

**Analyse spatio-temporelle de séquences d'images issues de la
vidéo-microscopie de fluorescence $3D + T$**

Thèse présentée

Par

Ikhlef Bechar

Pour l'obtention du titre de

Docteur de l'université de René Descartes, Paris V

En

Mathématiques Appliquées

Dans la spécialité de

Traitement d'Image

UFR de Mathématiques et Informatique

Université de René Descartes, Paris V

11, Décembre 2007

©2007 - Ikhlef Bechar

Tous droits réservés.

**Analyse spatio-temporelle de séquences d'images
issues de la vidéo-microscopie de fluorescence $3D + T$**

Résumé

Alors que la biologie moléculaire a permis de réaliser des progrès considérables dans la compréhension de l'organisation générale de la cellule, l'espace et le temps sont toutefois demeurés secondaires tout au long de sa démarche traditionnelle. Aujourd'hui, on sait que l'espace et le temps jouent un rôle fondateur dans la mise en œuvre des processus dynamiques à l'intérieur de la cellule, d'où l'importance de leur intégration, avec les connaissances préalablement acquises dans ce domaine, en vue d'une investigation de la cellule qui n'a jamais été aussi approfondie.

Avec les progrès réalisés d'une part dans le marquage de sondes fluorescentes, et d'autre part dans les techniques de microscopie photonique, on arrive aujourd'hui à acquérir des giga octets de données images *in vivo* et *in vitro* $nD + T$ ($n = 2, 3, etc. . .$) de processus cellulaires. Toutefois, la nature complexe des dynamiques cellulaires ainsi que leur aspect tridimensionnel fait de leur analyse quantitative un grand défi pour la communauté des chercheurs biologistes et biomédicaux.

Les grands progrès enregistrés par ailleurs dans les technologies de l'information aussi bien en capacité de calcul des ordinateurs qu'en traitement logiciel de grands volumes de données permettent aujourd'hui d'appréhender le traitement de giga octets de telles données spatio-temporelles. De telles techniques qui ont l'avantage d'être semi-automatiques ou complètement automatiques, et qui peuvent générer des résultats réutilisables, devraient permettre de réaliser des avancées spectaculaires aussi bien en recherche biologique qu'en recherche pharmaceutique ou médicale de manière générale.

Vu leur nature très complexe, nous sommes persuadés que l'atteinte des objectifs tracés par des projets, tels que le projet MODYNCELL5D¹ dans lequel s'inscrit cette

¹[http : //www.irisa.fr/vista/Themes/ACI/MoDynCell5D/MODYNCELL5D.html](http://www.irisa.fr/vista/Themes/ACI/MoDynCell5D/MODYNCELL5D.html)

thèse, qui ont pour objectif l'analyse des données image issues de la vidéo-microscopie de fluorescence et dans lesquels sont impliquées des équipes regroupant des compétences multi-disciplinaires (biophysique, biologie, microscopie, informatique, traitement d'images, et modélisation) ne peuvent être atteints de façon tout à fait satisfaisante que par le biais d'approches innovantes et exploitant au mieux l'information, souvent redondante spatio-temporellement, contenue dans de telles données image. La complexité de ces données provient d'abord du fait de leur nature $nD + T$ ($n \geq 3$) et de leur rapport signal sur bruit très faible, mais aussi de part la complexité des dynamiques cellulaires observées qui (i.e. la complexité) se résume dans :

- La forte densité des objets mobiles (plusieurs dizaines à quelques centaines d'objets mobiles) dans la scène biologique observée.
- L'absence de caractéristiques chez ces objets biologiques pouvant facilement les différencier dans la scène.
- De tels objets mobiles peuvent se déformer au cours du temps, mais aussi disparaître de la scène, faire une apparition soudaine dans la scène, fusionner, se diviser, etc. . .
- La présence de plusieurs régions fortement contrastées dans une séquence de vidéo-microscopie et correspondant aux structures quasi-stables de la cellule (Golgi, réticulums endoplasmiques, etc. . .) et qui pourraient mettre en défaut les algorithmes de détection de sondes fluorescentes mobiles dans la scène, ou même l'analyse de leur mouvement après la détection.

Toutefois, quelqu'un peut exploiter la redondance d'information qu'il y a dans ces données, ainsi que la nature statistique de la dynamique biologique observée (une telle statistique est à définir bien entendu), et éventuellement en ayant recours à la simulation sur ordinateur afin de combler le manque d'informations dans ces données, et ce, dans le but de répondre à des questions biologiques qui *a priori* ne sont pas évidentes.

En effet, ces dernières idées que nous venons de décrire succinctement constituent l'idée motrice de cette thèse, et tout au long de ce manuscrit, nous nous efforçons d'y apporter des réponses autant qu'on le peut ; un effort qui, souvent, nous a demandé d'aller fouiner dans une littérature qui jusqu'ici était très peu explorée ou carrément méconnue parmi la

communauté du traitement d'images et de vision par ordinateur.

Ce manuscrit est organisé comme suit. Après donc un premier chapitre qui décrit le contexte de cette thèse et qui la situe dans le projet MODYNCELL5D dans lequel elle s'inscrit, nous proposons d'étudier dans le premier chapitre quelques artefacts d'acquisition de séquences d'images issues de vidéo-microscopie de fluorescence $3D + T$, et nous proposons des méthodes permettant leur réduction. Dans le deuxième chapitre, dans la perspective d'une analyse spatio-temporelle d'une séquence de vidéo-microscopie, nous étudions la sélection simple de modèle dans un cadre gaussien non-asymptotique conformément à la théorie de sélection de modèle de Birgé et Massart (B & M). Nous nous basons pour cela sur les deux points de vue qu'ils ont décrits dans leurs articles parus respectivement en 2001 et en 2007. Nous en donnons également la version "sélection simultanée de modèles" qui permet de sélectionner, de façon simultanée, des modèles à plusieurs jeux de données afin de contrôler le nombre de sur-ajustements réalisés. Dans le troisième chapitre, nous appliquons les idées développées dans le troisième chapitre pour réaliser une segmentation spatio-temporelle d'une séquence de vidéo-microscopie $3D + T$, que nous appliquons par la suite, toujours dans le même chapitre, pour réaliser une détection spatio-temporelle des vésicules en mouvement dans la séquence. Dans le quatrième chapitre, nous proposons une méthode alternative au suivi automatique de spots fluorescents (tracking) en se basant sur la détection spatio-temporelle des vésicules réalisée dans le chapitre 4 pour l'analyse de dynamiques en vidéo-microscopie de fluorescence $3D + T$. Nous démontrons alors qu'il est possible, en utilisant des séquences simulées, et par le biais d'une certaine statistique calculée globalement dans la séquence dont nous donnons une formule possible, d'inférer la dynamique globale des vésicules dans une séquence de vidéo-microscopie sans avoir recours au suivi de chaque objet de la scène aux différents instants de la séquence. Enfin, nous terminons ce manuscrit par une petite synthèse du travail de thèse effectué ainsi que par quelques perspectives.

Table des matières

| | |
|--|-----------|
| Page de Titre | i |
| Résumé | iii |
| Table des matières | vi |
| 1 Contexte | 1 |
| 1.1 Trafic membranaire | 2 |
| 1.1.1 Principaux compartiments cellulaires ou organelles étudiés dans le projet | 3 |
| 1.1.2 Régulateurs de trafic membranaire | 8 |
| 1.2 Observations | 10 |
| 1.2.1 Marquage des protéines | 10 |
| 1.2.2 Microscopies | 10 |
| 1.3 Démarche du projet MODYNCELL5D | 12 |
| 1.4 Conclusion | 14 |
| 2 Prétraitement spatio-temporel d'une séquence de vidéo-microscopie | 15 |
| 2.1 Introduction | 16 |
| 2.2 Description de quelques méthodes de correction de la perte de fluorescence dans une séquence de vidéo-microscopie | 19 |
| 2.2.1 Correction du photo-blanchiment basée sur la moyenne instantanée des intensités | 20 |
| 2.2.2 Correction du photo-blanchiment basée sur la déformation de l'histogramme instantané des intensités | 21 |
| 2.2.3 Correction du photoblanchiment basée sur un modèle local de décroissance exponentielle de la fluorescence | 21 |
| 2.2.4 Correction du photoblanchiment basée sur un modèle de décroissance exponentielle avec prise en compte de la cohérence spatiale | 28 |
| 2.3 Stabilisation de la variance dans une séquence de vidéo-microscopie | 34 |
| 2.3.1 Calcul des jeux de moyennes et de variances | 36 |
| 2.3.2 Estimation robuste des coefficients a et b | 38 |
| 2.4 Soustraction de la partie déterministe quasi-stable d'une séquence de vidéo-microscopie | 40 |
| 2.5 Conclusion | 40 |
| 3 La sélection simple de modèle et la sélection simultanée de modèles dans un cadre gaussien | 47 |
| 3.1 Introduction | 49 |
| 3.1.1 Ajustement d'un polynome à un jeu de données | 49 |

| | | |
|----------|--|------------|
| 3.1.2 | Estimation d'une fonction constante par morceaux en utilisant la sélection de modèle | 49 |
| 3.1.3 | Estimation des composantes d'un modèle de mélange | 50 |
| 3.1.4 | Apreçu succinct sur la sélection de modèle basée sur un critère pénalisé | 50 |
| 3.2 | La sélection de modèle simple dans un cadre gaussien | 53 |
| 3.2.1 | Processus isonormal linéaire | 54 |
| 3.2.2 | Processus linéaire gaussien | 55 |
| 3.2.3 | Modèles linéaires et estimateurs de projection | 55 |
| 3.3 | Deux critères serrés pour la sélection de modèle dans un cadre gaussien | 59 |
| 3.3.1 | Le premier critère pénalisé (PCP) | 59 |
| 3.3.2 | Le second critère pénalisé (SCP) | 66 |
| 3.4 | Stratégie de choix des poids et des constantes dans la fonction de pénalité | 69 |
| 3.4.1 | Choix des poids L_m et des constantes K et θ dans la fonction de pénalité du premier critère | 70 |
| 3.4.2 | Choix des poids L_m et des constantes α et θ dans la fonction de pénalité du second critère | 73 |
| 3.5 | Comparaison des deux critères de sélection de modèles proposés | 75 |
| 3.5.1 | Calcul de θ et Δ | 76 |
| 3.6 | La Sélection simultanée de modèles dans un cadre gaussien | 77 |
| 3.6.1 | Choix des paramètres L_m , θ et α dans la fonction de pénalité dans le cas de la sélection simultanée de modèles | 81 |
| 3.7 | Expériences numériques | 82 |
| 3.7.1 | Formulation du problème d'estimation des points de ruptures dans un signal $1D$ comme un problème de sélection de modèle | 83 |
| 3.7.2 | Procédure de simulation des signaux $1D$ | 85 |
| 3.7.3 | Résultats obtenus et leur interprétation | 85 |
| 3.8 | Conclusion | 87 |
| 4 | Segmentation spatio-temporelle par sélection de modèle d'une séquence de vidéo-microscopie. Application à la détection automatique de vésicules | 89 |
| 4.1 | Approche de sélection de modèle pour la détection d'évènements en vidéo-microscopie de fluorescence $3D + T$ | 90 |
| 4.1.1 | Formulation du problème dans un cadre Gaussien | 92 |
| 4.2 | Détection des vésicules dans une séquence de vidéo-microscopie | 101 |
| 4.2.1 | Etat de l'art | 101 |
| 4.2.2 | Détection en se basant sur notre segmentation spatio-temporelle | 105 |
| 4.3 | Résultats expérimentaux | 106 |
| 4.3.1 | Résultats sur des signaux spatio-temporels synthétiques | 106 |
| 4.3.2 | Résultats sur des séquences d'images | 109 |
| 4.4 | Conclusion | 109 |
| 5 | Estimation de la dynamique de vésicules en vidéo-microscopie | 115 |
| 5.1 | Bref état de l'art sur l'estimation de dynamiques | 117 |
| 5.1.1 | Flot optique | 117 |
| 5.1.2 | Suivi individuel des objets en mouvement ou 'tracking' | 118 |
| 5.2 | Cadre général | 124 |
| 5.2.1 | Estimation du noyau H | 126 |
| 5.2.2 | Estimation de la distribution de probabilité empirique $f_v(\cdot)$ | 129 |

| | | |
|----------|--|------------|
| 5.3 | Méthodes de résolution du problème linéaire inverse avec une contrainte de non-négativité de la solution (5.14) | 131 |
| 5.3.1 | Résolution par le critère des moindres carrés pondérés | 131 |
| 5.3.2 | Résolution par la méthode de troncature des valeurs singulières | 133 |
| 5.3.3 | Résolution par pondération des valeurs singulières | 134 |
| 5.3.4 | Résolution par pénalisation du critère de moindres-carrés pondérés | 135 |
| 5.4 | Application de l'approche décrite ainsi que des estimateurs de densité à noyau pour estimer la dynamique globale de vésicules dans une séquence de vidéo-microscopie | 141 |
| 5.4.1 | Restauration de la distribution empirique du module de la vitesse | 142 |
| 5.4.2 | Restauration de la distribution empirique du champ de vitesse | 144 |
| 5.5 | Expériences numériques | 146 |
| 5.5.1 | Choix du noyau K et apprentissage de H à partir de séquences simulées | 147 |
| 5.5.2 | Estimation des écarts types σ_l dans le problème inverse (5.12) | 148 |
| 5.5.3 | Evaluation des performances de l'algorithme de résolution du problème inverse (5.12), Evaluation de la faisabilité de l'approche en utilisant des séquences simulées | 149 |
| 5.6 | Conclusion | 151 |
| 6 | Conclusion générale | 157 |
| A | Estimation robuste de type M-estimation | 172 |

Chapitre 1

Contexte

Ce chapitre a comme objectif de situer le travail effectué pour cette thèse vis à vis du projet plus global dans lequel il s'est inscrit ; projet qui regroupe des biologistes¹, des microscopistes², des spécialistes du traitement d'images³ et maintenant des modélisateurs⁴. Ce projet a pour principal objectif l'élucidation de certains mécanismes régissant le transport intracellulaire de membranes. Il s'appuie sur un matériel biologique (lignées cellulaires marquées) et un plateau de microscopie disponibles à l'Institut Curie. Les données que nous avons analysées ont été obtenues par video-microscopie 4D.

Dans une première section, nous présentons succinctement ce qu'on entend par trafic membranaire avant d'indiquer de manière tout aussi succincte quels sont les moyens dont on dispose pour observer ce trafic, nous terminerons par une description assez générale de la démarche pluridisciplinaire qui prévaut dans ce projet et nous situerons dans cette démarche. Il est bien entendu cependant, que le travail présenté dans cette thèse, même s'il a parfaitement sa place dans ce projet, est d'une portée plus générale. Enfin, le lecteur intéressé par les aspects biologiques de ce projet peut se reporter à plusieurs articles ou thèses [Echard et al. (1998), White et al. (1999), Hill et al. (2000), Opdam et al. (2000), Monier et al. (2002), Young et al. (2005), Del Nery et al. (2006), Racine et al. (2006, 2007a), Miserey et al. (2007)].

¹Institut Curie

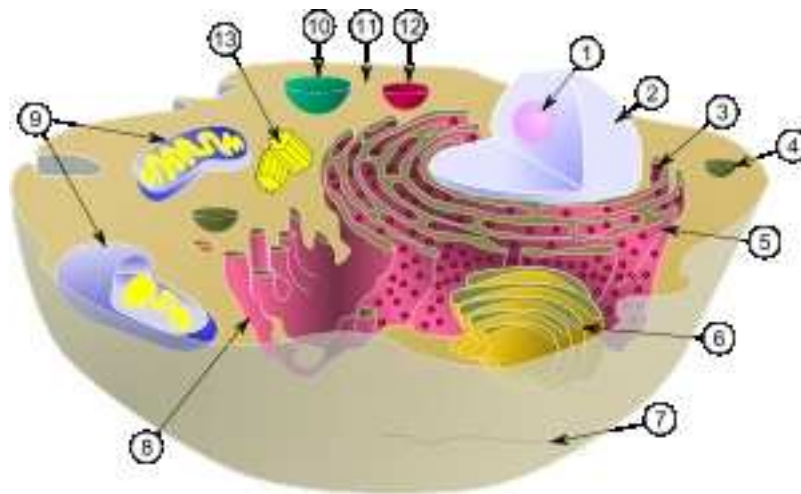
²Institut Curie

³IRISA, INRA

⁴IRISA, INRA

1.1 Trafic membranaire

Les cellules eucaryotes sont composées de plusieurs compartiments ayant des fonctionnalités propres (*c.f.* figure 1.1). Ces compartiments sont plus ou moins bien délimités spatialement par des membranes. Certains compartiments ne sont pas topologiquement séparés les uns des autres (e.g reticulum endoplasmique et noyau). Les principaux compartiments d'une cellule animale sont représentés sur la figure 1.1. Ces compartiments contiennent des



- (1) nucléole
- (2) noyau,
- (3) ribosome,
- (4) vésicule,
- (5) reticulum endoplasmique rugueux,
- (6) appareil de Golgi,
- (7) microtubule,
- (8) reticulum endoplasmique lisse,
- (9) mitochondrie,
- (10) vacuole,
- (11) cytoplasme,
- (12) lysosome,
- (13) centrosome.

FIG. 1.1: Organisation d'une cellule animale, (source Wikipédia).

molécules spécialisées et des enzymes⁵ spécifiques impliquées dans différentes fonctions plus ou moins spécifiques du compartiment. Ces compartiments (organelles) ne sont pas des entités figées mais au contraire très dynamiques et en situation d'échanges permanents avec leur environnement et en particulier d'autres compartiments dont le cytosol qui est en quelque sorte le complémentaire⁶ de tous les organelles dans la cellule. L'intérêt de cette compartimentation pour la cellule réside dans un confinement permettant la production ou la

⁵enzyme : molécule qui catalyse une réaction et qui reste intacte à la fin de la réaction.

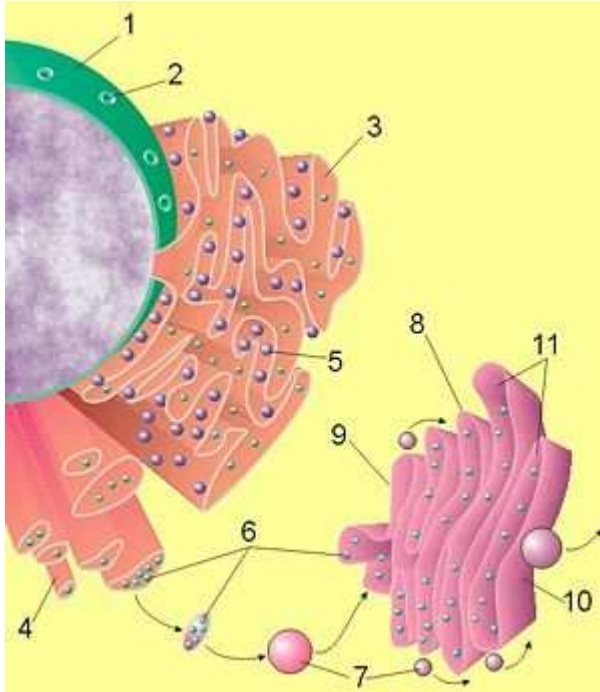
⁶au sens ensembliste

transformation de matériel utile à la vie de la cellule (comme une usine par rapport à une société humaine) mais aussi dans un 'packaging' facilitant le transport entre compartiments dans le cas des vésicules. Une autre fonction de cette compartimentation peut être la préservation d'une certaine intégrité par rapport aux variations du milieu extracellulaire (comme une maison protège ce qu'elle renferme par rapport à des intempéries). Pour comprendre pourquoi une cellule fonctionne 'anormalement', et pouvoir éventuellement prévenir ou corriger ce dysfonctionnement, il faut pouvoir remonter au moins partiellement un graphe de causalité et donc connaître l'organisation de la cellule : ce qui se fait dans les différents compartiments, ce qui s'échange et comment. Par ailleurs comme la cellule se renouvelle et se divise, il faut comprendre comment se crée les différents compartiments. De manière très imagée et simplificatrice, prenons l'exemple d'un embouteillage routier. Les causes peuvent être multiples, supposons que ce soit dû à un accident de circulation. Là encore les causes peuvent être multiples, supposons que ce soit un défaut de freinage d'un véhicule. Ce défaut peut être lié à l'état de la chaussée, à l'état des freins. Dans ce dernier cas, il peut s'agir d'un défaut des plaquettes mais peut être aussi d'une usure normale mais d'un manque d'entretien du véhicule. Par cet exemple proche de nous, on comprend mieux ce que signifie 'remonter le graphe de causalité' et le travail nécessaire pour construire un tel graphe puisqu'il faut connaître non seulement les noeuds de ce graphe mais leur fonctionnement (qui permet de construire les arêtes). L'étude du trafic membranaire relève d'une quête assez semblable dans un 'monde' où nous ne sommes pas intellectuellement les concepteurs des principaux acteurs (compartiments) et n'avons encore acquis qu'une connaissance partielle de ces acteurs, où tout se passe à une échelle pour laquelle nos moyens d'investigation sont limités. Les compartiments concernés par ce projet sont : le reticulum endoplasmique, l'appareil de Golgi, les vésicules, le cytosol et plus particulièrement le cytosquelette.

1.1.1 Principaux compartiments cellulaires ou organelles étudiés dans le projet

Le projet concerne une partie du trafic membranaire entre l'appareil de Golgi et le réticulum endoplasmique. Très schématiquement, les molécules échangées sont contenues dans des vésicules de transport (sortes de remorques) acheminées grace à des moteurs molé-

culaires (sortes de tracteurs) vers leur destination en empruntant le réseau de microtubules (sorte de réseau routier). Nous donnons ci-dessous les principales caractéristiques de ces 'acteurs' du trafic membranaire sans oublier les régulateurs de ce trafic que sont les protéines Rab.



- (1) noyau,
- (3) reticulum endoplasmique granuleux,
- (7) vésicule,
- (8) appareil de Golgi,
- (9) face cis de l'appareil de Golgi,
- (10) face trans de l'appareil de Golgi.

FIG. 1.2: Schématisation du reticulum endoplasmique avec les ribosomes à proximité du noyau et de l'appareil de Golgi (source : Wikipédia).

Le **reticulum endoplasmique (RE)** est un compartiment très proche du noyau (*c.f.* figure 1.2) dont certaines parties de la membrane sont en continuité avec la membrane externe du noyau et dont l'intérieur est en continuité avec l'espace intermembranaire du noyau (d'un point de vue topologique je ne sais si tout cela tient debout mais retenons que le reticulum endoplasmique et le noyau peuvent avoir une communication de proximité). Une partie de ce reticulum (REG) est couverte de ribosomes qui assemblent les acides aminés en chaînes protéiques en retour de 'commandes' produites au niveau du noyau. A l'intérieur du REG, des protéines sont donc assemblées avant d'être transportées vers

différents compartiments ou secrétées.

L'appareil de Golgi est un compartiment constitué d'un empilement de saccules membranaires ⁷ (*c.f.* figure 1.3). Les protéines du réticulum endoplasmique qui y parviennent continuent à y être modifiées entre sa face cis⁸ et sa face trans⁹. Ce compartiment est le point de passage obligé et le régulateur du trafic vésiculaire dans la voie d'exocytose¹⁰. Il régule la nature et la quantité des vésicules allant à la membrane plasmique ou vers d'autres membranes internes et participe au renouvellement membranaire.



