. Le principe est donc itératif, et repose sur les étapes suivantes :

- Choix d'un site s (par tirage selon une loi uniforme sur E)
- Tirage aléatoire d'un descripteur $\delta \in E$ selon une loi uniforme
- Calcul de la variation d'énergie induite par le changement d'état du site s (de son état à l'itération précédente à l'état δ) :

$$\Delta U = U_s (Z_s = \delta / Z_t = z_t^{q-1}, t \in V_s) - U_s (Z_s = z_s^{q-1} / Z_t = z_t^{q-1}, t \in V_s)$$

Algorithme :

Si ($\Delta U < 0$) Alors

 Le changement d'état est accepté pour le site s

Sinon

 Le changement est accepté avec une probabilité $\exp(-\Delta U)$

Fin Si

Les critères de convergence sont similaires à ceux de l'échantillonneur de Gibbs, mais cette dernière intervient cependant plus tardivement car les transitions ne sont ici pas toujours acceptées. En revanche, l'itération est plus rapide que celle de l'échantillonneur de Gibbs. car la variation d'énergie ne se calcule qu'entre deux configurations données.

Annexe B

Algorithme ICM non supervisé

L'algorithme Iterated Conditional Modes (ICM) ([Besag, 1986](#)) permet d'approximer l'estimateur du MAP en alternant étiquetage des voxels et estimation des paramètres pour résoudre le problème de la segmentation. Nous détaillons ces deux étapes dans les sections suivantes.

➤ Etape de segmentation

Le principe de l'étape de segmentation consiste à mettre à jour les pixels de manière itérative et les uns après les autres à partir de l'information contenue dans les observations y et dans les valeurs courantes z du champ Z sur les autres pixels. Ainsi, si

$$z^{(q-1)} = (z_1^{(q-1)}, \dots, z_N^{(q-1)})$$

représente la configuration des étiquettes à l'itération $(q-1)$, et $\theta^{(q-1)}$ les paramètres estimés à l'itération $(q-1)$, la classe de chaque pixel z_i est mise à jour à l'itération q par :

$$z_i^{(q)} = \underset{z_i}{\operatorname{argmax}} (p(z_i/z_{N(i)}^{(q-1)}, y_i, \theta^{(q-1)}))$$

Le calcul de ces probabilités locales est facile car il ne fait intervenir que les différents états possibles d'un pixel, au lieu de toutes les configurations de l'image comme dans le calcul de l'estimateur du MAP. On décompose avec la règle de Bayes en :

$$\begin{aligned} z_i^{(q)} &= \underset{z_i}{\operatorname{argmax}} (p(z_i/z_{N(i)}^{(q-1)}, y_i, \theta_z^{(q-1)}) p(y_i/z_i, \theta_y^{(q-1)})) \\ &= \underset{z_i}{\operatorname{argmax}} (\exp(-H(z_i/z_{N(i)}^{(q-1)}, y_i, \theta_z^{(q-1)})) + \log(y_i/z_i, \theta_y^{(q-1)})) \end{aligned}$$

La nouvelle classe e_i du voxel i est donc donnée par :

$$L = \underset{k=1\dots K}{\operatorname{argmax}} (-H(z_i = e_i/z_{N(i)}^{(q-1)}, y_i, \theta_z^{(q-1)}) + \log(y_i/z_i = e_i, \theta_y^{(q-1)})) \quad (\text{B.1})$$

On itérant (B.2) pour tous les voxels $i = 1 \dots N$ du volume on obtient $z^{(q)}$ la nouvelle configuration des étiquettes à l'itération q . Lorsque le parcours se fait de manière séquentielle, on peut montrer que la probabilité de la segmentation courante ne décroît jamais. Cette propriété assure la convergence de l'algorithme vers un maximum local de la distribution conditionnelle $p(z|y, \theta)$.

➤ Etape d'estimation

L'estimation des paramètres à l'étape q se décompose en deux étapes :

- Le paramètre θ_y est mis à jour par maximum de la log-vraisemblance conditionnelle :

$$\theta_y^{(q)} = \underset{\theta_y}{\operatorname{argmax}} \log p(y/z^q, \theta_y)$$

Avec des densité gaussiennes de moyenne μ_k et de variance σ_k^2 la maximisation est réalisée avec les expressions :

$$\mu_k^{(q)} = \frac{\sum_{i=1\dots k} (e_{k,z_i^{(q)}}) y_i}{\sum_{i=1\dots k} (e_{k,z_i^{(q)}})} \quad (\text{B.2})$$

$$\sigma_k^{(q)} = \frac{\sum_{i=1\dots k} (e_{k,z_i^{(q)}}) (y_i - \mu_k^{(q)})^2}{\sum_{i=1\dots k} (e_{k,z_i^{(q)}})} \quad (\text{B.3})$$

- **Le paramètre $\theta_z = \{\eta\}$** représentant l'influence de la corrélation spatiale dans $H(z|\theta_z)$ peut être estimé numériquement. Ils peut aussi être interprété comme l'inverse d'une température T qui décroît pour arriver à l'état stable.

Algorithme ICM non supervisé

Algorithme

Calculer une segmentation initiale $z^{(0)}$

q= 0

BOUCLE

q ← q + 1

Estimation

Mettre à jour θ_y : Calcul de $m_k^{(q)}$ et $(\sigma^2)_k^{(q)}$ en fonction de $z^{(q-1)}$ pour tout k avec (B.2) et (B.3).

Mettre à jour θ_z

Segmentation

Mettre à jour l'étiquette de chaque voxel :

Calcul de $z^{(q)}$ en fonction de $\mu_k^{(q)}$, $\sigma_k^{(q)}$ et $z^{(q-1)}$ avec Equation avec (B.1)

TANTQUE non convergence ALLER A BOUCLE

Le test de convergence peut se baser sur l'étude de l'évolution de la vraisemblance des paramètres au modèle (lorsqu'elle n'évolue plus beaucoup), ou sur le nombre de voxels dont l'étiquette change entre deux étapes de segmentation.

Dans ce chapitre nous présentons, l'implémentation et les résultats de la segmentation par les méthodes étudiées dans les chapitres précédant.

4.1 IMPLEMENTATION

4.1.1 Environnement de travail

Ce travail a été réalisé sur ordinateur portable (pc) modèle ACER doté du processeur pentium ® Dual-Core CPU T4400 @ 2.20GHz et d'une mémoire vive RAM de 2 Giga octets.

4.1.2 Présentation de logiciel Matlab

Matlab est un système interactif de programmation scientifique pour le calcul numérique et la visualisation graphique . développé à l'origine pour le calcul matriciel (le nom Matlab est dérivée de cette représentation Matlab = Matrix laboratory) , il offre aujourd'hui bien d'autres possibilités.il contient des bibliothèques spécialisées (toolbox) qui répondent à des besoins spécifique :analyse numérique, traitement du signal ,traitement de l'image ,etc.

Matlab est un langage très performant utilisé pour :

- Le calcul scientifique
- L'acquisition des données
- Le développement des algorithmes
- La modélisations et simulation
- L'Analyse, exploration et visualisation des données et des résultats

Nous avons utilisé la version de Matlab 7.8.0 (R2009a) ,le choix est justifié par la souplesse et la facilité de travail (*Figure 4.1*).

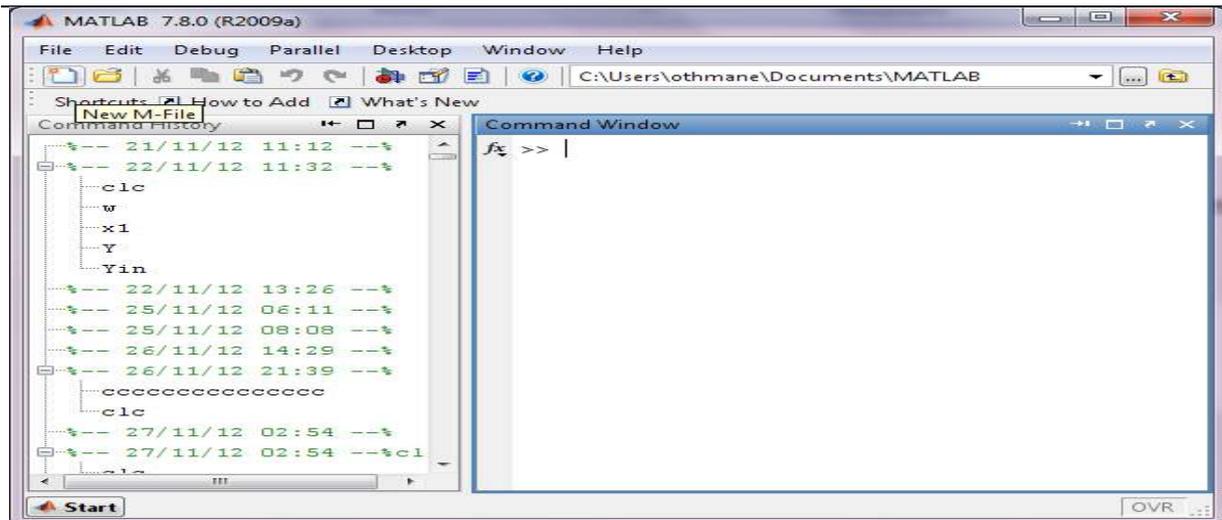


Figure 4.1 interface du logiciel Matlab

4.1.3 Logiciel MRicro

MRicro nous permet de visualiser les images 3D. L'avantage de ce logiciel c'est la précision de ces résultats le seul inconvénient est que les contrastes ne sont pas aussi détaillés que les résultats spm 2