FIG. 1.3: Appareil de Golgi d'un leucocyte (source : Wikipédia). On distingue bien l'empilement de saccules.

Les vésicules sont les principaux intermédiaires du transport membranaire. Ce sont des petits compartiments qui servent à stocker, transporter ou transformer des produits cellulaires. Beaucoup sont formées dans l'appareil de Golgi, dans le réticulum endoplasmique ou bien à partir de la membrane cytoplasmique. Les vésicules de transport sont d'abord créées à partir d'un compartiment donneur (e.g. Golgi ou RE).

Le cytosquelette est un réseau de fibres intracellulaires qui parcourt le cytoplasme. Il est constitué de trois catégories de fibres : les microtubules, les filaments intermédiaires et les microfilaments (*c.f.* figure 1.4).

Tous ces filaments sont des polymères allongés résultant de l'assemblage de protéines monomères. Les microtubules sont des structures très rigides d'un diamètre extérieur

⁷Est-ce alors un ou plusieurs compartiments membranaires

⁸du même côté que le RE

⁹de l'autre côté

¹⁰L'exocytose (du grec -exo "hors de" et de -kutos "cavité, cellule") est le mécanisme par lequel la cellule libère de larges biomolécules à travers sa membrane.

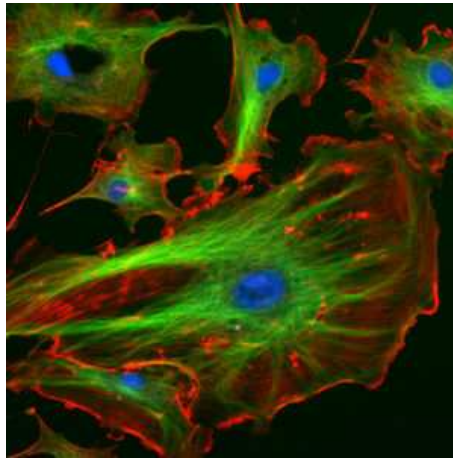


FIG. 1.4: Réseau des microtubules (en vert), réseau d'actine (en rouge), noyau (en bleu). (source : Wikipédia).

d'environ 24nm et d'une longueur qui varie d'une fraction de micron à quelques centaines de microns. Ce sont des structures polarisées avec des constantes de dissociation et association des sous-unités dimériques (α -tubuline et β -tubuline) différentes à leurs deux extrémités. L'extrémité présentant la croissance et décroissance la plus rapide est nommée extrémité plus alors que l'autre extrémité est appelée moins. Chez la plupart des cellules animales, les microtubules sont produits par nucléation de α -tubuline et β -tubuline en un endroit précis de la cellule appelé centrosome, localisé près du noyau. Cette nucléation s'effectue sous le contrôle d'une troisième tubuline (γ -tubuline) qui se trouve au centrosome et agit comme un germe pour la nucléation. Cependant, elle se comporte par rapport aux deux autres tubulines de façon à ce que l' α -tubuline soit à l'intérieur du centrosome et la β -tubuline orientée vers la périphérie de la cellule, ce qui conduit à cette polarisation avec l'extrémité la moins dynamique à l'intérieur du centrosome. Les microtubules peuvent avoir plusieurs rôles dans la cellule, celui qui nous intéresse ici concerne le transport de vésicules. Ils constituent alors de véritables rails que parcourent des moteurs moléculaires en traînant leur vésicule.

Les protéines moteur sont des enzymes spécialisées qui sont impliquées dans plusieurs mouvements cellulaires. Certaines génèrent un mouvement linéaire, comme les ARN polymérase qui se déplacent le long de l'ADN lors de la transcription, d'autres, telles les protéines moteur des filaments polarisés du cytosquelette, parcourent les filaments. On dis-

tingue trois familles de protéines moteur du cytosquelette qui se différencient au niveau de la séquence d'acides aminés de leur domaine moteur : la famille des myosines, la famille des kinésines (*c.f.* figure 1.5), la famille des dynéines. Par ailleurs les protéines moteurs d'une

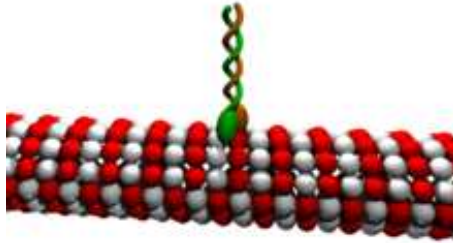


FIG. 1.5: Représentation schématique d'un microtubule et d'un moteur kinésine avec ses deux domaines globulaires à proximité du microtubule. (Source : Wikipedia).

même famille n'interagissent qu'avec une seule famille de filaments. Ainsi les kynésines et les dynéines interagissent avec les microtubules alors que les myosines interagissent avec les filaments d'actine. Ces protéines moteur sont des molécules allongées de 40 à 100nm qui comportent deux parties structurellement différentes : un domaine globulaire appelé domaine moteur à même de se fixer aux filaments et à des molécules d'ATP, une région allongée (queue) qui sert à l'attachement du moteur à d'autres protéines moteurs ou à différents cargos (vésicules,...). Une protéine moteur peut avoir plusieurs domaines globulaires. Si les séquences d'acides aminés des domaines globulaires sont très conservées dans les membres d'une même famille de moteurs, il n'en va pas de même des séquences de la queue; ce qui autorise une sélectivité des protéines transportées par un moteur.

La famille des **kinésines** est impliquée dans le transport de vésicules et la mitose. Le premier membre de cette famille a été identifié en 1985 au sein d'axones. Depuis plusieurs dizaines de kinésines ont été identifiées chez l'homme [Miki et al. (2001)] (*c.f.* figure 1.6). Les moteurs kinésine sont des moteurs de transport efficaces, leur vitesse maximale est de $2\mu\text{ms}^{-1}$ *in vivo* [Howard et al. (2001)] et leur déplacement généralement dirigé vers l'extrémité plus des microtubules mais il y a des exceptions. Plus récemment des observations de molécules uniques utilisant des nanoparticules (quantum dots) ont permis de mesurer des

paramètres de vitesse, processivité¹¹ de moteurs dans des lignées cellulaires Hela [Courty et al. (2006)].

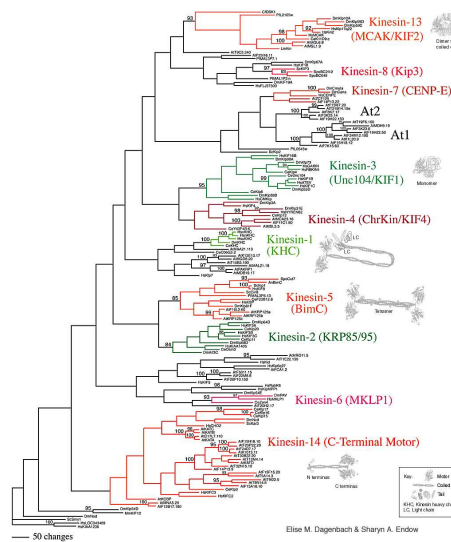


FIG. 1.6: Classification des kinésines.

L'étude du fonctionnement du couple moteur-microtubule a suscité de nombreux travaux expérimentaux [Nedelec et al. (2001), Seitz et al. (2006), Courty et al. (2006), Cai et al. (2007)] et théoriques [Klein et al. (2006)].

1.1.2 Régulateurs de trafic membranaire

Le transport des molécules d'un compartiment à un autre peut schématiquement être décrit par une succession d'évènements élémentaires que sont la création ou bourgeonnement de la vésicule, son transport vers le compartiment accepteur, son arrimage puis sa fusion à ce compartiment. La régulation de ces étapes est très complexe et mal connue. Elle est au coeur du travail des biologistes du projet.

Le processus de formation des vésicules nécessite le recrutement de complexes cytosoliques ou protéines manteau (COP I et II, clathrine, AP 1 et 2). Ce recrutement se fait via des GTPases telles que les protéines Arf, dynamine et Rab dans certains cas. L'assemblage contigu de plusieurs protéines manteau à la membrane du compartiment donneur conduit

¹¹temps d'attachement du moteur au filament

localement à sa déformation puis à un bourgeonnement du côté du manteau. La vésicule se détache du compartiment donneur, emportant un chargement de molécules sélectionnées (protéines solubles et membranaires, lipides, ions). Une fois formée, la vésicule se sépare de son manteau après hydrolyse du GTP par les GTPases activées par GAP (GTPase Active Protein).

L'arrimage des vésicules sur le compartiment accepteur serait catalysé par des facteurs d'arrimage (e.g. SNARE) activés sur les membranes donneuses et acceptrices. Les constituants de la vésicule sont alors intégrés au "pool" de la lumière cible où ils peuvent être recyclés puis redirigés, éventuellement vers le compartiment donneur.

Les Rab GTPases et leurs effecteurs¹² sont des régulateurs des voies de transport membranaire entre les organelles donneurs et accepteurs. La famille des protéines Rab peut directement, ou indirectement, réguler diverses étapes du transport vésiculaire, comme la sélection des cargos, le bourgeonnement des membranes, la motilité des vésicules, leur capture et amarrage, leur fusion à la membrane acceptrice. Elles peuvent intervenir dans l'accrochage des vésicules aux moteurs en permettant le recrutement de protéines effectrices qui leur sont spécifiques. Environ 60 protéines Rab ayant des activités spécifiques sont connues à ce jour [Stenmark et al. (2001), Seabra et al. (2002), Zerial et al. (2001)].

La compréhension moléculaire de l'activité des Rab a été obtenue par de multiples approches de biochimie, par l'étude statique des activités et des affinités entre les Rab, leurs effecteurs, et les autres protéines impliquées dans le transport (moteurs, cytosquelette, cargos). On connaît donc beaucoup de leurs partenaires potentiels mais on ne sait pas où et quand ils se rencontrent. Accéder à cette information spatio-temporelle sur les Rab et leurs partenaires semble nécessaire à une meilleure compréhension des fonctions des Rab. Cela est maintenant en partie possible par différentes techniques d'imagerie sur lesquelles nous reviendrons en sous-section 1.2.2. Les images qui ont servi de motivation et de support à ce travail ont été produites sur la plate-forme d'imagerie de l'Institut Curie et concernent deux régulateurs membranaires : Rab6A et Rab6A'. On les trouve liés à des structures vésiculaires, à l'appareil de Golgi ou libres dans le cytosol. Rab6A est un régulateur du trafic membranaire entre l'appareil de Golgi et le RE [White et al. (1999), Opdam et al.

¹²inhibiteurs ou activateurs

(2000), Monier et al. (2002)].

Pour observer ces Rab, on dispose de lignées cellulaires GFP-Rab6A et GFP-Rab6A'.

1.2 Observations

L'observation *in vivo* de protéines dans la cellule a connu ces dernières années un essor considérable, résultat d'avancées spectaculaires sur le marquage des protéines d'une part et en microscopie d'autre part.

1.2.1 Marquage des protéines

La découverte de la GFP (Green Fluorescent Protein) a été essentielle pour l'observation *in vitro* de protéines sélectionnées [White et al. (1999)]. Le gène de cette protéine peut être fusionné *in vitro* au gène d'une protéine cible qu'on veut étudier (*c.f.* figure 1.7). Le gène recombinant peut alors être introduit dans une cellule qui pourra synthétiser la protéine de fusion fluorescente. Cette protéine deviendra ainsi 'observable' en microscopie de fluorescence. Depuis l'usage de ces marquages s'est généralisé et il est possible d'observer simultanément plusieurs types de protéines marquées à l'aide de protéines fluorescentes différentes et d'étudier leurs interactions [Lippincott et al. (2003)].

1.2.2 Microscopies

La principale technique utilisée pour l'observation de protéines cibles dans la cellule est la microscopie de fluorescence. Elle consiste à exciter les fluorophores des protéines marquées avec une lumière d'excitation appropriée puis à collecter la lumière émise par ces fluorophores; tout cela au travers d'une optique très performante et à l'aide de capteurs CCD. La microscopie est dite de plein champ (wide field) quand tout l'échantillon à analyser est éclairé en même temps. L'excitation peut aussi être localisée (e.g. excitation laser) et balayer l'échantillon, c'est le cas en microscopie confocale. Dans les deux cas, il est possible d'obtenir des informations autorisant une observation 3D, c'est à dire aussi en profondeur dans l'échantillon. Dans les deux cas, un dispositif permet de sélectionner le plan focal et

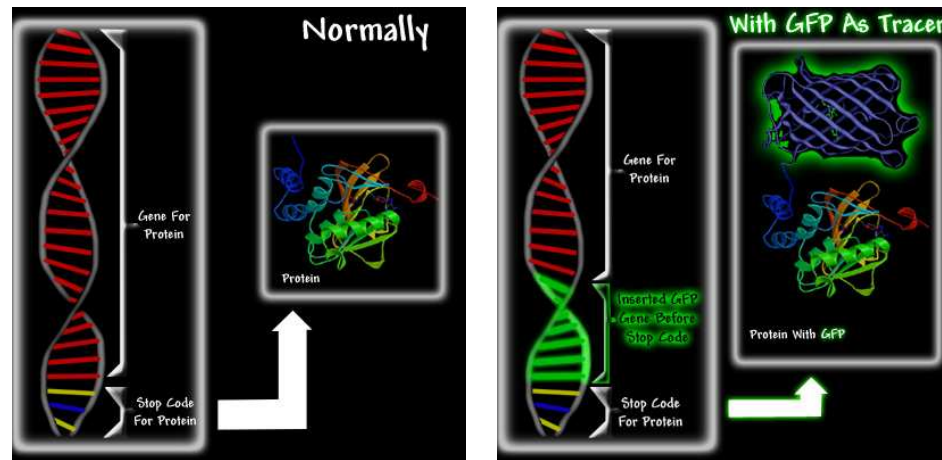


FIG. 1.7: Protéine normale (image de gauche), Protéine de fusion (image de droite). Source : <http://www.conncoll.edu/ccacad/zimmer/GFP-wv>.

d'obtenir plusieurs images restituant de manière prédominante l'information en provenance de différentes altitudes ou profondeurs z dans l'échantillon. En microscopie de plein champ, les différentes images acquises ne comportent pas une information suffisamment spécifique d'une altitude z donnée dans l'échantillon et un traitement numérique dit de déconvolution des images est nécessaire. En microscopie confocale, l'information est mieux séparée en z et ce traitement de déconvolution n'est pas toujours nécessaire. L'acquisition des images en microscopie de plein champ est plus rapide qu'en microscopie confocale. L'acquisition des images de fluorescence émises peut aussi être répétée dans le temps et on parle alors de video-microscopie. Lorsqu'un seul plan focal (resp ; plusieurs) est acquis à chaque pas de temps, on parle de vidéo-microscopie 2D+T (resp. 3D+T ou 4D). Un des enjeux de la microscopie est alors d'acquérir un maximum d'informations en un minimum de temps, tout en perturbant au minimum l'activité cellulaire. La durée d'illumination est contrainte par la lignée cellulaire et le niveau d'expression de la protéine marquée. Pratiquement, on recherche le meilleur compromis entre un bon rapport signal sur bruit (pour pouvoir détecter les structures fluorescentes), une bonne résolution temporelle (pour le suivi des mouvements rapides) et un minimum de temps à éclairer la cellule (pour éviter la création de radicaux libres engendrant des dommages cellulaires et le photoblanchiment). On reviendra sur ces aspects au chapitre 2.

On a coutume de dire que la résolution spatiale la plus forte en microscopie optique est de l'ordre de $0.2\mu m$ dans le plan latéral et de $0.6\mu m$ en z . On ne peut donc pas distinguer des structures fluorescentes si elles sont distantes de moins de ces valeurs. D'autres techniques de microscopies de fluorescence permettent cependant de sonder un voisinage beaucoup plus précis autour d'un point échantillonné du milieu. Ainsi, la technique de FRET¹³, permet de détecter la présence de molécules fluorescentes situées à une distance très faible, de l'ordre de quelques nm, en mesurant un transfert d'énergie entre ces molécules. La technique dite FLIM¹⁴, permet quant à elle, en mesurant les temps de vies des fluorophores de déceler des variations locales de la concentration en ions, du pH. La technique FLIM-FRET permet de déceler le passage à proximité de deux molécules fluorescentes. On peut visualiser des structures de très faibles taille en microscopie de fluorescence et même des molécules uniques [Courty (2006)], on peut aussi les localiser avec une très bonne précision de l'ordre de la dizaine de nm [Aguet (2005)]. Cependant ces molécules se présentent généralement par paquets et alors on ne peut pas séparer des paquets situés à moins de $0.2\mu m$.

Les vésicules marquées par Rab6 sont plus ou moins sphériques, de diamètre variant entre 60 et 150nm¹⁵. Les objets observés sur les images qu'on traitera dans le chapitre 4 s'étalent sur 5 à 25 pixels dans le plan xy ; ce sont le plus souvent des agrégats de vésicules ou des fragments du Golgi. Ces images sont obtenues en microscopie 4D de plein champ rapide à déconvolution [Sibarita et al. (2005)].

1.3 Démarche du projet MODYNCELL5D

Ce projet nommé MODYNCELL5D¹⁶ est donc focalisé sur le trafic membranaire entre le Golgi et le RE. L'objectif est de mieux comprendre la régulation de ce trafic et pour cela différents moyens sont mis en œuvre (*c.f.* figure 1.8) pour produire des lignées cellulaires permettant l'étude de ce trafic, observer ce trafic par différentes microscopies (video 4D, FRAP4D, FRET), extraire les informations pertinentes des énormes volumes de

¹³Forster Resonance Energy Transfer

¹⁴Fluorescence Life Time Imaging Microscopy

¹⁵Observations réalisées en microscopie électronique par M. Sachs et al, UMR144, Mécanismes moléculaires du transport intracellulaire.

¹⁶<http://www.irisa.fr/vista/Themes/ACI/MoDynCell5D/MODYNCELL5D.html>.

données générés, analyser ces informations à l'aide de descripteurs simples et à l'avenir de modèles qui devraient s'approcher de modèles physiques du fonctionnement. Ces modèles seront le reflet de la compréhension du système et leur confrontation aux données peut aider à préciser le rôle de différents régulateurs du trafic.

Notre travail concerne : d'une part, la représentation de l'information brute, dans notre cas sous la forme d'une segmentation spatio-temporelle des images, d'autre part le calcul de statistiques globales relatives à la dynamique observée. Ces statistiques : distributions de vitesses, ... peuvent alors être employées pour la comparaison de dynamiques. Le point fort de ce travail réside : d'une part, dans la prise en compte de la cohérence et spatiale et temporelle pour la détection des objets d'intérêt, d'autre part dans la tentative d'extraction d'informations relatives à la dynamique sans tracking individuel des objets. On peut noter que l'importance de cette prise en compte de la cohérence temporelle dans le cas de rapport signal à bruit faible est déjà mentionnée par Burgi *et al* [Burgi et al. (2000)] et rappelée dans l'article de revue récent de Meijering *et al* [Meijering et al. (2006)].

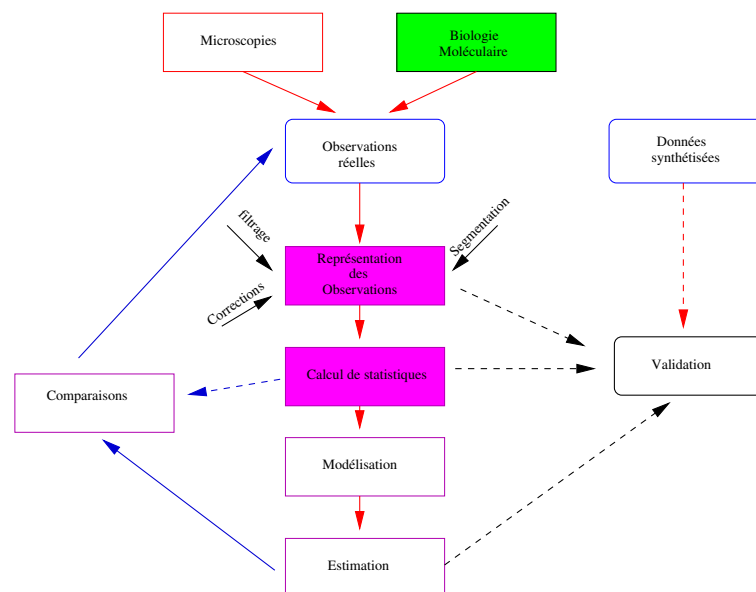


FIG. 1.8: Démarche mise en œuvre dans le projet MODYNCELL5D. En rouge, les aspects abordés par cette thèse.

1.4 Conclusion

Ce chapitre s'est voulu être un résumé scint du contexte et des motivations biologiques et médicales derrière ce travail de thèse qui n'est qu'un maillon dans une longue chaîne impliquant des équipes multi-disciplinaires de chercheurs et de doctorants. Nous avons alors essayé de présenter les grandes lignes du projet et nous y avons situé notre travail. Quelques notions de biologie ayant trait au sujet de thèse ont également été succinctement abordées. La suite du manuscrit est donc consacrée à la démarche d'analyse d'images que nous avons adoptée pour essayer de répondre à la problématique biologique présentée dans ce chapitre.

Chapitre 2

Prétraitement spatio-temporel d'une séquence de vidéo-microscopie

Dans la perspective d'une segmentation spatio-temporelle efficace en vidéo-microscopie de fluorescence, il est important de stabiliser le volume spatio-temporel et d'en réduire les artefacts d'acquisition qui sont nombreux avant de procéder à une analyse d'image quantitative de ce dernier. Nous nous intéressons dans ce chapitre à la réduction de deux types de distorsions importantes en vidéo-microscopie de fluorescence $3D + T$ à savoir le photoblanchiment et le bruit de Poisson. Aussi, dans le but de ne garder du volume spatio-temporel que l'information dynamique relative au trafic intra-cellulaire de vésicules, et par conséquent s'affranchir du problème de choix du seuil pour la détection du mouvement que nous traitons dans le chapitre 3, après alors avoir corrigé les profils temporels de l'intensité de fluorescence dans la séquence de vidéo-microscopie et stabilisé sa variance, nous réalisons la soustraction de sa partie déterministe quasi-stable (i.e. le fond qui correspond aux régions faiblement contrastées de la séquence et les structures quasi-stables qui correspondent aux régions fortement contrastées de la séquence).

2.1 Introduction

L'acquisition de séquences d'images *in vivo* ou *in vitro* en vidéo-microscopie de fluorescence (champ-large, confocal, FRAP, FLIM, FRET, etc...) est un processus assez complexe et produit des séquences d'images entachées de plusieurs types de distorsions. En effet, en plus de discrétisations spatiale (pixellisation) et temporelle (échantillonnage temporel) de la scène biologique observée (pour réaliser des images numérique) qui introduisent déjà une certaine distorsion dans les images reconstituées de la scène [Shanon (1949)], plusieurs autres distorsions existent en vidéo-microscopie de fluorescence. Ces distorsions sont dues à plusieurs facteurs liés à la technologie du dispositif d'acquisition utilisée, à la durée d'acquisition, à la durée d'exposition du spécimen biologique, au fluorophore utilisé, mais aussi aux propriétés biochimiques du specimen biologique, etc... Ici, nous ne considérons que le bruit statistique principalement dû au comptage de photons d'une part, et la perte de fluorescence au cours du temps (photo-blanchiment) d'autre part.