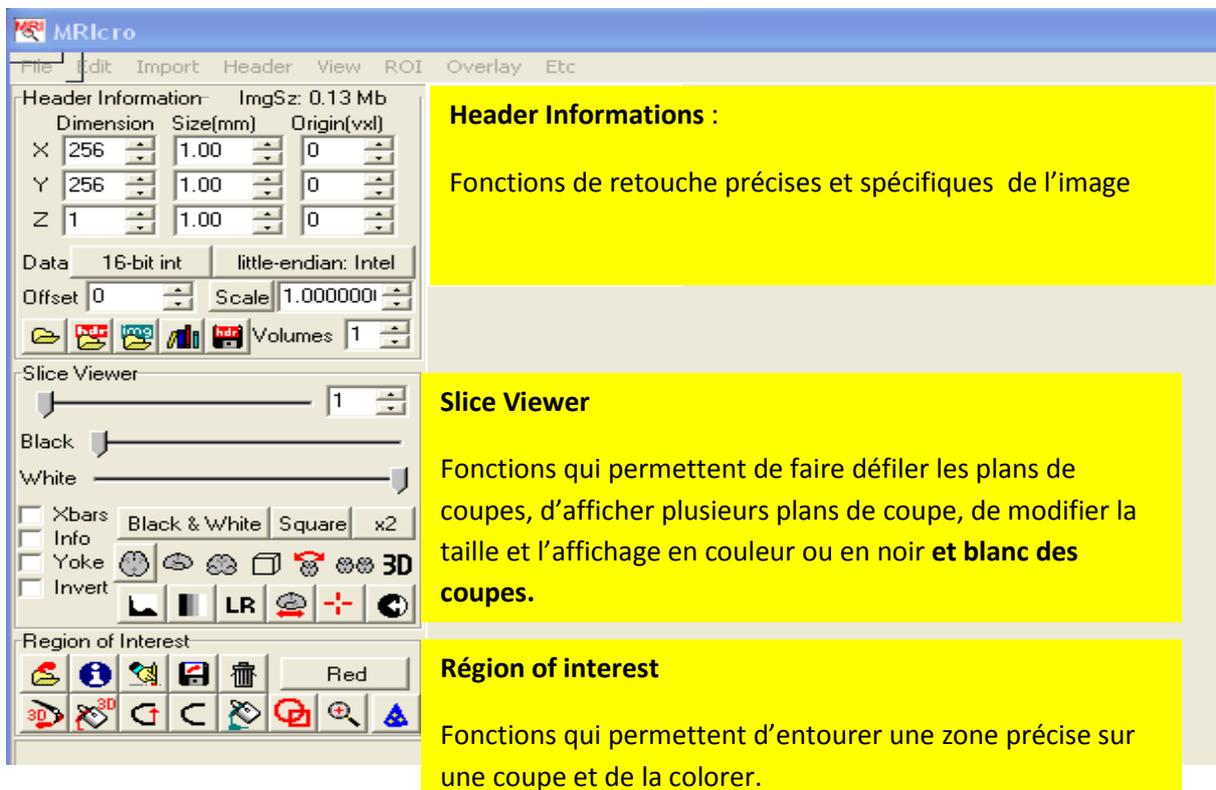


Figure 4.2 La fenêtre du logiciel MRicro

4.1.4 Implémentation de l'algorithme EM

Initialisation des paramètres

Les algorithmes d'estimation utilisés doivent être initialisés. Cette étape est importante car elle peut influencer sur la convergence de l'algorithme EM et conditionne la vitesse de convergence. Nous avons utilisé deux méthodes d'initialisation des paramètres :

❖ Première méthode d'initialisation

Pour initialiser les paramètres des moyennes (μ_j) et des écarts-types (σ_j avec $j \in \{1, \dots, k\}$ et k le nombre de classe que nous choisissons, (pour plus de détails voir chapitre 2) . « m » est le maximum de l'intensité de l'image ,pour avoir la moyennes d'une classe on multiplie m par la valeur de cette classe divisé par $k+1$.