Le photo-blanchiment est un phénomène physico-chimique peu compris parmi la communauté des biophysiciens et des biochimistes et se manifeste par une décroissance au cours du temps de la quantité de fluorescence émise par un fluorophore ¹. Si on considère un voxel contrasté appartenant à un objet d'intérêt (i.e. portant le fluorophore utilisé tel qu'un vésicule ou le Golgi) dans une séquence de vidéo-microscopie et que l'on observe son profil temporel d'intensité, on note une certaine décroissance de sa fluorescence au cours du temps (*c.f.* figure 2.3). Une remarque similaire peut être également faite à propos de l'histogramme instantané des intensités de fluorescence qui a tendance à "s'affaïsser" au cours du temps (*c.f.* figure 2.1). Un tel phénomène peut être réduit pendant l'acquisition de la séquence de vidéo-microscopie en limitant la durée d'exposition et par là, la taille temporelle de la séquence, et/ou en utilisant des fluorophores ayant une durée de vie assez longue. Une correction numérique de l'intensité de fluorescence dans une séquence de vidéo-microscopie peut toutefois s'avérer nécessaire pour que l'analyse d'image quantitative qui suit soit efficace et évite d'interpréter les changements d'intensité dûs au photo-blanchiment comme des changements d'intensités dûs au mouvement d'objets fluorescents (vésicules et structures

¹Le lecteur intéressé peut consulter le livre de James B. Pawley [Pawley (1989)] pour une description technique plus détaillée du sujet.

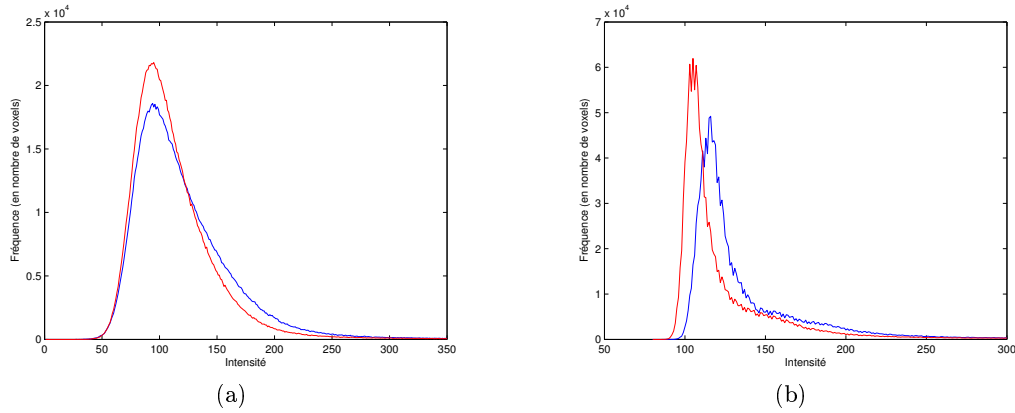


FIG. 2.1: Deux exemples montrant l'effet du photoblanchiment sur l'histogramme des intensités de deux séquences issues de la vidéo-microscopie de fluorescence 3D + T. En bleu, l'histogramme des intensités correspondant à l'instant $t = 1\text{sec.}$, et en rouge, l'histogramme des intensités de fluorescence correspondant à l'instant 120sec. de la séquence. Par souci de visualisation, la figure ne montre que la partie de l'histogramme des intensités correspondant à la plage $[0, 300]$ de niveaux de gris alors que le niveau d'intensité est codé sur 12 bits et peut donc atteindre quelques milliers.

quasi-stables) dans la séquence. Cela dit, pour qu'une telle correction du photo-blanchiment soit efficace, il est important qu'elle soit basée sur un modèle réaliste de décroissance de la fluorescence. Bien que ce problème ait été intensément étudié dans le cas de volumes 3D fixes pour corriger la perte de fluorescence dans la dimension "z" d'un volume issu de la microscopie 3D (Kervran et al. 2004; Lee et al, 2005), peu de travaux ont été faits jusqu'à maintenant pour la correction de l'intensité dans la dimension temporelle dans les séquences de vidéo-microscopie. Ceci peut s'expliquer par le fait que le traitement de séquences d'images, que ce soit pour le filtrage, la segmentation ou le suivi, est souvent local et dans l'espace et dans le temps. Dans le temps, la plupart des algorithmes exploitent l'information disponible à quelques instants précédents (souvent un seul instant), et quelques instants suivants (souvent zéro ou un seul instant). Dès lors qu'on s'autorise à considérer un nombre variable et éventuellement grand d'instants, il est évident que la perte de la fluorescence due au photo-blanchiment peut avoir des conséquences sur le résultat final.

Un exemple de telles approches est l'approche décrite au chapitre 4 qui traite tous les instants d'un segment, ou si on le souhaite, de toute une séquence de vidéo-microscopie à la fois. Par conséquent, ce chapitre se veut être une étude de la correction du photo-

blanchiment dans une séquence de vidéo-microscopie $3D + T$. Nous ne traitons pas le cas général de la correction du photo-blanchiment qui tente de corriger l'intensité pour chaque voxel de la séquence qu'il soit mobile ou statique, mais nous nous restreignons dans notre étude à la correction de l'intensité pour les voxels quasi-stables de la séquence qui peuvent perdre de leur fluorescence au cours du temps. Le but d'une telle approche est de corriger le profil temporel d'intensité pour chaque voxel de la séquence afin de pouvoir interpréter les changements d'intensité dans la séquence comme étant dûs au mouvement d'objets fluorescents dans la séquence et non au photo-blanchiment. Une approche de détection des régions en mouvement dans une séquence de vidéo-microscopie en se basant sur le changement dans le profil d'intensité est décrite en détails au chapitre 4.

Nous présentons d'abord deux méthodes de correction d'intensité dans une séquence d'images qui sont utilisées dans certains domaines de traitement d'images et nous argumentons qu'elles ne sont pas adaptées à notre application. Nous proposons ensuite, dans un premier temps, le modèle exponentiel simple que nous tentons d'ajuster au profil temporel d'intensité de chaque voxel de la séquence. L'estimation alors des paramètres du modèle se fait de manière robuste en utilisant la "M-estimation". Dans un second temps, dans le but d'accélérer les calculs et de rajouter de la robustesse à la méthode d'estimation des paramètres du modèle, au lieu d'estimer ces derniers pour chaque voxel de la séquence individuellement, nous proposons alors de prendre en compte la cohérence spatiale dans une image et de les estimer sur des voisinages spatiaux résultant de la subdivision de l'image en des petits blocs d'une certaine taille. L'estimation des paramètres d'un tel modèle se fait alors par le biais de la "M-estimation". Une des motivations derrière l'utilisation de l'estimation robuste pour estimer les paramètres du modèle de déclin du profil d'intensité d'un voxel est d'abord liée au bruit important qui entache les données image issues de la vidéo-microscopie de fluorescence. Une deuxième motivation très importante est que le profil temporel d'intensité d'un voxel peut subir plusieurs distorsions importantes dues au passage de vésicules, et il est important de ne pas en tenir compte dans l'estimation des paramètres du modèle en laissant le soin à la méthode d'estimation robuste de les détecter comme des "outliers".

Quant au bruit caractérisant les données image issues de la vidéo-microscopie de

fluorescence, il est principalement Poissonnien. Il en résulte une variance du bruit qui est inhomogène spatio-temporellement dans la séquence, puisqu'elle dépend de la moyenne du signal. La stabilisation de la variance dans une séquence de vidéo-microscopie de fluorescence peut alors constituer donc une étape primordiale avant une analyse quantitative efficace. Cela nous permettra d'utiliser des algorithmes de traitement efficaces qui sont développés dans le cas de bruit gaussien. La méthode de stabilisation de la variance dans une séquence de vidéo-microscopie décrite dans la deuxième partie de ce chapitre est inspirée du travail de Boulanger [Boulanger et al. (2006)], et notre apport concerne l'étape de calcul des jeux de moyennes et de variances dans la séquence qui servent à estimer les paramètres de la transformation d'Anscombe généralisée.

Une fois que les profils temporels d'intensité de tous les voxels d'une séquence de vidéo-microscopie sont corrigés, et que la stabilisation de la variance est réalisée dans cette dernière, nous décrivons une méthode utilisant l'estimation robuste pour soustraire la composante déterministe quasi-stable de la séquence de vidéo-microscopie afin de n'en garder que l'information relative à la dynamique des vésicules dans la séquence.

Enfin, notre méthodologie est appuyée par quelques résultats obtenus aussi bien sur des données synthétiques que sur des données réelles issues de la vidéo-microscopie de fluorescence $3D + T$.

2.2 Description de quelques méthodes de correction de la perte de fluorescence dans une séquence de vidéo-microscopie

Nous décrivons dans cette section quelques points de vue pour réduire les effets du phénomène de photo-blanchiment dans une séquence de vidéo-microscopie dans le but de normaliser temporellement la moyenne du profil temporel d'intensité de chaque voxel de la séquence.

2.2.1 Correction du photo-blanchiment basée sur la moyenne instantanée des intensités

Une première idée plutôt simpliste mais souvent utilisée en vidéo-microscopie pour réduire l'effet du photo-blanchiment dans une séquence de vidéo-microscopie consiste à normaliser chaque image pour que la somme de toutes des intensités soit constant dans le temps. Un tel objectif peut être atteint en multipliant chaque "t"-ième image par un facteur de correction " $\rho(t)$ " donné par la formule

$$\rho(t) = \frac{\text{Moyenne des intensités de l'image à l'instant "1"}}{\text{Moyenne des intensités de l'image à l'instant "t"}}$$

Cette correction du photo-blanchiment a l'avantage d'être facile à mettre en œuvre mais

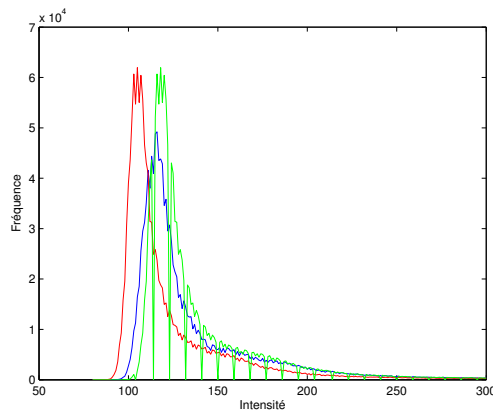


FIG. 2.2: En bleu, l'histogramme original des intensités d'une séquence de vidéo-microscopie considérée à l'instant $t = 1\text{sec.}$ En rouge, l'histogramme original des intensités de la même séquence de vidéo-microscopie considérée à l'instant $t = 120\text{sec.}$ En vert, l'histogramme des intensités après la correction d'intensité basée sur la moyenne instantanée des intensités. Les trous dans ce dernier sont dus à la discrétisation du niveau d'intensité après la multiplication de l'intensité de chaque voxel par le ratio de correction $\rho(t)$.

présente l'inconvénient de faire subir la même correction (linéaire) à tous les voxels de l'image quels que soient leur localisation spatiale et leur niveau de fluorescence. En effet, il s'avère que les voxels de l'image qui sont fortement contrastés ont tendance à subir une décroissance de la fluorescence qui est plus rapide que des voxels de l'image qui sont faiblement contrastés. La figure 2.2 montre le résultat assez décevant d'une telle correction sur une séquence de vidéo-microscopie après superposition de l'histogramme de l'image correspondant au dernier

instant de la séquence après la correction et l'histogramme de l'image correspondant au premier instant de la séquence. En effet, en plus de la distorsion importante (les creux multiples) de l'histogramme des intensités de l'image après la correction, on remarque les deux histogrammes demeurent très différents l'un de l'autre.

2.2.2 Correction du photo-blanchiment basée sur la déformation de l'histogramme instantané des intensités

Une deuxième façon de réduire les distorsions causées par le phénomène de photo-blanchiment serait la correction des histogrammes instantanés des images de la séquence par une technique qui opère directement sur les histogrammes instantanés plutôt que sur les voxels. L'idée alors consiste à déformer chaque histogramme instantané de la séquence de manière à le rapprocher d'un histogramme de référence (dans notre cas, ça pourrait bien être l'histogramme des intensités correspondant au premier instant de la séquence). Deux techniques appelées "spécification d'histogramme"² et "déformation d'histogramme"³ [Cox et al. (1995)] existent dans la littérature du traitement d'image pour rapprocher un histogramme des intensités d'une image donnée d'un autre histogramme de référence. Ces deux techniques, bien que couramment utilisées en photographie et en imagerie médicale pour augmenter le contraste d'une image et en améliorer la visualisation, ne sont pas très adaptées aux images issues de la vidéo-microscopie de fluorescence car elles ignorent la mécanique qu'il y a derrière les déformations des histogrammes instantanées. Par conséquent, elles risquent d'introduire d'autres types de distorsions qui peuvent compromettre l'analyse future d'une séquence de vidéo-microscopie.

2.2.3 Correction du photoblanchiment basée sur un modèle local de décroissance exponentielle de la fluorescence

La démarche que nous décrivons dans cette sous-section et dans la sous-section suivante ne constitue pas une méthode de correction du photo-blanchiment globalement dans toute la séquence de vidéo-microscopie. Elle constitue par contre une méthode de correction

²histogram specification en anglais

³histogram warping en anglais

des profils temporels d'intensité correspondant à tous les voxels de la séquence. Par conséquent, elle n'opère que sur le contenu quasi-statique de la séquence afin de le stabiliser et d'éviter que les fluctuations d'intensité qu'il subit mettent en défaut la méthode de détection de mouvement (ou plus exactement, la segmentation spatio-temporelle) que nous décrivons dans le chapitre 4. Elle est motivée par la constatation que les voxels appartenant surtout à des régions fortement contrastées de la séquence telles que les structures quasi-stables de la séquence (le Golgi et le réticulum endoplasmique (RE)) perdent de façon continue de la fluorescence au cours du temps. De telles structures quasi-stables comportent des objets fluorescents statiques ou (faiblement) dynamiques avec une grande densité, qui sont entièrement excités à chaque pas de temps par la lumière qu'émet le microscope en direction du spécimen biologique. Ce qui se passe alors, c'est qu'à chaque excitation, chaque objet fluorescent se trouvant à l'intérieur d'une structure quasi-stable verra son intensité de fluorescence diminuer d'un certain facteur. Et avec l'hypothèse de la constance de la moyenne d'intensité des objets fluorescents dans de telles structures, l'utilisation d'un modèle local de diminution de la fluorescence est bien justifiée. Il est alors important de stabiliser le profil temporel d'intensité de chaque voxel appartenant à une structure quasi-stable (sans forcément segmenter au préalable ces dernières dans la séquence bien sûr), afin d'éviter que leur perte continue de la fluorescence ne soit pas interprétée comme due au mouvement des objets fluorescents dans la séquence.

Cette même idée a déjà été utilisée dans [Boulangier (2006,2007)], mais l'auteur propose alors un modèle linéaire pour modéliser la décroissance de l'intensité de fluorescence d'un voxel. Donc il propose d'utiliser la régression linéaire robuste pour ajuster le profil d'intensité de chaque voxel du volume $3D$, et ceci de manière adaptative, dans ce sens que les paramètres de la régression linéaire robuste sont estimés sur un voisinage adaptatif d'un voxel donné en utilisant la règle de Lepski [Lepski (1990,1991)] de compromis biais-variance. Même si un tel modèle linéaire de décroissance de la fluorescence peut constituer une assez bonne approximation après quelques instants de la séquence, il a du mal à modéliser le phénomène du photo-blanchiment pour un voxel sur toute la plage temporelle de la séquence. En effet, il est connu dans la littérature qu'un tel phénomène est mieux décrit par un modèle exponentiel

de décroissance ⁴ et un tel constat a été confirmé par nos expériences. Conchello (1995) décrit une méthode de correction de l'intensité des différents plans $2D$ d'une image de microscopie $3D$ basée sur un modèle exponentiel de décroissance de la fluorescence. Pour cela, l'auteur considère que la constante de décroissance exponentielle de la fluorescence dans le volume $3D$ est homogène dans le volume $3D$, et propose donc une méthode basée sur l'algorithme "EM" (Expectation-Maximization) pour l'estimer globalement dans le volume $3D$. Notre méthode ne fait pas cette hypothèse de globalité des paramètres du modèle exponentiel de décroissance de la fluorescence, et son but ce n'est pas de corriger la perte de fluorescence dans toute la séquence, mais uniquement sur les profils temporels d'intensité. La méthode est inspirée du travail de [Boulangier (2006,2007)] en tentant d'ajuster un modèle $1D$ à la moyenne (robuste) du signal temporel correspondant à chaque voxel de la séquence. Nous traitons dans un premier temps le cas séparé, i.e. un modèle $1D$ est ajusté au profil temporel de chaque voxel de la séquence indépendamment des autres voxels. Dans un second temps, dans le but de rajouter de la robustesse à notre méthode et d'accélération des calculs, nous prenons en compte la cohérence spatiale et nous estimons les paramètres du modèle sur des voisinages spatiaux.

Nous supposons alors que le modèle de déclin du profil d'intensité de fluorescence s'écrit en chaque voxel x comme suit :

$$y_x(t) = \alpha_x \exp(-\lambda_x t) + \beta_x + \xi_t, t = 1, \dots, T \quad (2.1)$$

où α_x , λ_x et β_x sont des réels positifs ⁵ et ξ_t sont des gaussiennes centrées d'écart type σ_t . Nous proposons alors une méthode robuste de type "M-estimation" permettant leur estimation. L'utilisation d'un critère robuste pour l'estimation des paramètres α_x , λ_x et β_x est primordiale, pas seulement en raison de présence d'un bruit important dans les séquences de vidéo-microscopie, mais aussi pour éviter de prendre en compte dans leur estimation les instants correspondant à la visite d'un voxel par un objet fluorescent provoquant une déformation de son profil temporel d'intensité.

⁴Voir les pages 33-37 dans [Pawel (1989)] pour une discussion justifiant une telle approximation.

⁵Voir les pages 33-37 dans [Pawel et al. (1989)] pour une justification d'une telle approximation et le sens des coefficients α_x , λ_x et β_x du modèle.

Estimation robuste des paramètres du modèle

Les méthodes robustes de type “M-estimation” remplacent le critère quadratique par un critère robuste (i.e. moins sensible aux données aberrantes (*outlier*)) pour l'estimation des paramètres d'un modèle donné ⁶. Elles ont été intensivement étudiées dans la littérature [Huber et al.(1981), Hampel et al.(1986) ; Rousseeuw et al. (1987)] et ont trouvé un champ d'application très large en vision par ordinateur aussi bien pour des tâches de bas niveau en vision par ordinateur telles que la restauration d'images [Black et al (1996)], le recalage d'images [Nestares et al. (2001)], l'estimation du mouvement [Odobez et al (1995) , Cloutier et al. (1994)], le flot optique [Ye et al. (2000)], la stéréovision [Ruihua et al. (1993) , Zhang et al. (1994)], etc. . .

Dans notre cas, l'estimation robuste des paramètres α_x , λ_x et β_x du modèle (2.1) se fait en minimisant un critère robuste de la forme

$$\sum_{t=1}^T \rho\left(\frac{y_t - \alpha \exp(-\lambda t) - \beta}{\sigma_t}\right) \quad (2.2)$$

où $\rho(\cdot)$ est appelée fonction d'influence qui, contrairement à la fonction puissance carrée, limite l'influence des données s'écartant trop du modèle de données supposé (ici le modèle (2.1)), et σ_t est l'écart-type correspondant la t -ième observation $y_x(t)$. Le bon choix *a priori* d'une fonction robuste dépend de la nature du bruit sous-jacent aux données. Un descriptif du principe de la “M-estimation” est disponible dans l'annexe A.

Notre choix s'est porté sur la “Biweight de Tukey” ⁷ avec $c = 4.685$ en supposant un bruit gaussien⁸ qui a montré son efficacité et sa supériorité sur certains autres “M-estimateurs” pour tous les problèmes d'estimation robuste que nous considérons dans ce chapitre.

En écrivant la condition nécessaire d'optimalité (voir la formule A.3 de l'annexe A) pour le problème d'estimation robuste des paramètres α , λ et β du modèle de décroissance expo-

⁶Voir l'annexe A pour une présentation du principe de la “M-estimation”

⁷Voir l'annexe A

⁸Telle hypothèse n'est pas tout à fait vraie en vidéo-microscopie car le bruit est essentiellement Poissonien, mais ça n'empêche que ça constitue une bonne approximation pour traiter nos données.

entielle de la fluorescence, on trouve

$$\begin{aligned} \sum_{t=1}^T \frac{\exp(-\lambda t)}{\sigma_t} \rho' \left(\frac{y_t - \alpha \exp(-\lambda t) - \beta}{\sigma_t} \right) &= 0 \\ \sum_{t=1}^T \frac{t \exp(-\lambda t)}{\sigma_t} \rho' \left(\frac{y_t - \alpha \exp(-\lambda t) - \beta}{\sigma_t} \right) &= 0 \\ \sum_{t=1}^T \frac{1}{\sigma_t} \rho' \left(\frac{y_t - \alpha \exp(-\lambda t)}{\sigma_t} \right) &= 0 \end{aligned}$$

Supposant une variance homogène et posant :

$$w_t = \frac{\rho' \left(\frac{y_t - \alpha \exp(-\lambda t) - \beta}{\sigma} \right) \exp(-\lambda t)}{\left(\frac{y_t - \alpha \exp(-\lambda t) - \beta}{\sigma} \right)} = w \left(\frac{y_t - \alpha \exp(-\lambda t) - \beta}{\sigma} \right) \exp(-\lambda t) \quad (2.3)$$

où $w(\cdot)$ est la fonction poids correspondant à la fonction d'influence $\rho(\cdot)$, on obtient alors le schéma itératif suivant qui permet l'estimation des paramètres α , λ et β du modèle :

$$\begin{aligned} \sum_{t=1}^T w_t^{k-1} (y_t - \alpha^k \exp(-\lambda^k t) - \beta^k) &= 0 \\ \sum_{t=1}^T w_t^{k-1} t (y_t - \alpha^k \exp(-\lambda^k t) - \beta^k) &= 0 \\ \sum_{t=1}^T w_t^{k-1} \exp(\lambda^k t) (y_t - \alpha^k \exp(-\lambda^k t) - \beta^k) &= 0 \end{aligned} \quad (2.4)$$

où $w_t^{k-1}, \alpha^k, \lambda^k$ et β^k sont (respectivement) les estimations de w_t, α, λ et β à l'itération "k". Un tel schéma itératif a donc bien besoin d'une estimation initiale $w_t^0, t = 1, \dots, T$ des poids $w_t, t = 1, \dots, T$. Ainsi, les mises à jour des estimations de α, λ et β se font en résolvant, à chaque itération "k", le système d'équations (2.4), et celles des poids w_t se font à l'itération courante à l'aide de la formule (2.3) comme suit :

$$w_t^{k-1} = w \left(\frac{y_t^{k-1} - \alpha^{k-1} \exp(-\lambda^{k-1} t) - \beta^{k-1}}{\sigma} \right) \exp(-\lambda^{k-1} t)$$

Pour cela, on a donc besoin d'une estimation initiale α^0, λ^0 et β^0 de α, λ et β respectivement. Celle-ci doit être calculée automatiquement et doit être suffisamment proche de la solution optimale pour que le schéma itératif précédent puisse converger vers la solution recherchée.