Exemple : nombre de classe $k=3$ et $m= 255$ donc on aura 3 moyennes et 3 écarts-types pour chaque classe

$$\begin{aligned} \mu_1 &= \frac{1 \cdot 255}{4} & \sigma_1 = \sigma_2 = \sigma_3 = m = 255 \\ &= 63.75 \\ \mu_2 &= \frac{2 \cdot 255}{4} \\ &= 127.5 \\ \mu_3 &= \frac{3 \cdot 255}{4} \\ &= 191.25 \end{aligned}$$

❖ Deuxième méthode par algorithme k-means (k-Moyennes)

L'algorithme des K-Moyennes est une est très utilisé dans les techniques de clustering. Il s'agit d'un algorithme de quantification vectorielle générant son dictionnaire automatiquement à partir des observations. Le principe consiste à découper l'image en M imagettes (région) sur lesquelles on calcule les paramètres statistiques moyennes et écarts-types de chacune suivant C modalités. La $m^{ième}$ imagette sera associée au vecteur.

$$x_m = \{\mu_m^{(c)}, \sigma_m^{(c)} \forall c = 1, \dots, C\}$$

La quantification vectorielle consiste à associer à chaque vecteur x_m un élément choisi parmi un ensemble de K vecteurs appelé dictionnaire ou alphabet. Elle réalise une partition de

l'espace des observations en raison de la taille réduite de son dictionnaire. L'algorithme des K-Moyennes permet d'établir automatiquement ce dictionnaire. Chaque élément k de l'alphabet sera représenté par le centroïde C_k associé aux paramètres $c_k = \{\mu_m^{(c)}, \sigma_m^{(c)} \forall c = 1, \dots, C\}$. Le rattachement d'une imagerie m à un centroïde C_k se fait à l'aide d'une mesure de distorsion. Dans le cas de la distance euclidienne, l'algorithme des K-Moyennes consiste à minimiser le critère :

$$E = \sum_{k=1}^K \sum_{x_m \in c_k} |x_m - c_k|^2 \quad (4.1)$$

Pour une partition $\{C_k, k \in \{1, \dots, K\}\}$ donnée, l'indice de dispersion E est minimal si le centre de chaque groupement C_k est donné par la relation suivante :

$$c_k = \frac{1}{\text{card}(c_k)} \sum_{m \in c_k} x_m \quad (4.2)$$

Algorithme des K-Moyennes

1. Initialisation des K centroïdes à partir des K premières imageries

$$c_k^{[0]} = x_k \quad k \in \{1, \dots, K\}$$

2. Faire

– $q \leftarrow q + 1$

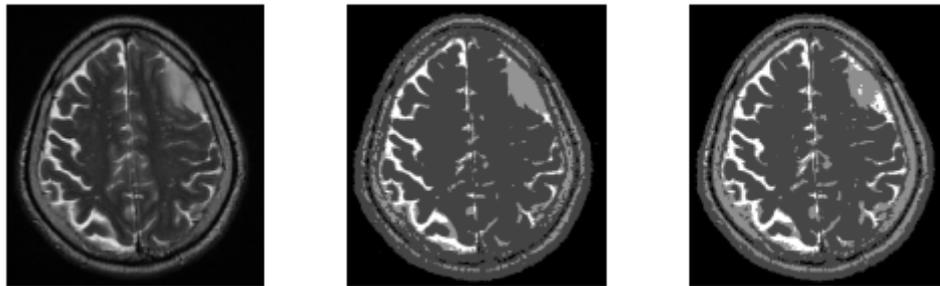
– Pour $m \in \{1, \dots, M\}$, on associe à l'imagerie m le centroïde l le plus proche au sens de la distance euclidienne.

$$C_k = \{m, |x_m - c_k| < |x_m - c_l|, \forall l \neq k\}$$

– Mise à jour des paramètres des centroïdes

$$c_k^{[q]} = \frac{1}{\text{card}(c_k)} \sum_{m \in c_k} x_m ; \forall k$$

Tant que $|c^{[q]} - c^{[q-1]}| < \varrho$ où ϱ est un seuil empirique.



(a)

(b)

(c)

Figure 4.3 segmentation par algorithme EM avec différentes initialisations des paramètres (a) Images original (b) initialisé par première méthode (c) initialisé par deuxième méthode (*k-means*)

Remarque : nous ne constatons que la meilleure segmentation de l'initialisation des paramètres est la 2^{ème} méthode (*k-means*).

4.2 Résultat

Dans cette section nous avons appliqué la segmentation par algorithme EM en deux dimensions ensuite le résultat de segmentation par l'algorithme EM avec champ de Markov pour constater la différence.

A la fin nous avons travaillé sur image trois dimensions avec logiciel MRICro

4.2.1 Segmentation par algorithme EM sur image 2D

Nous considérons un algorithme EM à 4 classes - matière blanche, matière grise, LCR, et le fond d'image.

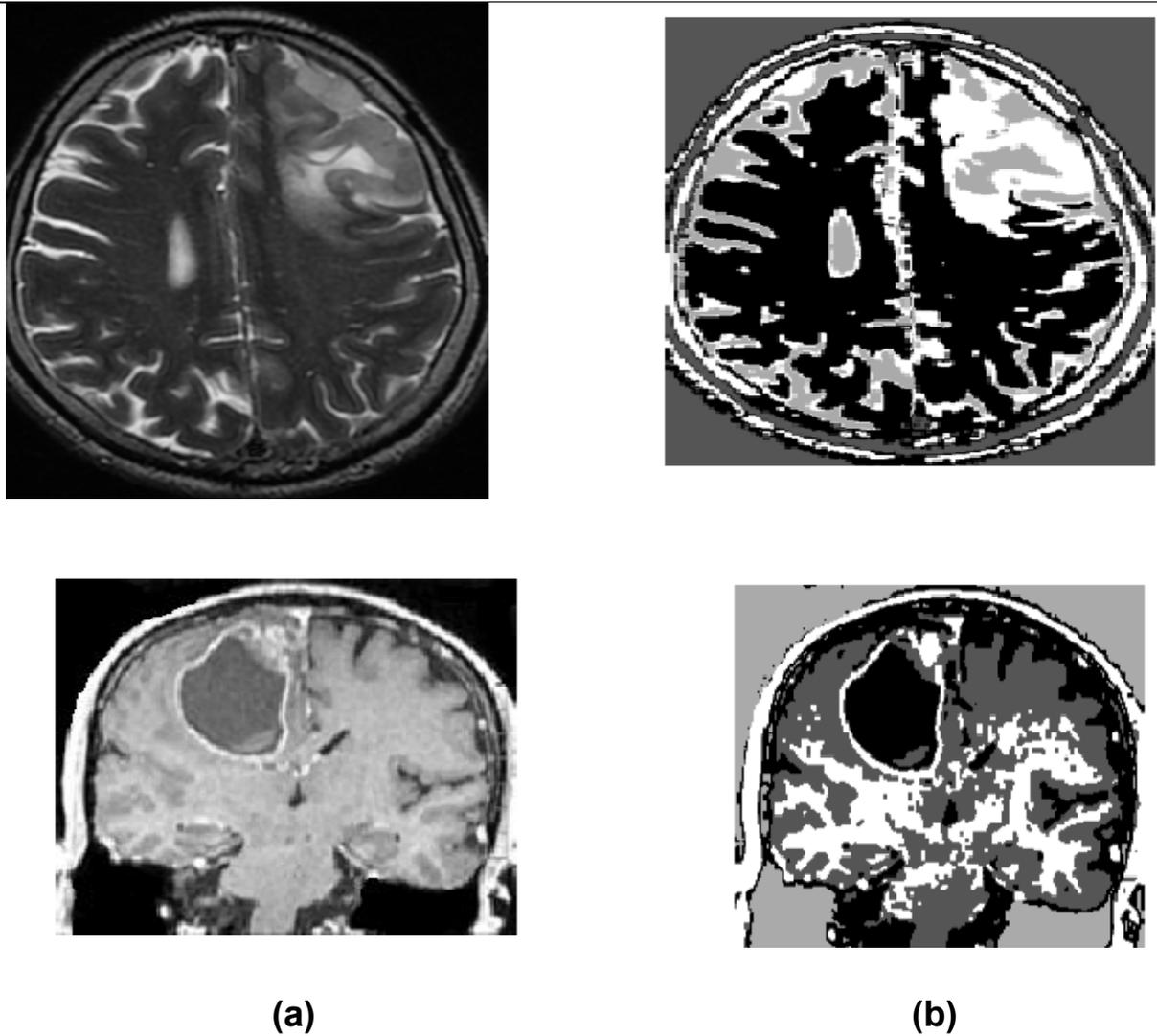


Figure 4.4 segmentation par algorithme EM (a) image originale (b) image segmenté

Discussion

La segmentation par algorithme EM n'est pas forcément la meilleur méthode , mais son avantage sur les autre algorithmes est de trouvée les données caché

4.2.2 Segmentation par algorithme EM avec champ de Markov sur image 2D

Nous avons introduit l'utilisation des champs de Markov dans l'algorithme EM pour une meilleure régularisation spatiale.