Pour trouver des bonnes estimations initiales α^0 , λ^0 et β^0 de α , λ et β respectivement, on procède alors comme suit. Notant d'abord que β correspond à l'espérance du signal quand "t" devient suffisamment grand. Cette valeur est atteinte pratiquement après quelques instants de la séquence, ce qui fait que pour la majorité des instants, l'espérance de l'intensité du signal est presque égale à β . Donc, on se propose de calculer l'estimation initiale β^0 de β comme une moyenne robuste des observations $y_t, t = 1, \dots, T$. L'estimation initiale β^0 de β minimise donc le critère robuste :

$$\sum_{t=1}^T \rho\left(\frac{y_t - \beta^0}{\delta}\right)$$

où δ est l'écart-type des données $y_t, t = 1, \dots, T$.

Cette estimation se calcule à l'aide du schéma itératif suivant :

$$(\beta^0)^k = \frac{\sum_{t=1}^T (w_\beta)_t^{k-1} y_t}{\sum_{t=1}^T (w_\beta)_t^{k-1}}$$

En prenant comme estimation initiale de β^0 la médiane des observations $y_t, t = 1, \dots, T$, la mise à jour des poids à l'itération courante se fait alors à l'aide de la formule suivante :

$$(w_\beta)_t^{k-1} = w\left(\frac{y_t - (\beta^0)^{k-1}}{\delta}\right)$$

Une fois que β^0 est estimé, il reste à estimer α^0 et λ^0 . On procède alors comme suit. On pose

$$z_t = \log(|y_t - \beta^0|)$$

Les z_t devraient alors suivre approximativement un modèle linéaire de la forme :

$$z_t = \log(|\alpha^0 \exp(-\lambda^0 t)|) = -\lambda^0 t + \log(\alpha^0) + \nu_t$$

où ν_t est le bruit résultant de l'approximation des z_t par un modèle linéaire avec un bruit additif.

Les estimations α^0 et λ^0 peuvent alors être obtenues en ajustant aux observations $z_t, t = 1, \dots, T$ un modèle linéaire robuste. Maintenant qu'on dispose de α^0 , λ^0 et β^0 , on revient à l'estimation de α , λ et β du modèle (2.1). Nous avons conçu une méthode rapide pour résoudre à chaque itération le système (2.4) en démarrant de l'estimation initiale α^0 , λ^0 et β^0 précédemment calculée. En effet, réécrivant le d'abord comme suit :

$$\alpha^k = \frac{\sum_{t=1}^T w_t^{k-1} (y_t - \beta^k)}{\sum_{t=1}^T w_t^{k-1} \exp(-\lambda^k t)}$$

$$\sum_{t=1}^T w_t^{k-1} t \alpha^k \exp(-\lambda^k t) = \sum_{t=1}^T w_t^{k-1} t (y_t - \beta^k)$$

$$\beta^k = \frac{\sum_{t=1}^T w_t^{k-1} \exp(\lambda^k t) (y_t - \alpha^k \exp(-\lambda^k t))}{\sum_{t=1}^T w_t^{k-1} \exp(\lambda^k t)}$$

Un tel système peut s'avérer difficile à résoudre, on adopte alors un autre schéma itératif

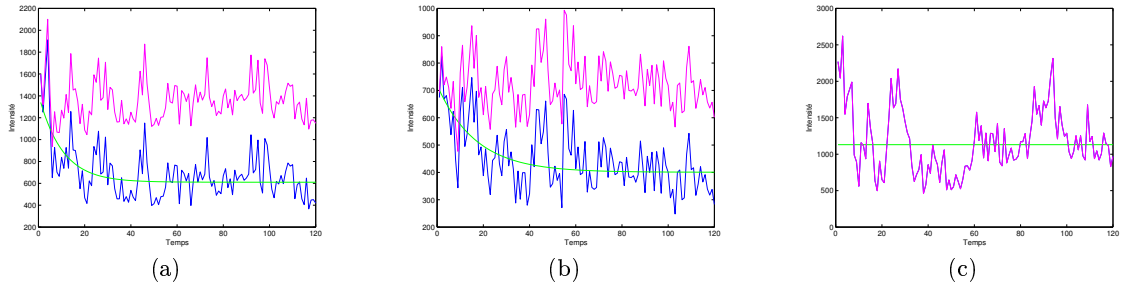


FIG. 2.3: La figure montre trois signaux voisins (27-voisinage) issus de la vidéo-microscopie 3D + T pris de la région de l'appareil de Golgi et auxquels on a tenté d'ajuster des modèles exponentiels robustes en considérant un modèle par voxel. En bleu le signal original, en vert le modèle exponentiel robuste ajusté, et en magenta le signal corrigé. Les modèles exponentiels ajustés sont : (a) $y_x(t) = 806.1381 \exp(-0.0993 t) + 610.4265$; (b) $y_x(t) = 318.8857 \exp(-0.0601 t) + 400.5825$; (c) $y_x(t) = 1113.8344$. Comme la méthode a estimé le terme exponentiel du modèle de correction d'intensité à zéro, donc le signal corrigé et le signal original se coïncident.

qui lui est voisin et qui converge vers la même solution finale ⁹ :

$$\alpha^k = \frac{\sum_{t=1}^T w_t^{k-1} (y_t - \beta^{k-1})}{\sum_{t=1}^T w_t^{k-1} \exp(-\lambda^{k-1} t)}$$

$$\sum_{t=1}^T w_t^{k-1} t \alpha^k \exp(-\lambda^k t) = \sum_{t=1}^T w_t^{k-1} t (y_t - \beta^{k-1})$$

$$\beta^k = \frac{\sum_{t=1}^T w_t^{k-1} \exp(\lambda^k t) (y_t - \alpha^k \exp(-\lambda^k t))}{\sum_{t=1}^T w_t^{k-1} \exp(\lambda^k t)}$$

La première équation de ce système d'équations fournit un schéma de mise à jour de l'estimation de α à l'itération "k" à partir des estimations de λ et β à l'itération "k-1".

La deuxième équation donne le schéma de mise à jour de l'estimation λ à l'itération "k" à

⁹pourvu que l'estimation initiale soit bonne bien entendu, ce qui est normalement le cas ici avec notre estimation initiale d'après les expériences numériques que nous avons réalisées.

partir des estimations courantes de α et de β à l'itération "k-1". Soulignant que la formule de mise à jour de λ suivante :

$$\sum_{t=1}^T w_t^{k-1} t \alpha^k \exp(-\lambda^k t) = \sum_{t=1}^T w_t^{k-1} t (y_t - \beta^{k-1})$$

n'est donnée que de façon implicite. Afin de pouvoir alors calculer la valeur de λ^k , on définit la fonction $\phi(\cdot)$ comme suit :

$$\phi(x) = \left(\sum_{t=1}^T w_t^{k-1} t \alpha^k \exp(-x t) - \sum_{t=1}^T w_t^{k-1} t (y_t - \beta^{k-1}) \right)^2$$

On remarque facilement que $\phi(\cdot)$ est une fonction strictement convexe et atteint son minimum unique pour $x = \lambda^k$. La recherche de λ^k revient alors à la recherche du minimum de la fonction $\phi(\cdot)$ à l'aide d'un algorithme d'optimisation unidimensionnelle tel que l'algorithme de la section d'or (golden search). Ce dernier a la propriété de converger vite vers la solution recherchée avec la précision souhaitée et ceci indépendamment de la forme de la fonction à optimiser du moment qu'elle est convexe.

La troisième équation donne la formule de mise à jour de l'estimation de β à l'itération "k" à partir des estimations courantes de α et de λ .

Enfin, la formule de correction de l'intensité en un voxel x donné s'effectue comme suit (si l'instant de référence est $t_0 \equiv 0$) :

$$y_x^c(t) := y_x(t) + \Delta y_x(t)$$

où $\Delta y_x(t)$ est la perte de fluorescence à l'instant "t" pour le voxel x et elle est donnée par la formule suivante :

$$\Delta y_x(t) = \alpha_x (1 - \exp(-\lambda_x t))$$

Quelques résultats sur des signaux qu'on a échantillonnés d'une séquence de vidéo-microscopie sont représentés en figure 2.3.

2.2.4 Correction du photoblanchiment basée sur un modèle de décroissance exponentielle avec prise en compte de la cohérence spatiale

Le modèle de correction exponentielle locale du photoblanchiment décrit dans la sous-section précédente considère chaque voxel indépendamment de son voisinage spatial.

Dans le but de rajouter de la robustesse à notre modèle d'une part, et d'accélérer des calculs d'autre part (puisque, comme on le verra, la mise à jour des paramètres se fera simultanément pour plusieurs voxels voisins), nous pouvons considérer que le taux de décroissance exponentielle de la fluorescence λ est le même pour tous les voxels d'un même petit voisinage spatial qui subissent une décroissance de la fluorescence (par exemple, considérer des voisinages de taille $3 \times 3 \times 3$), tandis que les paramètres α et β dépendent de chaque voxel de la séquence. Une telle hypothèse est bien justifiée en raison de la cohérence spatiale de l'intensité de fluorescence dans des petits voisinages spatiaux ¹⁰. Un tel modèle peut donc s'écrire comme suit :

$$y_i(t) = \alpha_i \exp(-\lambda t) + \beta_i + \xi_{i,t}, \quad t = 1, \dots, T; i = 1, \dots, N \quad (2.5)$$

où i désigne l'indice lexicographique du voxel dans le bloc (voisinage spatial) considéré, N est la taille du voisinage, et $\xi_{i,t}$ sont des gaussiennes centrées d'écart type $\sigma_{i,t}$. Un tel modèle est sans perte de généralité si une partie seulement des voxels d'un voisinage subissent une décroissance de la fluorescence et d'autres non. En effet, pour un tel cas de figure, on peut estimer une même valeur de λ pour tous les voxels, et on estime $\alpha_i = 0$ pour tous les voxels ne subissant pas une telle décroissance de la fluorescence et on obtient une erreur d'ajustement minimale (pourvu que le processus ait convergé vers les bonnes valeurs des paramètres bien entendu).

Estimation robuste des paramètres du modèle

L'estimation des paramètres α_i , λ et β_i , ($i = 1, \dots, N$) revient alors à minimiser un critère robuste de type "M-estimateur" de la forme :

$$\sum_{i=1}^N \sum_{t=1}^T \rho\left(\frac{y_i(t) - \alpha_i \exp(-\lambda t) - \beta_i}{\sigma_{i,t}}\right) \quad (2.6)$$

où $\sigma_{i,t}$ sont les écarts-types des observations $y_{i,t}$. On suppose que $\sigma_{i,t} \equiv \sigma_i$ (i.e. on suppose un seul écart-type par voxel pour pouvoir l'estimer).

¹⁰Il y a bien évidemment des exceptions à cela. Par exemple, dans une région frontière entre une structure quasi-stable qui est fortement contrastée (le Golgi par exemple) et le fond. Les voxels qui appartiennent donc à la structure quasi-stable subissent une décroissance importante de la fluorescence, tandis que les voxels qui appartiennent au fond subissent une décroissance de la fluorescence qui est négligeable. On reviendra plus tard à un tel et on verra comment on peut l'adresser ce façon élégante

Comme dans la sous-section précédente, on écrit les équations d'optimalité et on dérive le schéma itératif d'estimation des paramètres α_i , λ et β_i comme suit :

$$\begin{aligned}
\alpha_i^k &= \frac{\sum_{t=1}^T w_{i,t}^{k-1} (y_i(t) - \beta_i^{k-1})}{\sum_{t=1}^T w_{i,t}^{k-1} \exp(-\lambda^{k-1}t)} \\
\sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t \alpha_i^k \exp(-\lambda^k t) &= \sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t (y_i(t) - \beta_i^{k-1}) \\
\beta_i^k &= \frac{\sum_{t=1}^T w_{i,t}^{k-1} \exp(\lambda^k t) (y_i(t) - \alpha_i^k \exp(-\lambda^k t))}{\sum_{t=1}^T w_{i,t}^{k-1} \exp(\lambda^k t)}
\end{aligned} \tag{2.7}$$

où

$$w_{i,t}^{k-1} = w \left(\frac{y_i(t) - \alpha_i^{k-1} \exp(-\lambda^{k-1}t) - \beta_i^{k-1}}{\sigma_i} \right) \exp(-\lambda^{k-1}t)$$

Pour ce qui est de la mise à jour de λ , celle-ci est donnée de façon implicite par l'équation

$$\sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t \alpha_i^k \exp(-\lambda^k t) = \sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t (y_i(t) - \beta_i^{k-1})$$

ce qui revient (comme on a déjà eu l'occasion de l'expliquer plus haut) à minimiser la fonction unidimensionnelle $\psi(\cdot)$ suivante :

$$\psi(x) = \left(\sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t \alpha_i^k \exp(-x t) - \sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t (y_i(t) - \beta_i^{k-1}) \right)^2$$

en utilisant un algorithme standard de minimisation de fonctions unidimensionnelles strictement convexes.

Malheureusement, le schéma itératif (2.7) qui utilise à la fois l'information spatiale et l'information temporelle du signal spatio-temporel y peut donner des estimations finales de α_i , β_i et λ qui sont très biaisées. Cette remarque a été confirmée par les exemples numériques qu'on a réalisés sur des données synthétiques de simulation de signaux issus d'un même voisinage spatial et se comportant de façon similaire aux signaux issus de la vidéo-microscopie de fluorescence. En effet, on a constaté que lorsque seulement une petite partie (disant la minorité) des signaux d'un même voisinage spatial subissent une décroissance de la fluorescence, alors le paramètre λ est souvent estimé à 0. Une raison à cela est que les

“M-estimateurs” sont des estimateurs qui convergent vers un maximum de la densité de probabilité qui est local à la solution initiale choisie (Comaniciu et al., 2002)¹¹. Par conséquent, si on fournit au schéma (2.7) une solution initiale qui s'écarte trop de la vraie solution (i.e. le minimum global de (2.6)), alors ce dernier se trouve coincé dans un maximum local de la densité de probabilité (un minimum local de (2.6)).

Notre idée alors pour forcer le schéma itératif (2.7) à s'affranchir des minima locaux de (2.6) est de le rendre moins sensible à l'estimation initiale des paramètres α_i , β_i et λ . Pour cela, on utilise deux propriétés des données dont on dispose : leur redondance spatio-temporelle et leur profil temporel décroissant.

Pour ce faire, soit $w(\cdot)$ la fonction poids correspondant à une certaine fonction d'influence robuste $\rho(\cdot)$. On prend $w(\cdot)$ conformément aux poids de Tukey (voir l'annexe A), et on définit $w^*(\cdot)$ sa fonction poids complémentaire comme suit :

$$w^* = \begin{cases} [1 - (\frac{e}{c})^2]^2, & |e| > c \\ 0, & |e| \leq c \end{cases}$$

où $c = 4.685$, ensuite, on se donne un paramètre de temps t^* (par exemple $t^* = \frac{T}{6}$)¹² et on modifie le schéma itératif (2.7) comme suit :

$$\begin{aligned} \alpha_i^k &= \frac{\sum_{t=1}^T w_{i,t}^{k-1} (y_i(t) - \beta_i^{k-1})}{\sum_{t=1}^T w_{i,t}^{k-1} \exp(-\lambda^{k-1}t)} \\ \sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t \alpha_i^k \exp(-\lambda^k t) &= \sum_{i=1}^N \sum_{t=1}^T w_{i,t}^{k-1} t (y_i(t) - \beta_i^{k-1}) \\ \beta_i^k &= \frac{\sum_{t=1}^T v_{i,t}^{k-1} \exp(\lambda^k t) (y_i(t) - \alpha_i^k \exp(-\lambda^k t))}{\sum_{t=1}^T v_{i,t}^{k-1} \exp(\lambda^k t)} \end{aligned} \tag{2.8}$$

où les poids $w_{i,t}$ et $v_{i,t}$ et l'ordre de mise à jour des estimations des α_i , β_i et λ sont définis comme suit :

1. D'abord, on estime β_i^0 pour tout $i = 1, \dots, N$ comme suit : $\beta_i^0 := \text{mediane}_{t=1, \dots, T} \{y_i(t)\}$
2. On pose $\lambda^0 = 0$

¹¹ou de façon équivalente, vers un minimum local de (2.6)

¹²Un tel paramètre dit à la méthode d'estimation robuste que la composante exponentielle se situe en grande partie aux instants avant t^* .

3. On calcule les poids initiaux $w_{i,t}^0$ et $v_{i,t}^0$ comme suit :

$$- w_{i,t}^0 := \begin{cases} w^* \left(\frac{y_i(t) - \beta_i^0}{\sigma_i} \right), \forall t \leq t^* \\ w \left(\frac{y_i(t) - \beta_i^0}{\sigma_i} \right), \forall t > t^* \end{cases}$$

$$- v_{i,t}^0 := w \left(\frac{y_i(t) - \beta_i^0}{\sigma_i} \right), \forall t$$

4. On calcule α_i^0 en utilisant la première formule du schéma itératif (2.8) comme suit :

$$\alpha_i^0 := \frac{\sum_{t=1}^T w_{i,t}^0 (y_i(t) - \beta_i^0)}{\sum_{t=1}^T w_{i,t}^0}$$

5. Puis, à chaque itération $k + 1$, on calcule les poids $w_{i,t}^k$ et $v_{i,t}^k$ comme suit :

$$- w_{i,t}^k := \begin{cases} \max \left\{ w^* \left(\frac{y_i(t) - \alpha_i^k \exp(-\lambda^k t) - \beta_i^k}{\sigma_i} \right), w \left(\frac{y_i(t) - \alpha_i^k \exp(-\lambda^k t) - \beta_i^k}{\sigma_i} \right) \right\} \exp(-\lambda^k t), \forall t \leq t^* \\ w \left(\frac{y_i(t) - \alpha_i^k \exp(-\lambda^k t) - \beta_i^k}{\sigma_i} \right) \exp(-\lambda^k t), \forall t > t^* \end{cases}$$

$$- v_{i,t}^k := w \left(\frac{y_i(t) - \alpha_i^k \exp(-\lambda^k t) - \beta_i^k}{\sigma_i} \right), \forall t$$

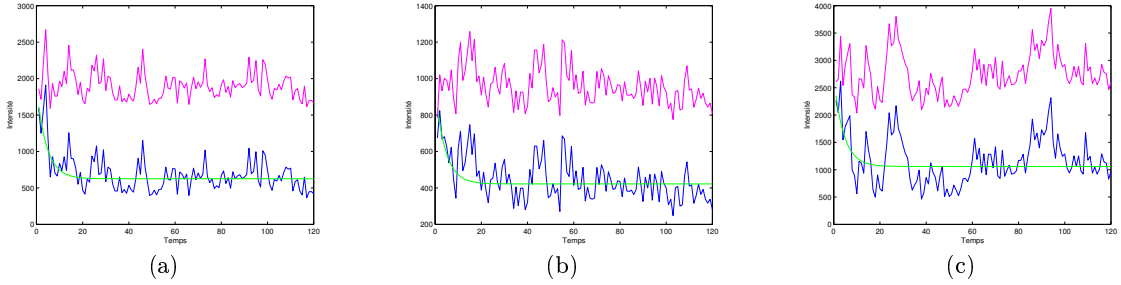


FIG. 2.4: La figure montre les mêmes signaux voisins de la figure 1.3. auxquels notre méthode a tenté d'ajuster des modèles exponentiels avec prise en compte du voisinage spatial (27 voisinage) et comment la prise en compte du voisinage spécial peut rajouter de la robustesse dans l'estimation des paramètres de décroissance de la fluorescence. En bleu le signal original, en vert le modèle exponentiel robuste ajusté, et en magenta le signal corrigé. Les modèles exponentiels ajustés sont : (a) $y_x(t) = 1249.7 \exp(-0.2333 t) + 626.0213$; (b) $y_x(t) = 527.1540 \exp(-0.2333 t) + 421.9424$; (c) $y_x(t) = 1634.5 \exp(-0.2333 t) + 1058.8$. Alors qu'avec un modèle de décroissance exponentielle local pour chaque voxel, on a estimé une décroissance nulle (voir figure 2.2.3), le fait de tenir compte de la cohérence spatiale a permis de rajouter plus de précision et de robustesse dans l'estimation des paramètres du modèle de décroissance exponentielle de la fluorescence.

Un tel schémas itératif, ainsi, d'ailleurs, que tous les schémas itératifs présentés dans ce chapitre, sont stoppés dès que la valeur absolue de la différence entre les estimations à deux itérations successives de chaque paramètre du modèle se situe en dessous d'un paramètre pré-spécifié. Enfin, les corrections des intensités s'effectuent comme précédemment à l'aide de la formule :

$$y_i^c(t) := y_i(t) + \Delta y_i(t)$$

où $\Delta y_i(t)$ est la perte de fluorescence à l'instant "t" pour le voxel i et elle est donnée par la formule

$$\Delta y_i(t) = \alpha_i(1 - \exp(-\lambda t))$$

La figure 2.4 montre sur trois signaux échantillonnés d'une séquence de vidéo-microscopie réelle comment la prise en compte de la cohérence spatiale permet de rajouter de la robustesse dans l'estimation des paramètres de décroissance exponentielle de la fluorescence au cours du temps localement pour chaque voxel d'une séquence de vidéo-microscopie.

Validation de l'approche

La validation de la méthode de correction d'intensité basée sur un modèle de décroissance exponentielle avec prise en compte de la cohérence spatiale qu'on a présentée plus haut a été réalisée d'abord en utilisant des signaux synthétiques. On fixe la variance σ^2 à une certaine valeur, puis on simule L voisinages spatiaux qui se rapprochent de la situation réelle en vidéo-microscopie de fluorescence comme suit :

- On fixe le nombre de signaux d'un même voisinage spatial à N signaux, et taille de chaque signal est T .
- Ensuite, pour chaque voisinage :
 - Le nombre de signaux n subissant une décroissance de la fluorescence suit une loi uniforme discrète $U[1, s]$, et les $N - n$ signaux restants ne subissent pas de décroissance de la fluorescence.
 - Le taux de décroissance de la fluorescence λ suit une loi uniforme continue $U[0, \frac{1}{3}]$.
 - Les paramètres α_i et β_i de chaque i -ième signal du voisinage suivent une loi uniforme continue $U[0, 2000]$

On essaie alors d'estimer le modèle (2.5) pour chaque voisinage en utilisant le schéma itératif d'estimation robuste (2.8). Soit alors un voisinage de signaux $(y_i(\cdot))_{i=1, \dots, N}$ simulés comme on vient de le décrire au dessus, et soient $(y_i^c(\cdot))_{i=1, \dots, N}$ les signaux reconstruits. On calcule alors l'erreur quadratique moyenne suivante :

$$\overline{\mathcal{R}}(\sigma^2) = \frac{1}{LNT} \sum_{l=1}^L \sum_{i=1}^N \sum_{t=1}^T \frac{(y_i(t) - y_i^c(t))^2}{\sigma^2}$$

On a obtenu alors pour $N = 27$, $T = 120$ sur $L = 1000$ jeux de signaux voisins et pour différentes valeurs de σ les différentes valeurs de l'erreur quadratique moyenne $\overline{R}(\sigma^2)$ qui sont résumées dans le tableau suivant :

| σ | $\overline{R}(\sigma^2)$ |
|----------|--------------------------|
| 1 | 9.9855 |
| 2.5 | 2.2706 |
| 5 | 1.3192 |
| 10 | 1.1149 |
| 20 | 1.0747 |
| 30 | 1.0829 |
| 40 | 1.0757 |
| 50 | 1.0880 |
| 75 | 1.1049 |
| 100 | 1.2399 |

Ces résultats signifient que l'erreur ponctuelle (i.e. commise sur une donnée atomique) qu'on commet dans l'estimation du modèle (2.5) sur des données simulées est en gros cadrée autour de σ même lorsque les données sont très corrompues par du bruit ($\sigma \geq 50$). Quand σ est trop petit par rapport à la moyenne du signal, $\sigma \leq 2$ par exemple, dans le cas de nos données simulées, on trouve une erreur quadratique moyenne qui est plusieurs fois plus grande que σ . La raison à cela est que le risque se décompose comme la somme d'un terme de biais et d'un terme de variance, et quand σ est trop petit par rapport à la moyenne du signal, c'est donc le terme de biais qui l'emporte sur le terme de variance.