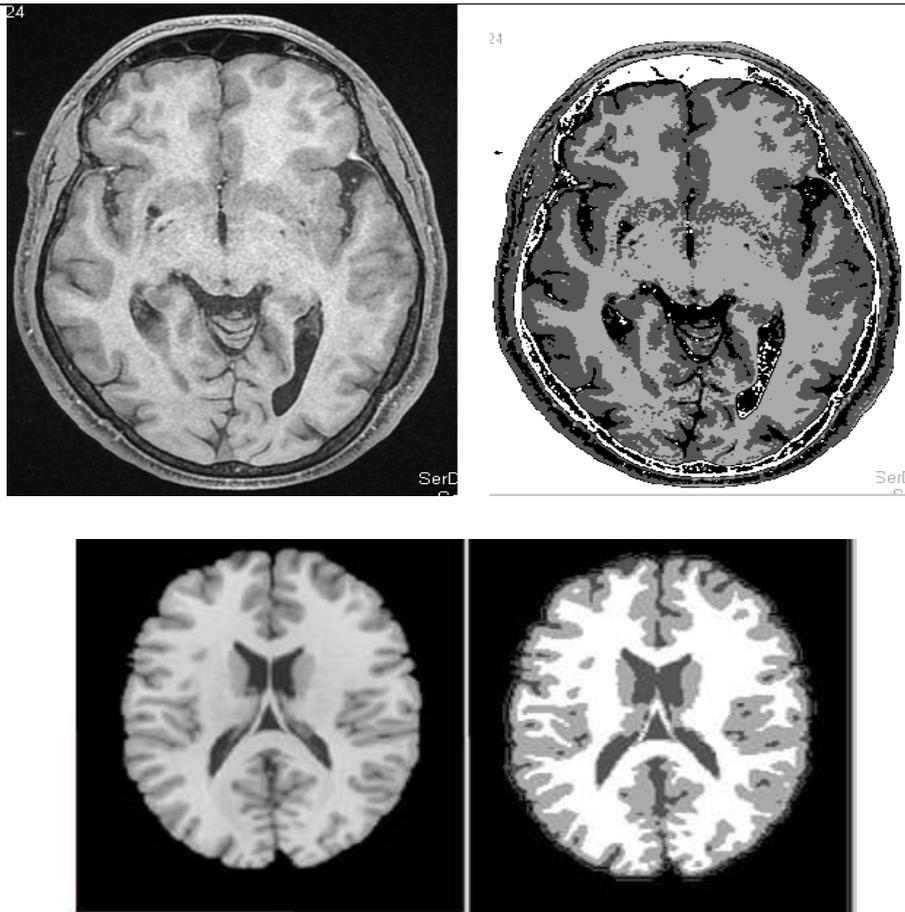
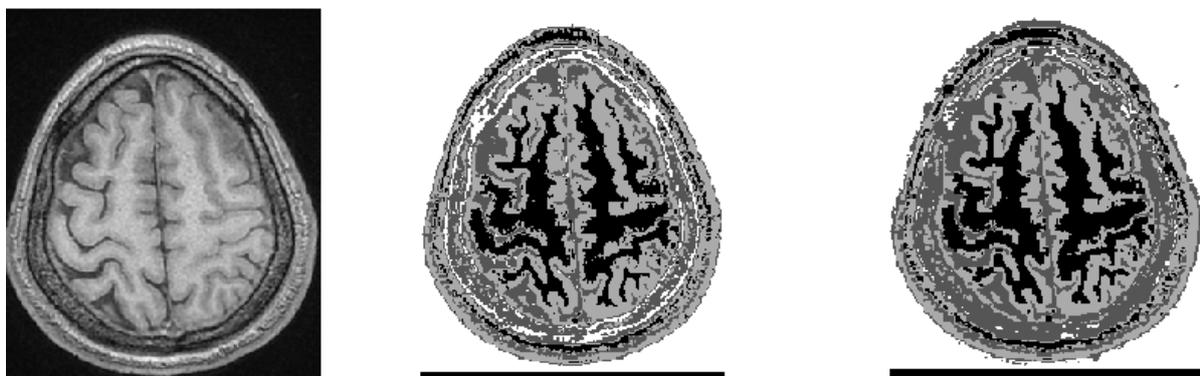
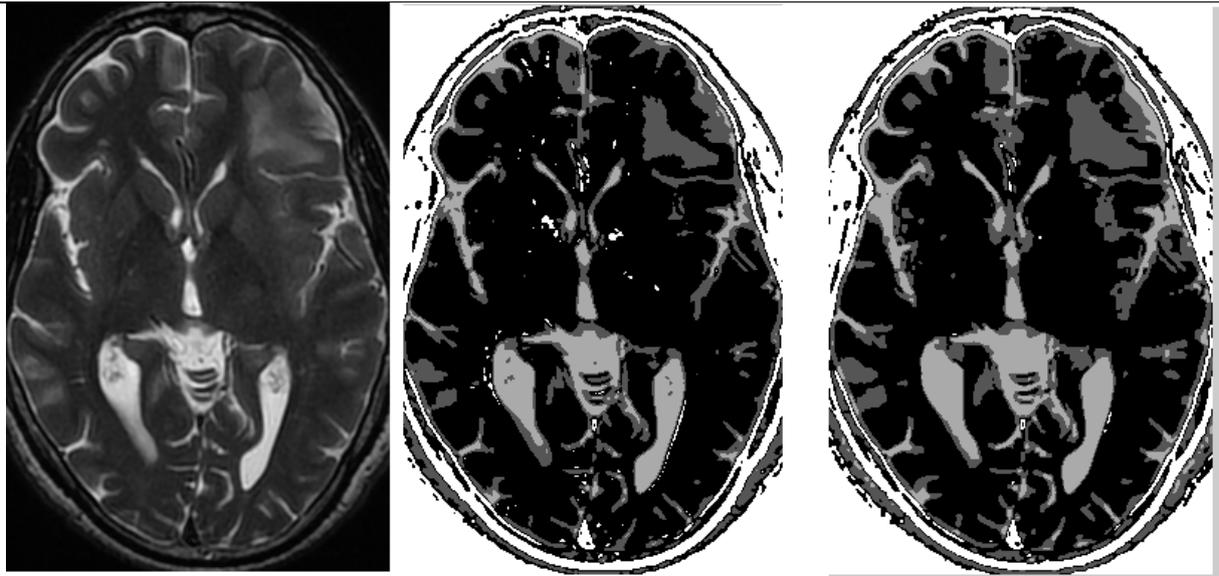