On a également validé visuellement l'algorithme sur des séquences de vidéo-microscopie de fluorescence réelles. D'abord, en réalisant plusieurs échantillonnages de signaux voisins dans des séquences de vidéo-microscopie et en remarquant la bonne qualité de l'ajustement par le modèle exponentiel qu'on a proposé. Cette constatation est par la suite confirmée par les résultats qu'on a obtenus en section après la sousstraction de la partie déterministe stable de la séquence (voir les résultats représentés dans les figures à la fin du chapitre).

2.3 Stabilisation de la variance dans une séquence de vidéo-microscopie

Le but des méthodes de stabilisation de la variance ou de gaussianisation d'un ensemble de données est d'homogénéiser sa variance (i.e. rendre sa variance indépendante

des données) afin de faciliter et de rendre plus efficace son analyse statistique par l'utilisation de méthodes développées dans le cas de bruit gaussien.

Une telle dépendance de la variance des données est bien connue en vidéo-microscopie de fluorescence. En effet, le bruit dans les séquences de vidéo-microscopie de fluorescence est essentiellement un bruit de comptage qui est mieux décrit par un processus de Poisson, auquel s'ajoutent d'autres types de bruit d'une importance moindre tels que le bruit électronique (bruit gaussien), et le bruit résultant de l'auto-fluorescence du fond de la cellule (bruit uniforme).

Le modèle généralement utilisé pour décrire les données image issues de la vidéo-microscopie de fluorescence est le suivant :

$$Y_i = \alpha_\gamma v_i + \xi_i + g_i, i = 1, \dots, N \quad (2.9)$$

avec v_i un processus de comptage (dû à la nature quantique des photons) qui obéit à un processus de Poisson de paramètre λ_i , α_γ étant le gain du capteur, ξ_i est un bruit gaussien de moyenne nulle et de variance σ_i^2 , et enfin, g_i est le niveau d'auto-fluorescence au voxel "i" qui est entaché d'un bruit uniforme mais dont l'importance est négligeable de sorte que g_i est traité comme une valeur déterministe.

Donc, pour stabiliser la variance dans toute la séquence de vidéo-microscopie, une transformation d'Anscombe généralisée [Murtagh et al.(1995)] est utilisée. Celle-ci s'exprime comme suit :

$$t(Y_i) = \frac{2}{\alpha_\gamma} \sqrt{\alpha_\gamma Y_i + \frac{3}{8} \alpha_\gamma^2 + \sigma_i^2} - \alpha_\gamma g_i, i = 1, \dots, N$$

Lorsque v_i tend vers l'infini (en pratique, pour $v_i \geq 30$), les données ainsi transformées $t(Y_i)_{i=1, \dots, N}$ suivent approximativement une loi normale de variance égale à 1.

Une méthode intéressante pour l'estimation de la transformation $t(\cdot)$ est décrite dans [Boulangier (2007)], et en voici le principe. Ecrivant d'abord les expressions de l'espérance et de la variance correspondant au modèle (2.9) et établissant la relation entre ces deux quantités. On a

$$\mathbb{E}(Y_i) = \alpha_\gamma \lambda_i + g_i$$

$$\mathbb{V}(Y_i) = \alpha_\gamma^2 \lambda_i + \sigma_i^2 = \alpha_\gamma \mathbb{E}(Y_i) + \sigma_i^2 - \alpha_\gamma g_i$$

en posant donc

$$a = \alpha_\gamma, \quad b_i = \sigma_i^2 - \alpha_\gamma g_i$$

on obtient la relation suivante entre l'espérance (E_i) et la variance (V_i) au point i :

$$V_i = aE_i + b_i$$

et la transformation d'Anscombe généralisée s'exprime en fonction de a et b_i comme suit :

$$t(Y_i) = \frac{2}{a} \sqrt{aY_i + \frac{3}{8}a^2 + b_i} \quad (2.10)$$

Donc l'idée de l'auteur de Boulanger (2007) c'est d'estimer les deux paramètres a et $b_i \equiv b$ (on suppose que b est le même dans toute la séquence) à l'aide d'une régression linéaire robuste basée sur un jeu de moyennes et un jeu de variances estimés à partir des données de la séquence de vidéo-microscopie.

2.3.1 Calcul des jeux de moyennes et de variances

Pour obtenir ces jeux de moyennes et variances empiriques, le même auteur suggère alors de diviser l'image en des petits blocs (l'auteur suggère d'utiliser une taille de $7 \times 7 \times 7$), et de calculer dans chaque bloc la moyenne et la variance empiriques. Le résultat sera alors un ensemble de couples de type (moyenne, variance), et une régression linéaire robuste est appliquée directement sur cet ensemble pour estimer les paramètres a et b de la transformation stabilisante (2.10).

Un inconvénient de cette méthode réside dans sa façon de calculer les jeux de moyennes et de variances empiriques. En effet, dans un bloc de taille donnée (par exemple $7 \times 7 \times 7$ comme le suggère l'auteur), il est fort probable qu'il n'existe pas qu'une unique moyenne et une unique variance pour les données. En effet, par exemple, dans les blocs de l'image comprenant à la fois une partie peu contrastée appartenant au fond de l'image et une partie contrastée appartenant à une structure quasi-stable de la cellule (telle que le Golgi) ou à une vésicule, il existe au moins deux moyennes et deux variances. Pour expliciter le problème, considérant un bloc de l'image dont les données proviennent de deux classes C_1 et C_2 avec les proportions λ et $1 - \lambda$ respectivement. La moyenne et la variance de la classe

C issue du mélange de C_1 et C_2 s'écrivent alors

$$\mathbb{E}(C) = \lambda\mathbb{E}(C_1) + (1 - \lambda)\mathbb{E}(C_2) , \quad V(C) = \lambda^2V(C_1) + (1 - \lambda)^2V(C_2)$$

Remplaçant $V(C_1)$ par $a\mathbb{E}(C_1) + b$ et $V(C_2)$ par $a\mathbb{E}(C_2) + b$ dans cette dernière expression, on obtient la relation

$$V(C) = \lambda^2(a\mathbb{E}(C_1) + b) + (1 - \lambda)^2(a\mathbb{E}(C_2) + b)$$

et on voit bien que $V(C) \neq a\mathbb{E}(C) + b$. Le problème est que de telles fausses estimations peuvent être tellement en fort taux dans les jeux de moyennes et de variances empiriques que la régression linéaire robuste peut donner des estimations assez biaisées des coefficients de régression de b sur a . Pour réduire ce biais, nous proposons alors de modifier cette procédure de calcul des jeux de moyennes et de variances empiriques en rajoutant une étape de classification préalable des données de chaque bloc de l'image. Cette étape de classification tente de classer (clusteriser) les voxels dans chaque bloc en un certain nombre de classes (ou clusters) ; chaque classe étant caractérisée par une moyenne et une variance. Nous proposons donc d'effectuer cette classification en utilisant l'algorithme "EM" (Expectation Maximisation) d'estimation de mélange de gaussiennes combiné avec le critère "MDL" (Minimum Description Length) (Bouman[[27]]) pour l'estimation du nombre de classes dans un bloc de l'image. Ce critère si le nombre de classes est égal à K s'écrit :

$$\text{MDL}_y(K, \theta) = -\log p_y(y/K, \theta) + \frac{3K - 1}{2} \log(N) \quad (2.11)$$

où $p_y(y/K, \theta)$ désigne la distribution de probabilité des données Y_i si le nombre de classes est K , θ étant le vecteur des paramètres correspondant, N est la taille des données (égale au nombre de voxels dans un bloc), et enfin $3K - 1$ est nombre total de paramètres à estimer qui coïncide avec la taille de θ puisque pour chaque classe C_k , on estime sa moyenne μ_k , sa variance σ_k^2 ainsi que sa proportion dans le mélange q_k ¹³. On suppose que $p_y(y/K, \theta)$ est une gaussienne multivariée qui s'écrit donc comme suit :

$$p_y(y/K, \theta) = \prod_{k=1}^K \prod_{i \in C_k} g\left(\frac{y_i - \mu_k}{\sigma_k}\right)$$

¹³Comme $\sum_{k=1}^K q_k = 1$, donc on a exactement $K - 1$ de telles proportions q_k à estimer ; ce qui fait qu'on a au total un nombre de $3K - 1$ paramètres à estimer.

où $g(\cdot)$ est une gaussienne standard, C_k désigne la k -ième classe, et $\theta = ((\mu_k, \sigma_k)_{k=1, \dots, K})$, avec μ_k et σ_k sont respectivement la moyenne et l'écart type de la classe C_k . Le but est alors de minimiser le critère (2.11) par rapport à K et θ . Ceci se fait en fixant le nombre de classes K à $1, 2, \dots, K_{max}$ et en estimant à chaque fois la moyenne et la variance de chacune des K classes considérées, et à chaque estimation des K classes, on évalue le critère MDL correspondant, et à la fin, on aura gardé les valeurs de K et θ réalisant le minimum du critère (2.11). Notons que l'utilisation du critère "MDL" apparaît ici comme une heuristique seulement. En effet, même si l'on en a démontré l'efficacité la consistance pour une variété de problèmes d'estimation, toutefois, à notre connaissance, un tel résultat de consistance n'a pas été démontré dans le cas de la classification. Toutefois, d'après Bouman [[27]], le critère "MDL" garantit généralement d'assez bons résultats pour l'estimation du nombre de composantes d'un mélange de gaussiennes.

2.3.2 Estimation robuste des coefficients a et b

L'estimation des coefficients de régression linéaire entre le jeu de moyennes $(E_i)_{i=1, \dots, N}$ et le jeu de variances $(V_i)_{i=1, \dots, N}$ décrits dans la sous-section précédente se fait à l'aide de la régression robuste de type "M-estimation". Ceci revient donc à la minimisation du critère robuste

$$\sum_{i=1}^n \rho\left(\frac{V_i - aE_i - b}{\sigma}\right)$$

où σ modélise l'écart type du bruit dans les jeux de moyennes et de variances. On écrit alors le schéma itératif permettant l'estimation des coefficients de régression a et b comme suit :

$$\begin{aligned} \sum_{i=1}^N w_i^{k-1} (V_i - a^k i - b^k) &= 0 \\ \sum_{i=1}^N w_i^{k-1} i (V_i - a^k i - b^k) &= 0 \end{aligned}$$

(2.12)

et on en dérive le système de deux équations permettant la mise à jour des estimations de a et b :

$$\begin{aligned} \left(\sum_{i=1}^N w_i^{k-1} i\right) a^k + \left(\sum_{i=1}^N w_i^{k-1}\right) b^k &= \sum_{i=1}^N w_i^{k-1} V_i \\ \left(\sum_{i=1}^N w_i^{k-1} i^2\right) a^k + \left(\sum_{i=1}^N w_i^{k-1} i\right) b^k &= \sum_{i=1}^N w_i^{k-1} i V_i \end{aligned} \quad (2.13)$$

et pour ne pas avoir à résoudre à chaque itération k un tel système linéaire aux inconnues a^k et b^k (et éviter le problème du déterminant nul qui pourrait se poser), on considère à la place le schéma suivant :

$$\begin{aligned} a^k &= \frac{\sum_{i=1}^N w_i^{k-1} V_i - \left(\sum_{i=1}^N w_i^{k-1}\right) b^{k-1}}{\sum_{i=1}^N w_i^{k-1} i} \\ b^k &= \frac{\sum_{i=1}^N w_i^{k-1} i V_i - \left(\sum_{i=1}^N w_i^{k-1} i^2\right) a^k}{\sum_{i=1}^N w_i^{k-1} i} \end{aligned} \quad (2.14)$$

où les poids w_i^{k-1} à l'itération courante sont donnés par la formule :

$$w_i^{k-1} = w\left(\frac{V_i - a^{k-1} i - b^{k-1}}{\sigma}\right)$$

avec $w(\cdot)$ la fonction poids correspondant à la fonction d'influence $\rho(\cdot)$. Comme le montre ce schéma itératif, on a bien besoin des estimations initiales a^0 et b^0 de a et b respectivement. Celles-ci peuvent donc être calculées à l'aide des formules suivantes :

$$a^0 = \operatorname{mediane}_{i=1, \dots, N} \left\{ \frac{V_i - \operatorname{mediane}_{j=1, \dots, N} \{V_j\}}{E_i - \operatorname{mediane}_{j=1, \dots, N} \{E_j\}} \right\}, \quad b^0 = \operatorname{mediane}_{i=1, \dots, N} \{V_i - a^0 E_i\}$$

La figure .26 montre les modèles de régression linéaire ajustés aux jeux de moyennes et de variances correspondant à deux séquences de vidéo-microscopie de fluorescence réelles avant la stabilisation de la variance, ainsi que les jeux de moyennes et de variances correspondant aux mêmes séquences mais après stabilisation de la variance.

2.4 Soustraction de la partie déterministe quasi-stable d'une séquence de vidéo-microscopie

Une fois la correction d'intensité évoquée en section 2.2.1 réalisée sur la séquence de vidéo-microscopie, et la variance stabilisée spatio-temporellement dans toute la séquence, la soustraction du fond, i.e. l'enlèvement de la partie déterministe quasi-statique de la séquence, peut alors se faire de manière efficace en calculant sur chaque voxel x une moyenne robuste μ_x^c qui est sensée être l'espérance de l'intensité du fond au voxel x ¹⁴, et en soustrayant cette dernière du signal $y_x^c = (y_x^c(t))_{t=1, \dots, T}$ du voxel x . Le calcul d'une telle moyenne robuste μ_x^c se fait donc en minimisant un critère robuste de la forme :

$$\sum_{t=1}^T \rho\left(\frac{y_x^c(t) - \mu_x^c}{\sigma_x^c}\right)$$

où σ_x^c est l'écart-type du bruit du signal y_x^c . On en dérive alors le schéma itératif suivant permettant l'estimation de μ_x^c :

$$(\mu_x^c)^k = \sum_{t=1}^T w_t^{k-1} y_x^c(t)$$

où les poids w_t^{k-1} sont donnés par la formule :

$$w_t^{k-1} = w\left(\frac{y_x^c(t) - (\mu_x^c)^{k-1}}{\sigma_x^c}\right), k = 1, 2, 3, \dots$$

et en démarrant d'une estimation initiale $(\mu_x^c)^0$ qui est calculée comme suit :

$$(\mu_x^c)^0 = \text{mediane}_{t=1, \dots, T} \{y_x^c(t)\}$$

Les figures 2.7, 2.8, 2.9 et 2.10 présentent les résultats obtenus sur deux séquences de vidéo-microscopie après stabilisation de leur variance.

2.5 Conclusion

Dans ce chapitre, nous nous sommes intéressés dans un premier temps à la réduction de quelques artefacts en vidéo-microscopie de fluorescence qui sont le photo-blanchiment

¹⁴On désigne ici par fond tout ce qui correspond à une partie statique de la séquence y compris les structures quasi-stationnaires de la cellule telles que le Golgi et les RE

et le bruit de Poisson. Le but était alors de produire un volume spatio-temporel en vidéo-microscopie pour lequel les fluctuations d'intensité ne sont pas dues à de tels artefacts mais à la dynamique étudiée. Un tel volume spatio-temporel s'apprête mieux que le volume spatio-temporel original à l'analyse quantitative spatio-temporelle que nous allons décrire dans le chapitre 4. Dans un second temps, nous nous sommes intéressés à la soustraction de la composante spatio-temporelle déterministe quasi-statique après avoir effectué les deux étapes précédentes sur une séquence de vidéo-microscopie, et ceci est dans le but de n'en garder que l'information dynamique à du bruit gaussien près. Une telle démarche nous affranchira après la segmentation spatio-temporelle d'une séquence de vidéo-microscopie décrite dans le chapitre 3 de la délicate procédure de choix du seuil à partir duquel on considèrera qu'une partie spatio-temporelle est statique ou est en mouvement.

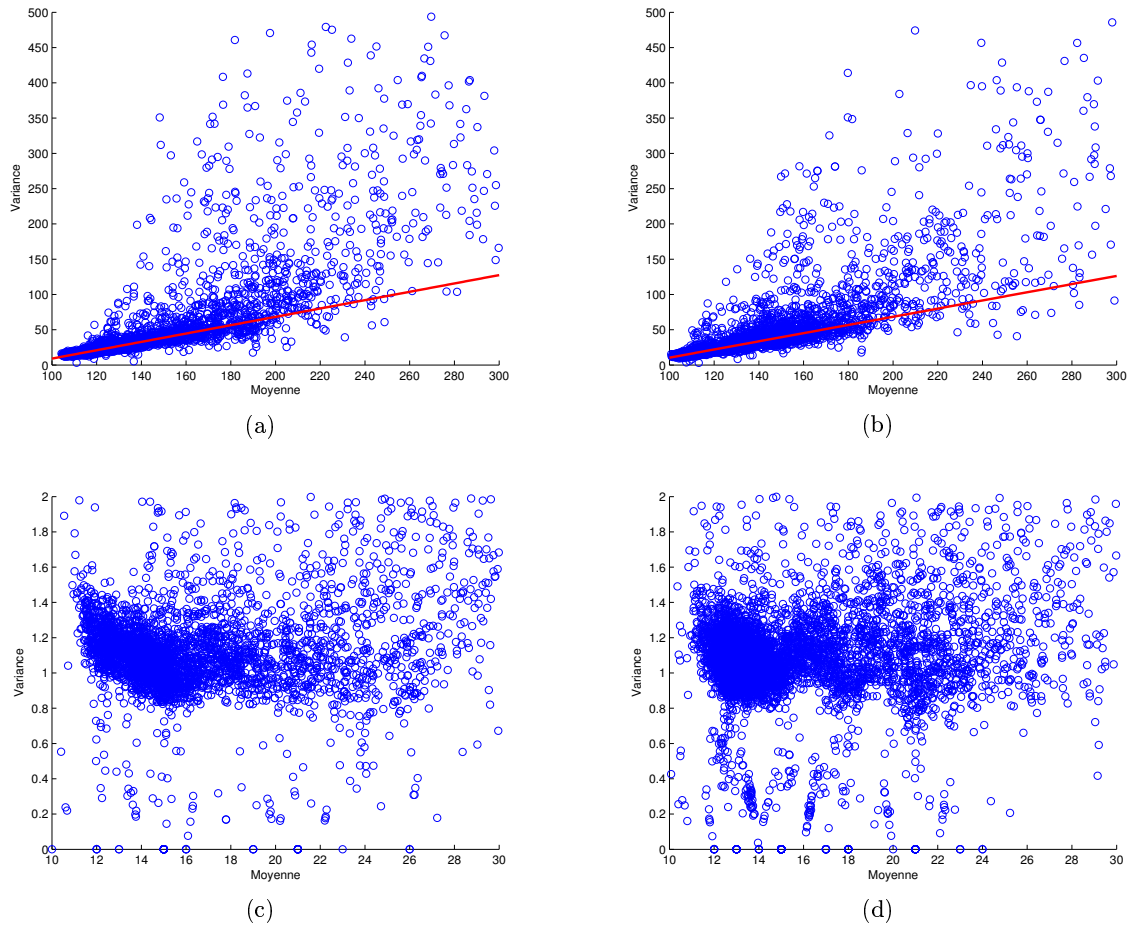
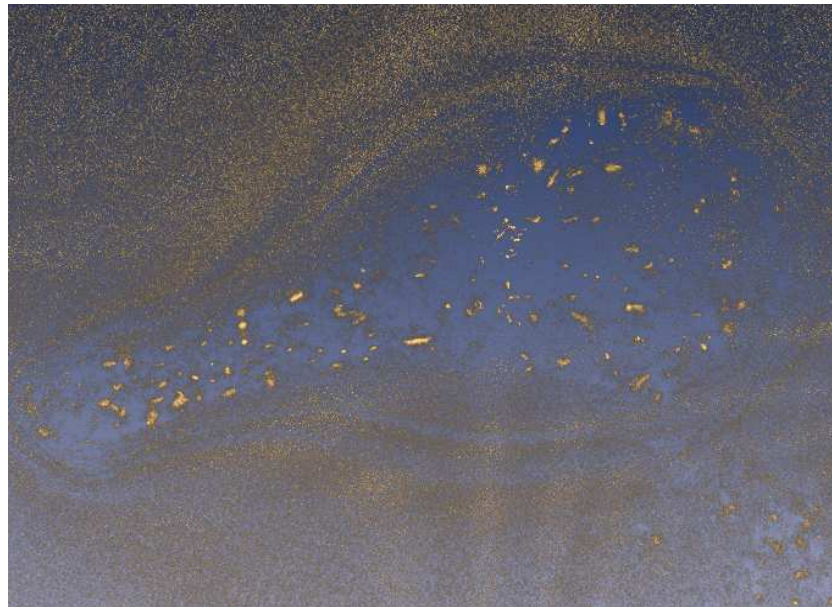


FIG. 2.5: (a)-(b) Régression linéaire robuste des jeux de moyennes et de variances pour deux séquences de vidéo-microscopie de fluorescence 3D + T. Les modèles de régression linéaire robuste estimés sont : (a) $V = 0.591677E - 50.0451$; (b) $V = 0.579375E - 47.5166$; (c)-(d) Stabilisation de la variance dans les deux séquences de vidéo-microscopie. Les deux figures montrent que la variance dans les deux séquences après transformation est distribuée approximativement autour de 1.



(a)

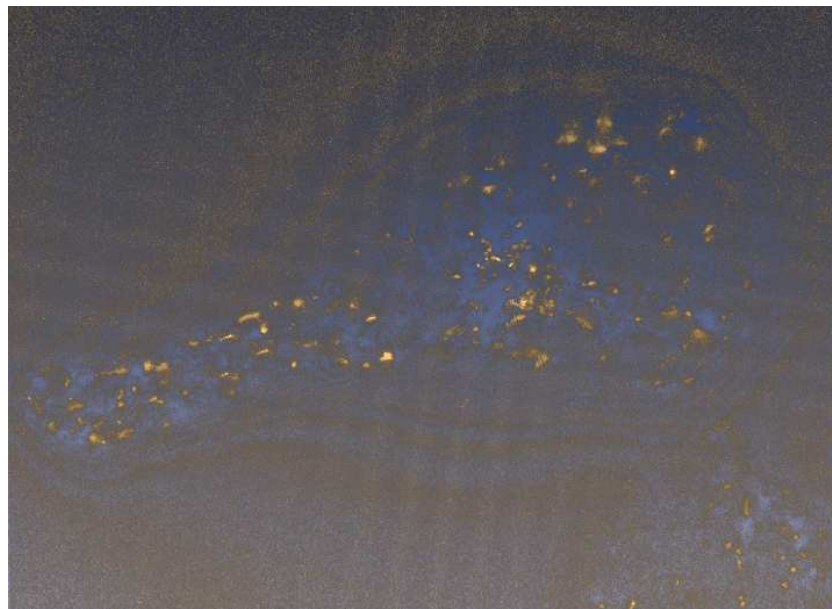


(b)

FIG. 2.6: (a) Rendu volumique (logiciel Amira) de la première image de taille $696 \times 520 \times 6$ d'une première séquence issue de la vidéo-microscopie de fluorescence 3D + T de 121 instants (secondes) ($\approx 525MO$). (b) La même image après correction d'intensité avec prise en compte de la cohérence spatiale, stabilisation de la variance et soustraction du fond.

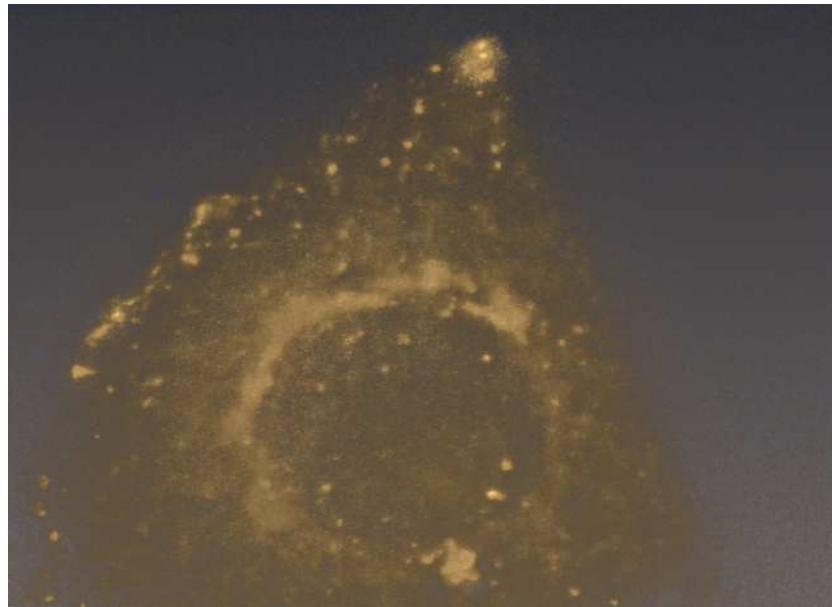


(a)

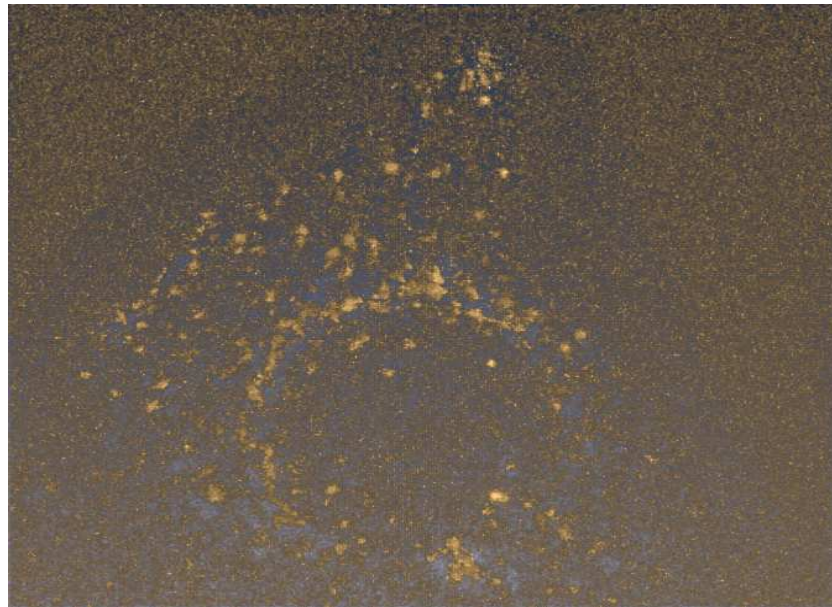


(b)

FIG. 2.7: (a) Rendu volumique de la dernière (121-ième) image de taille $696 \times 520 \times 6$ d'une première séquence issue de la vidéo-microscopie de fluorescence $3D + T$ de 121 instants (secondes) ($\approx 525MO$). (b) La même image après correction d'intensité avec prise en compte de la cohérence spatiale, stabilisation de la variance et soustraction du fond.

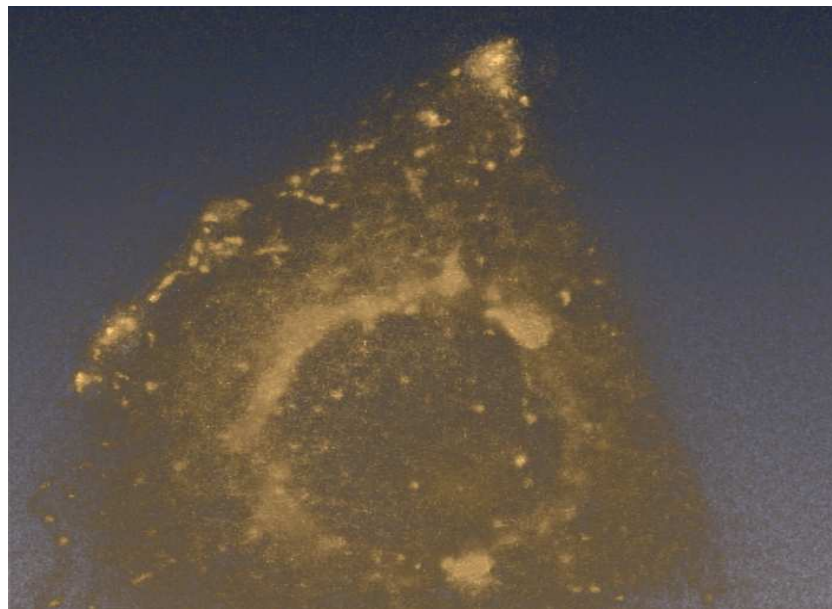


(a)

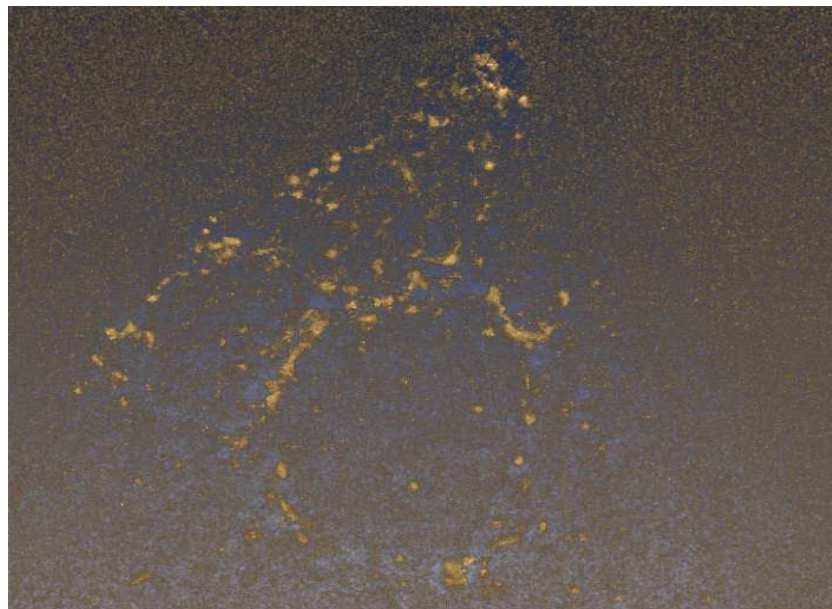


(b)

FIG. 2.8: (a) Rendu volumique de la première image de taille $395 \times 345 \times 10$ d'une seconde séquence issue de la vidéo-microscopie de fluorescence $3D + T$ de 121 instants (secondes) ($\approx 330MO$). (b) La même image après correction d'intensité avec prise en compte de la cohérence spatiale, stabilisation de la variance et soustraction du fond.



(a)



(b)

FIG. 2.9: (a) Rendu volumique de la dernière (121-ième) image de taille $395 \times 345 \times 10$ d'une deuxième séquence issue de la vidéo-microscopie de fluorescence $3D + T$ de 121 instants (secondes) ($\approx 330MO$). (b) La même image après correction d'intensité avec prise en compte de la cohérence spatiale, stabilisation de la variance et soustraction du fond.

Chapitre 3

La sélection simple de modèle et la sélection simultanée de modèles dans un cadre gaussien

Nous étudions dans ce chapitre la sélection de modèle dans un cadre gaussien basée sur la pénalisation d'un critère de type "moindres carrés", et nous nous appuyons pour cela sur une théorie de sélection de modèle très puissante due à Barron, Birgé & Massart (1999) et à Birgé & Massart (B & M) (2001, 2004, 2007). Nous nous inspirons dans cette étude des deux points de vue décrits dans [B & M (2001)] et [B & M (2004,2007)] auxquels et nous y proposons quelques modifications permettant de dériver des stratégies intéressantes de choix des constantes dont dépend le critère de sélection de modèle.

Une des motivations pour une telle étude est que les critères de sélection de modèle classiques tels que AIC [Akaike (1973)], BIC [Shwarz (1978)], MDL [Raissanen (1978)] ne permettent pas de prendre en compte la richesse d'une famille de modèles¹, et leur application pour choisir un modèle qui décrit au mieux des données observées résulte généralement en le choix d'un modèle les sur-ajustant. Au fait, un tel sur-ajustement des données est d'autant plus important que la famille de modèles est riche. C'est le cas par exemple pour

¹Par "richesse d'une famille de modèles", nous entendons la relation entre le nombre de modèles d'une famille de modèles et la taille des données qui se résume dans la quantité $\text{Card}_{\mathcal{M}}(\cdot)$ auquel on aura l'occasion de revenir dans la suite de ce chapitre.

les problèmes de sélection de modèle dont la taille de la famille de modèles dépend de façon exponentielle de la taille de l'ensemble des données [B & M (2001)].

Nous supposons que, dans un premier temps, les données pour lesquelles nous cherchons à choisir un modèle parmi une famille de modèles possibles proviennent toutes d'un même processus gaussien. Une difficulté des critères décrits dans [B & M (2001)] et [B & M (2004,2007)] réside dans le choix automatique des constantes dont dépend le critère de sélection de modèle. Nous proposons alors deux nouveaux critères pénalisés de sélection de modèle inspirés respectivement de [B & M (2001)] et [B & M (2004, 2007)], et nous décrivons une stratégie de choix automatique des constantes dont dépendent les deux critères. Une telle stratégie tente de minimiser uniformément pour tous les modèles la borne supérieure du risque correspondant à chaque critère. Nous comparons les deux critères par une telle stratégie de choix des constantes dans le critère de sélection de modèle et nous démontrons alors la supériorité du second critère par cette stratégie.

Une deuxième motivation très importante d'une telle étude est que les critères classiques ne fournissent pas un cadre théorique intéressant permettant leur généralisation au cas de la sélection simultanée de modèles, i.e. lorsque l'on dispose de plusieurs ensembles de données provenant de plusieurs processus indépendants et que l'on veut choisir un modèle pour chaque ensemble de données. La théorie de l'inférence simultanée nous suggère alors de corriger le critère pénalisé de choix de modèles si l'on veut éviter de trop sur-ajuster des données. Une telle situation survient forcément lorsque la taille des données traitées simultanément par sélection de modèle dépasse de loin leur dimension effective². Donc, dans un second temps de notre étude, nous étendons notre résultat dans le cas de la sélection simultanée de modèles, et nous verrons qu'une telle extension se fait sans grande difficulté.

Nous démontrons la faisabilité de notre approche à la fin de ce chapitre sur un exemple classique en sélection de modèle qui est l'estimation d'une fonction constante par morceaux, et au chapitre 4 lorsque nous traiterons de la segmentation spatio-temporelle d'une séquence de vidéo-microscopie dans le chapitre 3.

²cette idée sera mieux explicitée au chapitre 4 lorsque nous traiterons de la segmentation spatio-temporelle globale d'une séquence de vidéo-microscopie par aggrégation de plusieurs (dizaines de milliers) de segmentations spatio-temporelles locales

3.1 Introduction

Mettre au point une stratégie de choix d'un modèle qui ajuste bien des données sans ni trop les sur-ajuster ni trop les sous-ajuster est au coeur de la théorie de sélection de modèle. Pour introduire ce concept de sélection de modèle, nous présentons trois problèmes classiques très fréquemment rencontrés en traitement de signal et d'image qui sont l'ajustement d'un polynôme à un jeu de données, l'estimation d'une fonction constante par morceaux, et l'estimation des composantes d'un modèle de mélange.

3.1.1 Ajustement d'un polynôme à un jeu de données

Supposant que l'on dispose d'un jeu de n données observées d'un processus bruité et qu'on veuille ajuster ces données par un polynôme d'un degré inférieur ou égal à p . Alors, si l'on choisissait d'ajuster à ces données un modèle constant, l'erreur empirique serait très grossière, ce qui pourrait signifier que le modèle constant réalise un sous-ajustement des données. Par contre, si l'on choisissait de lui ajuster un polynôme de degré " p ", l'erreur empirique comise serait très petite, mais l'on risquerait alors de sur-ajuster les données. Au fait, ces deux choix extrêmes n'ont aucun intérêt en pratique car ils ne nous renseignent point sur la structure du processus qui a généré les données, ce qui peut être dans une étude la question centrale. La théorie de sélection de modèle nous permet donc de choisir le degré du polynôme qui ajuste au mieux de telles données et d'en estimer les paramètres.

3.1.2 Estimation d'une fonction constante par morceaux en utilisant la sélection de modèle

Soit un système caractérisé par un certain nombre d'états $\{E_i, i = 1, \dots, n\}$. Un tel système peut rester dans un état pendant un certain temps inconnu avant de basculer dans un autre état. Supposons maintenant que l'on dispose d'un appareil qui enregistre à chaque seconde pendant un temps T un signal décrivant l'état du système. En absence de bruit, l'appareil enregistre un même signal " C_i " (qu'on suppose inconnu) si le système est dans l'état " E_i " de sorte que le signal en l'absence de bruit est une fonction constante par morceaux (appelée parfois aussi histogramme). Bien évidemment, un tel signal est corrompu

par du bruit et le problème est d'estimer l'histogramme des états du système à partir du signal enregistré sachant que l'on ne connaît ni la localisation des états desquels est passé le système, ni leur nombre.

3.1.3 Estimation des composantes d'un modèle de mélange

Le problème est, étant donné un ensemble de données appartenant à K classes (K étant inconnu), comment estimer K ainsi que chaque classe (ou composante) de données parmi ces données observées. Un exemple bien classique est l'estimation des classes d'un mélange de données provenant de K gaussiennes, K étant inconnu et chacune de ces gaussiennes étant caractérisée par une moyenne et une matrice de variances-covariances inconnues. Un tel problème est généralement résolu avec un algorithme de classification (*EM* ou *k-means*), combiné avec un critère de sélection de modèle pour estimer le nombre de classes K .

3.1.4 Aperçu succinct sur la sélection de modèle basée sur un critère pénalisé

La sélection de modèle basée sur la pénalisation de la vraisemblance remonte aux travaux de Mallows et d'Akaike dans les années soixante dix qui ont proposé (respectivement) le critère de Mallows (Mallows' C_p) et le critère d'Akaike (AIC). Le critère Mallows' C_p [Mall1971] est une heuristique basée sur un estimateur sans biais du risque d'un estimateur sous l'hypothèse que le bruit est gaussien. Le critère AIC [Akaike (1971)] est une généralisation du Mallows' C_p dans le cas d'un bruit qui n'est pas nécessairement gaussien, basée sur la pénalisation de la vraisemblance. Notant que le critère AIC et le Mallows' coïncident dans le cas d'un bruit gaussien avec une variance connue. Ces deux travaux pionniers ont permis par la suite à beaucoup de chercheurs statisticiens et d'autres chercheurs dans des domaines relevant de l'ingénierie (statistique appliquée à la médecine et à la société, traitement de signal, traitement d'image et vision par ordinateur, etc...) d'étudier et d'écrire des centaines d'articles traitant de la sélection de modèle basée sur un critère pénalisé, aussi bien dans un cadre théorique plus ou moins général, que dans un cadre purement applicatif. Shwarz (1978) et Raissanen (1978) ont proposé respectivement le critère BIC (Bayesian Information

Criterion) et le critère MDL (Minimum Description Length) en donnant une justification dans un cadre Bayésien pour le premier, et dans un cadre de la théorie de l'information pour le second. Tandis que le Mallows' C_p et l'AIC ne font pas intervenir la taille des données dans le terme de pénalité, le BIC et le MDL dépendent linéairement du logarithme de la taille des données et leur consistance pour un certain nombre de problèmes a été démontrée. Notant que la nécessité de dépendance d'un critère pénalisé du logarithme de la taille des données a d'ailleurs été démontrée dans [B & M (2001)] pour éviter le problème de sur-ajustement. Sugiura (1978) et Hurvitch & Tsai (1989) ont proposé respectivement le c-AIC (consistent AIC) et le AICC (AIC corrected) pour des tailles d'échantillons qui sont trop petites par rapport au nombre de paramètres du modèle, Ronchetti (1982, 1985, 1994) a étudié la sélection de modèle robuste en présence de données aberrantes, Haughton (1988) a étendu le BIC pour le choix d'un modèle de courbe ajustant des données provenant d'une famille de distribution de probabilité exponentielle, Shi & Tsai (1998) et Burnham & Anderson (2002) ont proposé respectivement le c-QUAIC et QAIC (Quasi-AIC) pour la sélection de modèle dans le cas de données dispersées, Kanatani et al. ont proposé des versions géométriques des critères AIC et MDL et les ont appliquées pour adresser des problèmes de sélection de modèle rencontrés en vision par ordinateur tels que la stabilisation de la calibration d'une caméra en mouvement à partir d'images [Mastun (1999)], la stabilisation d'une mosaïque d'images [Kanatani (1999)], la segmentation du mouvement [Kanatani (2002)], etc ...

Les performances de ces différents critères de sélection de modèle ont fait l'objet de plusieurs études. Les conclusions sur la supériorité d'un critère pénalisé sur les autres a souvent dépendu du problème étudié et des critères auxquels il est comparé. Par exemple, Celeux & Soromenho (1999) puis McLachlan & Peel (2000) ont affirmé la supériorité du critère BIC sur le critère AIC pour le problème d'estimation du nombre de composantes dans un modèle de mélange. Castellan (1999) puis Lebarbier (2002) s'accordent également à dire que pour le problème de détection de ruptures dans les signaux $1D$, alors que la dimension du modèle choisie avec AIC explose, le critère BIC s'approche lui de la consistance et permet de sélectionner des modèles avec des dimensions plus raisonnables. Gheissari & Bab-Hadiashar (2003) ont montré en utilisant des données synthétiques que le critère SSC qu'ils proposent (Surface Selection Criterion) fait mieux que les autres (AIC, BIC, MDL,

etc. . .) pour le problème d'estimation du mouvement (flot optique) et de sélection d'un modèle de surface à des données $3D$ géométriques.

Malgré leur application avec succès à des problèmes de sélection de modèle relevant de plusieurs domaines d'ingénierie (statistique appliquée à la médecine, apprentissage, traitement du signal, traitement d'image et vision par ordinateur, etc. . .) avec éventuellement quelques adaptations au problème étudié, les critères de sélection de modèle traditionnels ne donnent aucune garantie de marcher sur de nouvelles classes de problèmes de sélection de modèle. En effet, la majorité de ces critères sont soit basés sur des idées heuristiques (comme c'est le cas pour le Mallows' C_p , l'AIC et le MDL), soit sur une approximation du logarithme de la vraisemblance et la maximisation de cette dernière (comme c'est le cas pour le BIC), et leur consistance n'a été démontrée que pour quelques problèmes particuliers. Barron, Birgé & Massart (1999) décrivent une théorie intéressante de sélection de modèle dans un cadre gaussien basée sur la pénalisation d'un critère de type "moindres carrés" qui peut être vue comme une extension du Mallows' C_p , et la théorie a été mise en œuvre avec succès par B & M (2001) où ils proposent un critère de sélection de modèle dans un cadre gaussien quand la variance est connue, qui permet de contrôler une certaine borne supérieure du risque dont ils donnent la formule explicite. Les auteurs utilisent au fait la théorie de l'inférence simultanée pour dériver une fonction de pénalité qui dépend de la richesse de la famille de modèles, ce que ne permettent pas les critères de sélection de modèles classiques. Malgré l'intrêrêt théorique que présente une telle théorie, son application pour résoudre des problèmes de sélection de modèle qui se posent en pratique a montré ses limites. En effet, en plus de la difficulté de pouvoir dériver une stratégie optimale de choix des constantes dans la fonction de pénalité qu'ils proposent, la borne supérieure du risque que garantit l'estimateur pénalisé correspondant est assez grossière et a tendance à sélectionner des modèles de petites dimensions par rapport aux vrais modèles ayant généré les données. Les mêmes auteurs, se basant alors sur leurs travaux antérieurs, proposent dans [B&M (2007)] un nouveau critère pénalisé constituant une amélioration significative de la borne supérieure du risque de l'estimateur pénalisé correspondant. Toutefois, la fonction de pénalité dépend d'un certain nombre de constantes dont la dérivation des valeurs optimales n'est pas une simple tâche.

Dans ce chapitre, nous nous basons sur les travaux de B & M (2001, 2004, 2007)

pour proposer deux nouveaux critères pénalisés non-asymptotiques dans un cadre gaussien quand la structure de variances-variances est connue. L'avantage des critères pénalisés que nous présentons dans ce chapitre est que, en plus de leur garantie d'une borne supérieure du risque qui est serrée, permettent une dérivation facile des valeurs optimales des constantes dans la fonction de pénalité³ indépendamment du problème de sélection de modèle considéré. Ceci est très important lorsque l'on veut analyser de gros volumes de données de façon complètement automatique. Un deuxième avantage non-négligeable de ces critères est qu'ils s'étendent de façon naturelle à la sélection simultanée de modèles via la correction de la fonction de pénalité⁴. Un critère de sélection simultanée de modèles, combiné avec l'étape de pré-traitement spatio-temporel présentée dans le chapitre 1, va nous permettre d'effectuer une segmentation spatio-temporelle automatique globale efficace d'une séquence de vidéo-microscopie à partir de plusieurs segmentations spatio-temporelles locales.