Figure 4.5 Segmentation d'une IRM cérébrale : a gauche image originale, a droite. Segmentation en utilisant un algorithme EM avec un a priori markovien Discussion L'utilisation du modélisation markovien avec l'algorithme EM donne un résultat plus performant que celui de l'algorithme EM seul.

4.2.3 Différence entre la segmentation par algorithme EM et segmentation par algorithme EM avec modélisation Markovienne





(a)

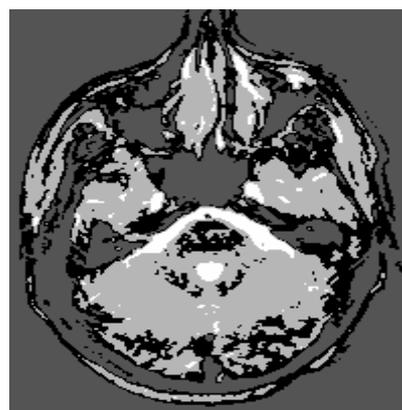
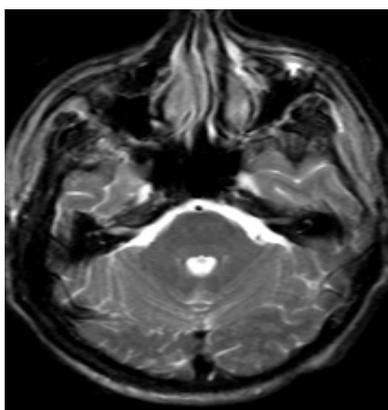
(b)

(c)

Figure 4.6 Segmentation d'une IRM cérébrale : (a). image originale, (b). Segmentation en utilisant un algorithme EM, (c). segmentation en utilisant un algorithme EM avec un a priori markovien

Nous constatons que la segmentation par algorithme EM avec champ de Markov donne de meilleurs résultats, grâce à la régularisation par champs de markov qui permet de considérer le contexte spatial.