3.2 La sélection de modèle simple dans un cadre gaussien

Le problème de sélection de modèle que nous traitons dans cette section est le suivant. On observe des réalisations y_i d'un processus Y indexé par un paramètre i . Sans perte de généralité, on suppose que le paramètre i du processus prend ses valeurs dans le sous-ensemble des entiers $\{1, \dots, n\}$. Donc on écrit le processus Y comme suit :

$$Y_i = s(x_i) + \epsilon_i, i = 1, \dots, n \quad (3.1)$$

où $s(x_i)$ est la valeur déterministe (inconnue) du processus Y au point x_i (x_i étant déterministe), et $\epsilon_i, i = 1, \dots, N$ suivent un processus gaussien de moyenne nulle et de matrice de variances-covariances Σ connue.

Le problème est donc d'estimer la fonction (le vecteur) déterministe $s(\cdot)$ à partir d'une réalisation $(y_i)_{i=1, \dots, n}$ du processus Y qui est supposée appartenir à un certain \mathbb{S} qui est un sous-espace d'un certain espace de Hilbert \mathbb{H} qui est à son tour un sous-espace de \mathbb{R}^n .

Comme un tel problème étant mal posé, l'utilisateur impose donc un certain modèle *a priori* sur les données. Un tel modèle *a priori* permet au fait d'introduire une régularisation

³dans un sens que nous précisons plus loin dans ce chapitre

⁴via la correction d'une seule constante comme nous allons le démontrer dans la suite de ce chapitre.

dans le problème d'estimation de la fonction $s(\cdot)$ dans (3.1) pour en garantir une restauration efficace à partir d'une réalisation du processus Y .

Dans une telle perspective, nous allons alors d'abord fixer quelques notations, puis, nous introduisons quelques définitions et quelques outils que nous avons empruntés de [BM2001] dans le cas homoscedastique (i.e. variance homogène) et que nous avons tenu à généraliser dans le cas hétéroscedastique (i.e. variance inhomogène). De tels outils vont nous servir pour faire la démonstration de notre résultat principal du théorème (1).

Nous définissons les deux produits scalaires $\langle \cdot, \cdot \rangle$ et $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$ comme suit :

$$\langle t, u \rangle = t^T u, \quad \text{quels que soient } t, u \in \mathbb{S}$$

$$\langle t, u \rangle_{\Sigma^{-1}} = t^T \Sigma^{-1} u, \quad \text{quels que soient } t, u \in \mathbb{S}$$

où t^T signifie la transposée du vecteur t .

Aussi, nous notons par $\|\cdot\|$ la norme euclidienne sur \mathbb{S} et nous définissons la norme $\|\cdot\|_{\Sigma^{-1}}$ sur \mathbb{S} par :

$$\|t\|_{\Sigma^{-1}}^2 = \langle t, t \rangle_{\Sigma^{-1}} \quad \text{quel que soit } t \in \mathbb{S}.$$

3.2.1 Processus isonormal linéaire

Etant donné un sous-espace linéaire \mathbb{S} d'un certain espace de Hilbert \mathbb{H} muni du produit scalaire $\langle \cdot, \cdot \rangle$, un processus isonormal linéaire Z indexé par \mathbb{S} est presque sûrement un processus gaussien centré avec une matrice de variances-covariances ayant la structure $\text{Cov}(Z(t), Z(u)) = \langle t, u \rangle$.

La linéarité presque sûre (*p.s.*) de Z signifie que l'on peut trouver un sous-ensemble Ω' de Ω tel que $\mathbb{P}(\Omega') = 1$ et $\alpha Z(t)(w) + \beta Z(u)(w) = Z(\alpha t + \beta u)(w)$ quels que soient $w \in \Omega'$, $\alpha, \beta \in \mathbb{R}$ et $t, u \in \mathbb{S}$. Donc, étant donné un vecteur $t = (t_1, t_2, \dots, t_n) \in \mathbb{S}$ on peut écrire :

$$Z(t) = \sum_{i=1}^n \xi_i t_i$$

où $\xi_i, i = 1, \dots, n$ sont des gaussiennes indépendantes identiquement distribuées (i.i.d.) centrées réduites (i.e. avec une moyenne 0 et un écart type de 1) .

3.2.2 Processus linéaire gaussien

Etant donné un sous-espace \mathbb{S} d'un certain espace de Hilbert \mathbb{H} , on appelle un processus linéaire gaussien sur \mathbb{S} avec une moyenne s et une matrice de variances-covariances Σ tout processus Y indexé par \mathbb{S} de la forme :

$$Y(t) = \langle s, t \rangle + Z(\Sigma^{\frac{1}{2}}t), \quad \text{quel que soit } t \in \mathbb{S} \quad (3.2)$$

où Z représente un processus isonormal linéaire indexé par \mathbb{S} .

Remarque : Notant que (3.2) est une écriture compactée de (3.1). En effet, en considérant le modèle des données (3.1), il suffit alors de considérer le vecteur e_i comme le i -ème vecteur canonique de \mathbb{R}^n (i.e. dont toutes les composantes sont nulles sauf la i -ième composante qui est égale à 1) pour $i = 1, \dots, n$, et de définir Y_i comme étant $Y_i \triangleq Y(e_i)$. Il en découle alors par linéarité du produit scalaire que :

$$\begin{aligned} Y(t) &= \sum_{i=1}^n t_i Y(e_i) \equiv \sum_{i=1}^n t_i Y_i = \sum_{i=1}^n t_i (s(x_i) + \xi_i) \\ &= \sum_{i=1}^n t_i s(x_i) + \sum_{i=1}^n t_i \xi_i \equiv \langle s, t \rangle + Z(\Sigma^{\frac{1}{2}}t) \end{aligned}$$

3.2.3 Modèles linéaires et estimateurs de projection

Notre objectif, étant donné un processus linéaire gaussien Y sur \mathbb{S} , avec une moyenne inconnue $s \in \mathbb{H}$ et une structure de variances-covariances connue Σ , est de restaurer s à partir d'une réalisation du processus Y .

Modèles linéaires

On entend par modèle linéaire tout sous-espace de dimension finie de l'espace de Hilbert \mathbb{H} . On peut alors associer à un modèle linéaire \mathcal{S}_m une base linéaire orthonormée $\{\varphi_\lambda\}_{\lambda \in \Lambda_m}$ par rapport au produit scalaire $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$, i.e.

$$\langle \varphi_\lambda, \varphi_{\lambda'} \rangle_{\Sigma^{-1}} = \delta_{\lambda=\lambda'}$$

où δ est le symbole de Krönecker. Il en découle que quel que soit $t \in \mathcal{S}_m$, il existe un ensemble de coefficients $\beta_{\lambda \in \Lambda_m}$ tels que $t = \sum_{\lambda \in \Lambda_m} \beta_\lambda \varphi_\lambda$.

On considère ici une famille dénombrable de tels sous-espaces linéaires (modèles) $\{\mathcal{S}_m\}_{m \in \mathcal{M}}$. En pratique, de tels modèles linéaires sont spécifiés par l'utilisateur soit par connaissance *a priori* de la classe du modèle du vecteur déterministe s du processus étudié Y , ou pour décrire un motif dont il cherche les possibles occurrences dans les observations $y_i, i = 1, \dots, n$ du processus Y , et permettent donc d'introduire de la régularité dans le problème d'estimation de s .

Estimateur de projection associé à un modèle

Etant donné un processus linéaire gaussien Y , on définit pour tout vecteur $t \in \mathbb{S}$ la quantité suivante :

$$\gamma(t) = \|t\|_{\Sigma^{-1}}^2 - 2Y(\Sigma^{-1}t) \quad (3.3)$$

Cette quantité $\gamma(t)$ peut également s'écrire comme suit :

$$\gamma(t) = \|t - s\|_{\Sigma^{-1}}^2 - 2Z(\Sigma^{-\frac{1}{2}}t) \quad (3.4)$$

L'estimateur de projection de s par rapport au modèle \mathcal{S}_m noté \hat{s}_m est défini comme le vecteur qui minimise la quantité $\gamma(\cdot)$ sur \mathcal{S}_m et est donné par la formule suivante :

$$\hat{s}_m \triangleq \Pi_m Y \triangleq \arg \min_{t \in \mathcal{S}_m} \gamma(t)$$

Etant donné maintenant un modèle \mathbb{S}_m de dimension D_m , l'estimateur de projection \hat{s}_m peut être facilement calculé comme suit. On considère une base orthonormée $\{\varphi_\lambda\}_{\lambda \in \Lambda_m}$ par rapport au produit scalaire $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$, et posant $t := \sum_{\lambda \in \Lambda_m} \beta_\lambda \varphi_\lambda$. Par linéarité, minimiser $\gamma(t)$ revient à minimiser la quantité

$$\sum_{\lambda \in \Lambda_m, \lambda' \in \Lambda_m} \beta_\lambda \beta_{\lambda'} \langle \varphi_\lambda, \varphi_{\lambda'} \rangle_{\Sigma^{-1}} - 2 \sum_{\lambda \in \Lambda_m} \beta_\lambda Y(\Sigma^{-1} \varphi_\lambda)$$

Posant $\hat{s}_m := \sum_{\lambda \in \Lambda_m} \hat{\beta}_\lambda \varphi_\lambda$ comme le vecteur qui minimise $\gamma(t)$. Il s'ensuit que les coefficients $\{\hat{\beta}_\lambda, \lambda \in \Lambda_m\}$ sont solution du système linéaire d'équations suivant :

$$2\hat{\beta}_\lambda \|\varphi_\lambda\|_{\Sigma^{-1}}^2 + \sum_{\lambda' \in \Lambda_m, \lambda' \neq \lambda} \hat{\beta}_{\lambda'} \langle \varphi_\lambda, \varphi_{\lambda'} \rangle_{\Sigma^{-1}} = 2Y(\Sigma^{-1} \varphi_\lambda), \lambda \in \Lambda_m$$

Et comme $\|\varphi_\lambda\|_{\Sigma^{-1}}^2 = 1$ et $\langle \varphi_\lambda, \varphi_{\lambda'} \rangle_{\Sigma^{-1}} = 0$ quel que soit $\lambda \neq \lambda'$, donc on obtient $\hat{\beta}_\lambda = Y(\Sigma^{-1}\varphi_\lambda), \forall \lambda \in \Lambda_m$, puis on dérive que

$$\hat{s}_m = \sum_{\lambda \in \Lambda_m} \hat{\beta}_\lambda \varphi_\lambda = \sum_{\lambda \in \Lambda_m} \langle s, \Sigma^{-1}\varphi_\lambda \rangle + \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)\varphi_\lambda$$

et on déduit enfin l'expression suivante de \hat{s}_m :

$$\hat{s}_m = s_m + \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)\varphi_\lambda \quad (3.5)$$

où s_m est la projection de s sur \mathcal{S}_m par le produit scalaire $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$.

Risque d'un estimateur de projection, estimateur pénalisé

On définit le risque d'un estimateur de projection \hat{s}_m de s par rapport à un modèle $m \in \mathcal{M}$ et qu'on note par $\mathcal{R}(\hat{s}_m)$ comme suit :

$$\mathcal{R}(\hat{s}_m) \triangleq \mathbb{E}\|\hat{s}_m - s\|_{\Sigma^{-1}}^2 \quad (3.6)$$

On peut alors définir le meilleur estimateur de s par rapport à $\mathcal{R}(\cdot)$ comme étant l'estimateur de projection qui minimise $\mathcal{R}(\hat{s}_m)$ sur $\{\hat{s}_m, m \in \mathcal{M}\}$, et on note son risque par :

$$\mathcal{R}(s, \mathcal{S}) \triangleq \arg \min_{m \in \mathcal{M}} \mathcal{R}(\hat{s}_m) \quad (3.7)$$

Toutefois, un tel estimateur ne peut pas être calculé en pratique. En effet, de la formule (3.5), et comme $Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)$ sont des gaussiennes centrées réduites, et que $|\Lambda_m| = D_m$, on obtient

$$\mathbb{E}\|\hat{s}_m - s\|_{\Sigma^{-1}}^2 = \|s_m - s\|_{\Sigma^{-1}}^2 + D_m \quad (3.8)$$

On retrouve ainsi un résultat classique en statistique qui est que le risque se décompose bien comme un terme de biais et un terme de variance, et on vient d'en démontrer la version hétéroscédastique quand la matrice de variances-covariances Σ est connue.

Idéalement, on cherche un estimateur \hat{s}^* qui minimise le risque (3.6) et par là, la partie de droite de (3.8). Mais malheureusement, comme le montre la décomposition biais-variance (3.8) du risque, un tel estimateur n'existe pas en pratique puisqu'il dépend de l'inconnu s . C'est pourquoi on minimise à la place un critère pénalisé de la forme :

$$\text{Crit}(\hat{s}_m) \triangleq \gamma(\hat{s}_m) + \text{pen}(m) \quad (3.9)$$

où $\text{pen}(m)$ est une fonction qui dépend seulement du modèle \mathcal{S}_m et non des données, et elle croît avec la complexité d'un modèle ⁵, de sorte qu'un modèle \mathcal{S}_m ayant une plus grande complexité se voit naturellement attribuer une pénalité $\text{pen}(m)$ plus forte. Au fait, on souhaite par une telle démarche sélectionner un estimateur \tilde{s} dont le risque ne s'éloigne pas trop de (3.6). L'estimateur \tilde{s} , appelé l'estimateur pénalisé de s , et qui dépend bien évidemment de s , de Σ ainsi que de la famille de modèles $\{\mathcal{S}_m\}_{m \in \mathcal{M}}$ et est donc défini par :

$$\tilde{s} \triangleq \arg \min_{m \in \mathcal{M}} \text{Crit}(\hat{s}_m) \quad (3.10)$$

Diverses fonctions de pénalités $\text{pen}(m)$ existent dans la littérature de sélection de modèle ⁶. Par exemple, Mallows [Mall1971] propose un critère pénalisé basé sur un estimateur sans biais du risque. Pour dériver son heuristique de sélection de modèle (appelée Mallows' C_p), Mallows avait procédé comme suit : D'après la formule de décomposition biais-variance (??) du risque, une procédure de sélection de modèle idéale minimiserait alors sur $\{\mathcal{S}_m\}_{m \in \mathcal{M}}$ la quantité :

$$\|s_m - s\|_{\Sigma^{-1}}^2 + D_m = (\|s_m - s\|_{\Sigma^{-1}}^2 - \|s\|_{\Sigma^{-1}}^2 + D_m) + \|s\|_{\Sigma^{-1}}^2 = -\|s_m\|_{\Sigma^{-1}}^2 + D_m + \|s\|_{\Sigma^{-1}}^2$$

ou de façon équivalente, la quantité :

$$-\|s_m\|_{\Sigma^{-1}}^2 + D_m$$

or, d'après la formule suivante dérivant de (3.5) :

$$\mathbb{E}[\|\hat{s}_m\|_{\Sigma^{-1}}^2] = \|s_m\|_{\Sigma^{-1}}^2 + D_m$$

un estimateur sans biais de $\|s_m\|_{\Sigma^{-1}}^2$ est donc

$$\|\hat{s}_m\|_{\Sigma^{-1}}^2 - D_m$$

d'où l'expression suivante à minimiser (qui est l'estimateur sans biais du risque $\mathcal{R}(\hat{s}_m)$ à un terme additif près qui n'intervient pas dans la minimisation) :

$$-\|\hat{s}_m\|_{\Sigma^{-1}}^2 + 2D_m$$

⁵Il est difficile de donner une définition générale du mot "complexité", mais en théorie de sélection de modèle, la complexité d'un modèle signifie sa capacité à représenter (modéliser) un ensemble de vecteurs dans \mathbb{S} , de sorte qu'un ensemble grand de tels vecteurs implique une complexité plus grande

⁶en voir une liste (non-exhaustive) dans l'introduction du présent chapitre

ou de façon équivalente, minimiser la quantité :

$$\|\hat{s}_m - Y\|_{\Sigma^{-1}}^2 + 2D_m$$

ce qui résulte en une fonction de pénalité égale à $2D_m$. Il est clair qu'une telle fonction de pénalité ne dépend pas de la richesse de la famille de modèles $\{\mathcal{S}_m\}_{m \in \mathcal{M}}$, tel est le cas aussi pour d'autres critères classiques (AIC, BIC, MDL, etc...).

Le résultat que nous présentons dans le théorème (1), et que nous nous efforcerons de démontrer tout dans la suite de ce chapitre repose sur une théorie bien solide inspirée de la théorie de sélection de modèle de Birgé & Massart. Au fait, il permet de choisir une forme de la pénalité $\text{pen}(m)$ qui prend en compte la richesse de la famille de modèles $\{\mathcal{S}_m\}_{m \in \mathcal{M}}$, et ce, pour garantir que le risque de l'estimateur sélectionné \tilde{s} ne soit pas trop grand comparé au risque théorique (3.6), et par là, éviter de sur-ajuster les données surtout dans le cas de familles de modèles $\{\mathcal{S}_m\}_{m \in \mathcal{M}}$ très larges.

3.3 Deux critères serrés pour la sélection de modèle dans un cadre gaussien

3.3.1 Le premier critère pénalisé (PCP)

Avant d'énoncer le premier théorème permettant le choix de la forme de la fonction de pénalité $\text{pen}(\cdot)$, nous énonçons d'abord le lemme suivant dont la preuve peut être trouvée dans [B & M (2004), B & M (2007)], et qui nous fournit une inégalité, appelée inégalité de concentration de Laurent & Massart (2000), qui va nous servir pour la démonstration du théorème 1 :

Lemme 1 *Soit V une variable aléatoire qui suit un $\chi^2(D)$. Alors, pour tout réel positif x , et pour tout $b \geq 2$, on a la borne suivante de la probabilité :*

$$\mathbb{P}\left[V \geq D + 2\sqrt{Dx} + bx\right] \leq \exp(-x)$$

Théoreme 1 On considère le processus Y défini dans (3.1) et la famille de modèles $\{m \in \mathcal{M}\}$ auxquels on associe des poids positifs L_m vérifiant :

$$\Delta \triangleq \sum_{m \in \mathcal{M}} \exp(-L_m D_m) < \infty \quad (3.11)$$

et soit un réel positif K vérifiant $K > 1$.

Il s'ensuit que si la fonction de pénalité d'un modèle $m \in \mathcal{M}$ est la suivante :

$$\text{pen}(m) \geq K D_m \left((1 + \theta) + \left(2 + \frac{1}{\theta}\right) L_m \right) \quad (3.12)$$

alors l'estimateur pénalisé \tilde{s} vérifie que

$$\begin{aligned} \mathbb{E} \|s - \tilde{s}\|_{\Sigma^{-1}}^2 \leq & \\ & \left(\frac{K}{K-1} \right)^2 \inf_{m \in \mathcal{M}} \left\{ \left[\sqrt{\|s - s_m\|_{\Sigma^{-1}}^2 + \left(\frac{K-1}{K} \right) \left(D_m (K(1+\theta) - 1) + \text{pen}(m) + K \left(2 + \frac{1}{\theta} \right) \Delta \right)} \right. \right. \\ & \left. \left. + \frac{\|s - s_m\|_{\Sigma^{-1}}}{K} \right]^2 \right\} \quad (3.13) \end{aligned}$$

■ **Preuve** : La preuve que nous donnons du théorème 1 est largement inspirée de [B&M (2001)].

Donc pour prouver (3.13), on considère un modèle $m \in \mathcal{M}$ arbitraire et on définit $\mathcal{M}' = \{m' \in \mathcal{M} | \text{Crit}(m') \leq \text{Crit}(m)\}$. On considère maintenant un autre modèle $m' \in \mathcal{M}'$. On suppose alors que

$$\text{Crit}(\hat{s}_{m'}) \leq \text{Crit}(\hat{s}_m)$$

et par définition de $\text{Crit}(\cdot)$, on obtient

$$\gamma(\hat{s}'_m) + \text{pen}(m') \leq \gamma(\hat{s}_m) + \text{pen}(m)$$

Notant donc par s_m la projection de s sur m , on obtient

$$\gamma(\hat{s}'_m) + \text{pen}(m') \leq \gamma(s_m) + \text{pen}(m) + [\gamma(\hat{s}_m) - \gamma(s_m)]$$

Posant

$$r_m := \gamma(\hat{s}_m) - \gamma(s_m)$$

Comme $\gamma(\hat{s}_m) \leq \gamma(s_m)$ puisque \hat{s}_m minimise $\gamma(t)$, pour tout $t \in \mathcal{S}_m$, donc on a presque sûrement $r_m \leq 0$. Utilisant maintenant le fait que

$$\gamma(t) = \|t - s\|_{\Sigma^{-1}}^2 - 2Z(\Sigma^{-\frac{1}{2}}t) - \|s\|_{\Sigma^{-1}}^2$$

on déduit donc que

$$\|s - \hat{s}'_m\|_{\Sigma^{-1}}^2 \leq \|s - s_m\|_{\Sigma^{-1}}^2 + 2[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)] + \text{pen}(m) - \text{pen}(m') + r_m \quad (3.14)$$

En utilisant l'inégalité (3.15) suivante :

$$2ab \leq \frac{a^2}{\eta} + b^2\eta, \quad \text{quel que soit } \eta > 0 \quad (3.15)$$

d'abord appliquée à

$$2[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)] \equiv 2\|\hat{s}'_m - s_m\|_{\Sigma^{-1}} \cdot \frac{[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)]}{\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}}$$

avec $\eta := \nu_m$, on a alors pour tout $\nu_m > 0$,

$$2[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)] \leq \frac{\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}^2}{\nu_m} + \nu_m \frac{[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)]^2}{\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}^2} \quad (3.16)$$

et comme $\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}^2 \leq \|\hat{s}'_m - s\|_{\Sigma^{-1}}^2 + \|s - s_m\|_{\Sigma^{-1}}^2 + 2\|\hat{s}'_m - s\|_{\Sigma^{-1}} \cdot \|s - s_m\|_{\Sigma^{-1}}$, alors on l'applique une seconde fois à $2\|\hat{s}'_m - s\|_{\Sigma^{-1}} \cdot \|s - s_m\|_{\Sigma^{-1}}$ avec $\eta := \mu_m$, on déduit que pour tout $\mu_m > 0$:

$$2[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)] \leq \frac{1}{\nu_m}(1 + \mu_m)\|s - \hat{s}'_m\|_{\Sigma^{-1}}^2 + \frac{1}{\nu_m}(1 + \frac{1}{\mu_m})\|s - \hat{s}_m\|_{\Sigma^{-1}}^2 + \nu_m \frac{[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)]^2}{\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}^2} \quad (3.17)$$

on trouve enfin que

$$(1 - \frac{2}{\nu_m}(1 + \mu_m))\|s - \hat{s}'_m\|_{\Sigma^{-1}}^2 \leq (1 + \frac{2}{\nu_m}(1 + \frac{1}{\mu_m}))\|s - s_m\|_{\Sigma^{-1}}^2 + \text{pen}(m) - \text{pen}(m') + \nu_m \frac{[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)]^2}{\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}^2} + r_m \quad (3.18)$$

On utilise alors l'inégalité du lemme (3.3.1) pour contrôler uniformément pour tout $t \in \mathcal{S}_{m'}$ la statistique suivante :

$$\frac{[Z(\Sigma^{-\frac{1}{2}}\hat{s}'_m) - Z(\Sigma^{-\frac{1}{2}}s_m)]^2}{\|\hat{s}'_m - s_m\|_{\Sigma^{-1}}^2}$$