4.2.4 Segmentation par algorithme EM avec champ de Markov sur image 3D



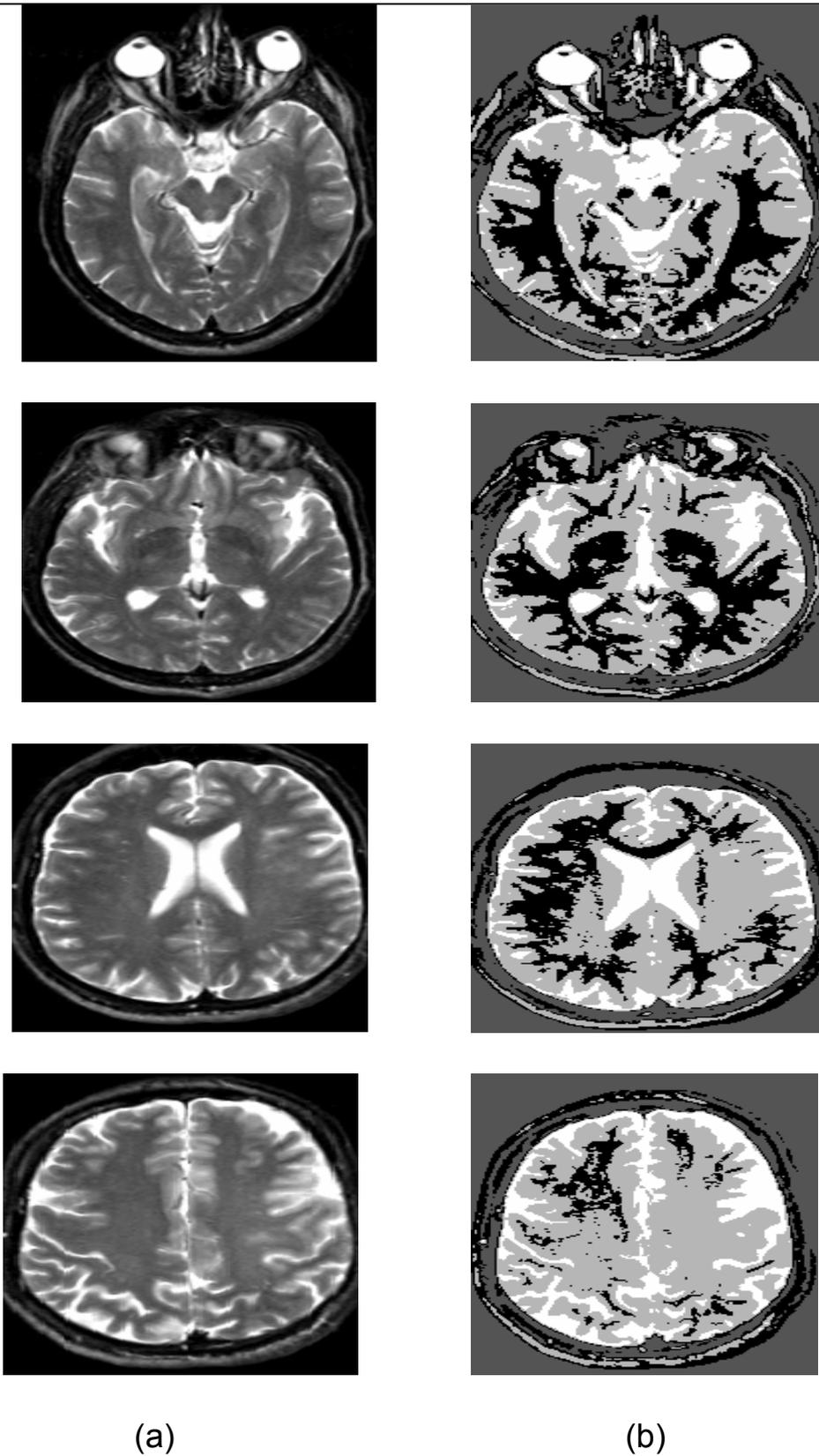


Figure 4.7 : (a) images original (b) segmentation des IRM par algorithme EM avec a priori

Markovienne en 3D

Discussion

Le fait de considérer plusieurs coupes dans la segmentation 3D nous donne une précision meilleure que celle obtenue en 2D.

Conclusion :

Dans ce chapitre, nous avons utilisé trois techniques, la première utilise l'algorithme k-means pour initialiser seulement les paramètres de l'algorithme EM, l'algorithme K-means seul est moins performant devant l'algorithme EM qui possède l'avantage d'estimer les données manquantes, la troisième technique consiste à utiliser les champs de markov combinée avec l'algorithme EM afin de considérer la corrélation spatial des voxels et permettre une robustesse vis-à-vis du bruit.