Pour cela, considérant l'espace linéaire de dimension D défini par : $\mathcal{S}_{m,m'} = \mathcal{S}_m + \mathcal{S}_{m'}$, et auquel on associe une base ψ_1, \dots, ψ_D orthonormée par rapport au produit scalaire $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$, et en utilisant l'inégalité de Cauchy-Schwarz ⁷, on obtient alors

$$\begin{aligned} \sup_{t \in \mathcal{S}_{m'}} \left| \frac{Z(\Sigma^{-\frac{1}{2}}t) - Z(\Sigma^{-\frac{1}{2}}s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right| &\leq \sup_{u \in \mathcal{S}_{m,m'}} \frac{|Z(\Sigma^{-\frac{1}{2}}u)|}{\|u\|_{\Sigma^{-1}}} = \sup_{\alpha \in \mathbb{R}^D} \frac{|\sum_{j=1}^D \alpha_j Z(\Sigma^{-\frac{1}{2}}\psi_j)|}{\|\alpha\|} \\ &\leq \sup_{\alpha \in \mathbb{R}^D} \frac{\sum_{j=1}^D |\alpha_j| \cdot |Z(\Sigma^{-\frac{1}{2}}\psi_j)|}{\|\alpha\|} \leq \left[\sum_{j=1}^D Z^2(\Sigma^{-\frac{1}{2}}\psi_j) \right]^{\frac{1}{2}} \end{aligned}$$

d'où

$$\sup_{t \in \mathcal{S}_{m'}} \left[\frac{Z(\Sigma^{-\frac{1}{2}}t) - Z(\Sigma^{-\frac{1}{2}}s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right]^2 \leq \sum_{j=1}^D Z^2(\Sigma^{-\frac{1}{2}}\psi_j)$$

et remarquant que $\sum_{j=1}^D Z^2(\Sigma^{-\frac{1}{2}}\psi_j)$ est une $\chi^2(D)$, en appliquant alors le lemme 3.3.1, avec $b = 2$ et $x = (L_{m'}D_{m'} + \xi)$, avec $\xi > 0$, et tenant compte du fait que $D \leq D_{m'} + D_m$, on obtient

$$\mathbb{P} \left[\sup_{t \in \mathcal{S}_{m'}} \left[\frac{Z(\Sigma^{-\frac{1}{2}}t) - Z(\Sigma^{-\frac{1}{2}}s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right]^2 \geq U(m', m, \xi) \right] \leq \exp(-L_{m'}D_{m'} - \xi) \quad (3.19)$$

où $U(m', m, \xi) \equiv D_{m'} + D_m + 2\sqrt{(D_{m'} + D_m)(D_{m'}L_{m'} + \xi)} + 2(D_{m'}L_{m'} + \xi)$. Par conséquent

$$\mathbb{P} \left[\sup_{t \in \mathcal{S}_{m'}} \left| \frac{Z(\Sigma^{-\frac{1}{2}}t) - Z(\Sigma^{-\frac{1}{2}}s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right| \leq U(m', m, \xi) \right] \geq 1 - \exp(-L_{m'}D_{m'} - \xi) \quad (3.20)$$

Appliquant maintenant l'inégalité (3.15) pour $2\sqrt{D_{m'} + D_m}\sqrt{L_{m'}D_{m'} + \xi}$ avec $a = \sqrt{D_{m'} + D_m}$, $b = \sqrt{L_{m'}D_{m'} + \xi}$ et $\eta = \theta$, on obtient

$$U(m', m, \xi) \leq D_m(1 + \theta) + D_{m'} \left((1 + \theta) + \left(2 + \frac{1}{\theta}\right)L_{m'} \right) + \left(2 + \frac{1}{\theta}\right)\xi$$

Posant

$$\begin{aligned} U'(m', m, \xi) &\triangleq D_m(1 + \theta) + D_{m'} \left((1 + \theta) + \left(2 + \frac{1}{\theta}\right)L_{m'} \right) + \left(2 + \frac{1}{\theta}\right)\xi \\ &= \frac{1}{K} \text{pen}(m') + D_m(1 + \theta) + \left(2 + \frac{1}{\theta}\right)\xi \end{aligned} \quad (3.21)$$

Par conséquent

$$\mathbb{P} \left[\sup_{t \in \mathcal{S}_{m'}} \left| \frac{Z(\Sigma^{-\frac{1}{2}}t) - Z(\Sigma^{-\frac{1}{2}}s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right| \leq U'(m', m, \xi) \right] \geq 1 - \exp(-L_{m'}D_{m'} - \xi) \quad (3.22)$$

⁷Cette inégalité peut également être obtenu en écrivant le problème de Lagrange correspondant au problème d'optimisation : $\sup_{\alpha \in \mathbb{R}^D} \frac{\sum_{j=1}^D |\alpha_j| \cdot |Z(\Sigma^{-\frac{1}{2}}\psi_j)|}{\|\alpha\|}$

On définit maintenant l'évènement

$$\Omega_{\xi, m'} = \left\{ \sup_{t \in \mathcal{S}_{m'}} \left| \frac{Z(\Sigma^{-\frac{1}{2}} t) - Z(\Sigma^{-\frac{1}{2}} s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right| \leq U'(m', m, \xi) \right\}$$

ainsi que l'évènement simultané Ω_{ξ} donné par

$$\Omega_{\xi} = \bigcap_{m' \in \mathcal{M}} \Omega_{\xi, m'}$$

et notant par Ω_{ξ}^c son évènement complémentaire, on obtient alors

$$\begin{aligned} \mathbb{P}(\Omega_{\xi}^c) &\leq \sum_{m' \in \mathcal{M}} \mathbb{P} \left[\sup_{t \in \mathcal{S}_{m'}} \left| \frac{Z(\Sigma^{-\frac{1}{2}} t) - Z(\Sigma^{-\frac{1}{2}} s_m)}{\|t - s_m\|_{\Sigma^{-1}}} \right| \geq U'(m', m, \xi) \right] \\ &\leq \sum_{m' \in \mathcal{M}} \exp(-L_{m'} D_{m'} - \xi) \end{aligned} \quad (3.23)$$

Il s'ensuit que simultanément pour tout $m' \in \mathcal{M}$, sauf sur un ensemble Ω_{ξ} de probabilité bornée par $\Delta \exp(-\xi)$, on a

$$\frac{[Z(\Sigma^{-\frac{1}{2}} t) - Z(\Sigma^{-\frac{1}{2}} s_m)]^2}{\|t - s_m\|_{\Sigma^{-1}}^2} \leq \frac{1}{K} \text{pen}(m') + D_m(1 + \theta) + (2 + \frac{1}{\theta})\xi$$

En combinant avec (3.18), il s'ensuit que sauf sur Ω_{ξ} , pour tout $m' \in \mathcal{M}'$, on a

$$\begin{aligned} (1 - \frac{2}{\nu_m}(1 + \mu_m)) \|s - \hat{s}_{m'}\|_{\Sigma^{-1}}^2 &\leq (1 + \frac{2}{\nu_m}(1 + \frac{1}{\mu_m})) \|s - s_m\|_{\Sigma^{-1}}^2 + (\frac{\nu_m}{K} - 1) \text{pen}(m') \\ &\quad + \nu_m D_m(1 + \theta) + \text{pen}(m) + \nu_m(2 + \frac{1}{\theta})\xi + r_m \end{aligned}$$

Posant $\nu_m := K$, et utilisant cette dernière inégalité avec $\hat{s}_{m'} := \tilde{s}$, donc sauf sur Ω_{ξ} ,

$$\begin{aligned} (1 - \frac{1}{K}(1 + \mu_m)) \|s - \tilde{s}\|_{\Sigma^{-1}}^2 &\leq (1 + \frac{1}{K}(1 + \frac{1}{\mu_m})) \|s - s_m\|_{\Sigma^{-1}}^2 \\ &\quad + K D_m(1 + \theta) + \text{pen}(m) + K(2 + \frac{1}{\theta})\xi + r_m \end{aligned}$$

En prenant μ tel que $K > 1 + \mu_m$, on déduit que sauf sur Ω_{ξ} ,

$$\begin{aligned} \|s - \tilde{s}\|_{\Sigma^{-1}}^2 &\leq \frac{(1 + \frac{1}{K}(1 + \frac{1}{\mu_m}))}{(1 - \frac{1}{K}(1 + \mu_m))} \|s - s_m\|_{\Sigma^{-1}}^2 \\ &\quad + \frac{1}{(1 - \frac{1}{K}(1 + \mu_m))} [K D_m(1 + \theta) + \text{pen}(m) + K(2 + \frac{1}{\theta})\xi + r_m] \end{aligned} \quad (3.24)$$

par conséquent, il existe une variable aléatoire V telle que $\mathbb{P}(V \geq \xi) \geq \Delta \exp(-\xi)$, $\mathbb{E}(V) \leq \Delta$ ⁸, et que l'inégalité suivante soit vérifiée *p.s.* :

$$\begin{aligned} \|s - \tilde{s}\|_{\Sigma^{-1}}^2 &\leq \frac{(1 + \frac{1}{K}(1 + \frac{1}{\mu_m}))}{(1 - \frac{1}{K}(1 + \mu_m))} \|s - s_m\|_{\Sigma^{-1}}^2 \\ &\quad + \frac{1}{(1 - \frac{1}{K}(1 + \mu_m))} [KD_m(1 + \theta) + \text{pen}(m) + K(2 + \frac{1}{\theta})V + r_m] \end{aligned} \quad (3.25)$$

Calculant maintenant l'espérance de r_m . On a

$$r_m \triangleq \gamma(\hat{s}_m) - \gamma(s_m)$$

De l'équation (3.4), on dérive que

$$\gamma(\hat{s}_m) = \|\hat{s}_m - s\|_{\Sigma^{-1}}^2 - \|s\|_{\Sigma^{-1}}^2 - 2Z(\Sigma^{-\frac{1}{2}}\hat{s}_m) \quad (3.26)$$

et de l'équation (3.5) on a

$$Z(\Sigma^{-\frac{1}{2}}\hat{s}_m) = Z(\Sigma^{-\frac{1}{2}}s_m + \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)\Sigma^{-\frac{1}{2}}\varphi_\lambda)$$

et par linéarité, on trouve que

$$Z(\Sigma^{-\frac{1}{2}}\hat{s}_m) = Z(\Sigma^{-\frac{1}{2}}s_m) + \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)^2 \quad (3.27)$$

On déduit aussi de l'équation (3.5) que

$$\|\hat{s}_m - s\|_{\Sigma^{-1}}^2 = \|s_m - s\|_{\Sigma^{-1}}^2 + \sum_{\lambda \in \Lambda_m} \langle s_m - s, Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)\varphi_\lambda \rangle_{\Sigma^{-1}} + \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)^2 \quad (3.28)$$

d'où, en combinant (3.26), (3.27) et (3.28)

$$\gamma(\hat{s}_m) - \gamma(s_m) = \sum_{\lambda \in \Lambda_m} \langle s_m - s, Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)\varphi_\lambda \rangle_{\Sigma^{-1}} - \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\varphi_\lambda)^2$$

d'où enfin

$$\mathbb{E}(\gamma(\hat{s}_m) - \gamma(s_m)) = -D_m$$

Essayant maintenant d'optimiser le second côté de cette dernière inégalité par rapport à μ_m .

Pour cela, posant

$$\begin{aligned} A_m &:= \|s - s_m\|_{\Sigma^{-1}}^2 \\ B_m &:= KD_m(1 + \theta) + \text{pen}(m) + K(2 + \frac{1}{\theta})\Delta - D_m \end{aligned}$$

⁸En utilisant le fait que $\mathbb{E}(V) = \int_v \mathbb{P}(V \geq v)dv$

On a donc

$$\|s - \tilde{s}\|_{\Sigma^{-1}}^2 \leq \left[\frac{(1 + \frac{1}{K}(1 + \frac{1}{\mu_m}))}{(1 - \frac{1}{K}(1 + \mu_m))} A_m + \frac{1}{(1 - \frac{1}{K}(1 + \mu_m))} B_m \right] + \left[\frac{K(V - \Delta)}{(1 - \frac{1}{K}(1 + \mu_m))} (2 + \frac{1}{\theta}) + \frac{r_m + D_m}{(1 - \frac{1}{K}(1 + \mu_m))} \right] \quad (3.29)$$

On définit maintenant la fonction $f_m(\mu_m)$ par

$$f_m(\mu_m) = \frac{(1 + \frac{1}{K}(1 + \frac{1}{\mu_m}))}{(1 - \frac{1}{K}(1 + \mu_m))} A_m + \frac{1}{(1 - \frac{1}{K}(1 + \mu_m))} B_m$$

et cherchant son minimum par rapport à μ . Alors, une condition nécessaire pour que $\mu^*(m)$ soit le point minimum de $f(\mu)$ est que $f'(\mu_m^*) = 0$. On a

$$f'_m(\mu_m) = \frac{1}{K} \left[\frac{\frac{-1}{\mu_m^2} A_m}{(1 - \frac{1}{K}(1 + \mu_m))} + \frac{1}{(1 - \frac{1}{K}(1 + \mu_m))} \frac{(1 + \frac{1}{K}(1 + \frac{1}{\mu_m})) A_m + B_m}{(1 - \frac{1}{K}(1 + \mu_m))} \right]$$

On déduit alors que

$$f'_m(\mu_m) = \frac{1}{K(1 - \frac{1}{K}(1 + \mu_m))} \left[\frac{-1}{\mu_m^2} A_m + f_m(\mu_m) \right]$$

Et puisque $f'_m(\mu_m^*) = 0$, on déduit que $f_m(\mu_m^*) = \frac{A_m}{(\mu_m^*)^2}$ d'un côté, et de l'autre côté $f'_m(\mu_m^*) = 0$. On trouve alors

$$\left[\left(1 + \frac{1}{K}\right) + \frac{2}{K\mu_m^*} + \left(\frac{1}{K} - 1\right) \frac{1}{(\mu_m^*)^2} \right] A_m + B_m = 0 \quad (3.30)$$

et

$$f_m(\mu_m^*) = \frac{1}{(\mu_m^*)^2} A_m = \frac{K}{K-1} \left[B_m + \left(\left(1 + \frac{1}{K}\right) + \frac{2}{K\mu_m^*} \right) A_m \right]$$

En résolvant par rapport à μ_m^* l'égalité (3.30), on obtient

$$\mu_m^* = \frac{\sqrt{A_m^2 + (1 - \frac{1}{K}) A_m B_m - \frac{A_m}{K}}}{(1 + \frac{1}{K}) A_m + B_m} \quad (3.31)$$

On montre facilement que $f''_m(\mu_m^*) = \frac{2A_m}{(\mu_m^*)^3} > 0$ ce qui signifie que μ_m^* est le minimum de $f(\mu)$. De l'égalité (3.30), on déduit que

$$\begin{aligned} \left(1 + \frac{1}{K}\right) A_m + \frac{2\mu_m^*}{K} \frac{A_m}{(\mu_m^*)^2} + \left(\frac{1}{K} - 1\right) \frac{A_m}{(\mu_m^*)^2} + B_m &= \frac{2\mu_m^*}{K} f_m(\mu_m^*) + \left(\frac{1}{K} - 1\right) f_m(\mu_m^*) \\ &\quad + B_m + \left(1 + \frac{1}{K}\right) A_m \\ &= \left(\frac{2\mu_m^*}{K} + \left(\frac{1}{K} - 1\right)\right) f_m(\mu_m^*) \\ &\quad + B_m + \left(1 + \frac{1}{K}\right) A_m \\ &= 0 \end{aligned}$$

ceci implique que

$$\left(-\frac{2\mu_m^*}{K} + \left(1 - \frac{1}{K}\right)\right) f_m(\mu_m^*) = B_m + \left(1 + \frac{1}{K}\right) A_m$$

Puisque $B_m + \left(1 + \frac{1}{K}\right) A_m \geq 0$ et $f(\mu) \geq 0$ quel que soit $\mu > 0$, on déduit alors que $-\frac{2\mu_m^*}{K} + \left(1 - \frac{1}{K}\right) > 0$, d'où $\mu_m^* \leq \frac{K-1}{2} \leq K-1$. Calculant maintenant $f_m(\mu_m^*)$. On a

$$f_m(\mu_m^*) = \frac{A_m}{(\mu_m^*)^2} = \frac{A_m \left(\left(1 + \frac{1}{K}\right) A_m + B_m \right)^2}{\left(\sqrt{A_m^2 + \left(1 - \frac{1}{K}\right) A_m B_m} - \frac{A_m}{K} \right)^2} \quad (3.32)$$

En multipliant et en divisant par le conjugué du dénominateur, on obtient

$$f_m(\mu_m^*) = \frac{A_m \left(\left(1 + \frac{1}{K}\right) A_m + B_m \right)^2 \left(\sqrt{A_m^2 + \left(1 - \frac{1}{K}\right) A_m B_m} + \frac{A_m}{K} \right)^2}{\left(\left(1 - \frac{1}{K^2}\right) A_m^2 + \left(1 - \frac{1}{K}\right) A_m B_m \right)^2} \quad (3.33)$$

$$= \left(\frac{K}{K-1} \right)^2 \left(\sqrt{A_m + \left(1 - \frac{1}{K}\right) B_m} + \frac{\sqrt{A_m}}{K} \right)^2 \quad (3.34)$$

Enfin, en remplaçant A_m et B_m par leurs expressions respectives dans l'expression (3.33), en remplaçant cette dernière dans l'inégalité (3.29), et en appliquant l'opérateur de l'espérance à chaque côté de cette dernière, on retrouve ainsi l'inégalité (3.13) du théorème 1.

3.3.2 Le second critère pénalisé (SCP)

Nous présentons maintenant un second critère de sélection de modèle dans un cadre gaussien. Pour cela, on considère l'inégalité de concentration du lemme 2 suivant due à Birgé & Massart [BM2004, BM2007] qui va nous permettre de démontrer un tel résultat :

Lemme 2 *Soit V et U deux variables aléatoires indépendantes suivant une distribution du $\chi^2(D)$ et une distribution Gaussienne standard $\mathcal{N}(0, 1)$ respectivement. Alors pour tous réels positifs x et ρ , et pour tout réel $b \geq 2$, on a la borne suivante de la probabilité :*

$$\mathbb{P}\left[V + \rho U \geq D + \frac{\rho^2}{2b} + 2\sqrt{Dx} + bx\right] \leq \exp(-x)$$

Théorème 2 *On considère le processus Y défini dans (3.1) et la famille de modèles $\{m \in \mathcal{M}\}$ auxquels on associe des poids positifs L_m vérifiant :*

$$\Delta \triangleq \sum_{m \in \mathcal{M}} \exp(-L_m D_m) < \infty \quad (3.35)$$

et soit deux réels $\theta > 0$ et $\alpha \in (0, 1)$.

Il s'ensuit que si la fonction de pénalité d'un modèle $S_m \in \mathcal{M}$ est définie comme suit :

$$\text{pen}(m) \geq \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m$$

alors l'estimateur pénalisé \tilde{s} vérifie que :

$$\mathbb{E} \|s - \tilde{s}\|_{\Sigma^{-1}}^2 \leq \frac{1}{1 - \alpha} \inf_{m \in \mathcal{M}} \left\{ \|s - s_m\|_{\Sigma^{-1}}^2 + \text{pen}(m) - D_m + \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta} \right) \Delta \right\}$$

■ **Preuve** : La preuve que nous donnons du théorème que 2 est quasiment identique à celle du théorème 1 principal dans [BM2004, BM2007]. Toutefois, nous introduisons dans la fonction de pénalité un paramètre α qui va nous permettre par la suite de dériver une stratégie optimale (dans le sens de la borne supérieure du risque) de choix des poids et des constantes dans la fonction de pénalité. Voici alors la preuve du théorème 2 :

D'abord, on associe à chaque $m \in \mathcal{M}$ une base orthonormée $\{\phi_\lambda\}_{\lambda \in \Lambda_m}$ par rapport au produit scalaire $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$, avec $|\Lambda_m| \equiv D_m$. On a alors pour tout $t \in \mathcal{S}_m$

$$Z(\Sigma^{-\frac{1}{2}}t) = \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\phi_\lambda) \langle t, \phi_\lambda \rangle_{\Sigma^{-1}} \triangleq \langle t, \zeta_m \rangle_{\Sigma^{-1}} \quad \text{avec} \quad \zeta_m \triangleq \sum_{\lambda \in \Lambda_m} Z(\Sigma^{-\frac{1}{2}}\phi_\lambda) \phi_\lambda \in \mathcal{S}_m$$

et on déduit que

$$\zeta_m \sim \mathcal{N}(0, I_m) \quad \text{et} \quad V_m \triangleq \|\zeta_m\|^2 \sim \chi^2(D_m) \tag{3.36}$$

où $\mathcal{N}(0, I_m)$ est la gaussienne standard D_m -dimensionnelle et $\chi^2(D_m)$ est la distribution du Chi-deux avec D_m degrés de liberté. On a

$$\begin{aligned} \gamma(t) &= \|s - t\|_{\Sigma^{-1}}^2 + \|t - s_m\|_{\Sigma^{-1}}^2 - \|s\|_{\Sigma^{-1}}^2 - 2Z(\Sigma^{-\frac{1}{2}}t) \\ &= \|s - t\|_{\Sigma^{-1}}^2 + \|t - s_m\|_{\Sigma^{-1}}^2 - \|s\|_{\Sigma^{-1}}^2 - 2\langle t, \zeta_m \rangle_{\Sigma^{-1}} \end{aligned}$$

et de l'égalité (3.5), on déduit que

$$\gamma(\hat{s}_m) = \|s - s_m\|_{\Sigma^{-1}}^2 - \|s\|_{\Sigma^{-1}}^2 - V_m - 2Z(\Sigma^{-\frac{1}{2}}s_m) \tag{3.37}$$

et

$$\|\hat{s}_m - s\|_{\Sigma^{-1}}^2 = \|s - s_m\|_{\Sigma^{-2}}^2 + V_m \tag{3.38}$$

Posant $\tilde{s} = \hat{s}_{\tilde{m}}$, de la définition de \tilde{s} , on a alors

$$\|s\|_{\Sigma^{-1}}^2 + 2Z(\Sigma^{-\frac{1}{2}}s) + \gamma(\tilde{s}) + \text{pen}(\tilde{m}) = \inf_{m \in \mathcal{M}} \{ \|s\|_{\Sigma^{-1}}^2 + 2Z(\Sigma^{-\frac{1}{2}}s) + \gamma(\hat{s}_m) + \text{pen}(m) \}$$

et de (3.37) et (3.38), on a

$$\|s\|_{\Sigma^{-1}}^2 + 2Z(\Sigma^{-\frac{1}{2}}s) + \gamma(\tilde{s}) = \|s - \tilde{s}\|_{\Sigma^{-1}}^2 - 2V_{\tilde{m}} - 2Z(\Sigma^{-\frac{1}{2}}(s_{\tilde{m}} - s))$$

et en utilisant (3.37), on déduit que

$$\begin{aligned} \|\tilde{s} - s\|_{\Sigma^{-1}}^2 &= 2V_{\tilde{m}} + 2Z(\Sigma^{-\frac{1}{2}}(s_{\tilde{m}} - s)) - \text{pen}(\tilde{m}) + \\ &\quad \inf_{m \in \mathcal{M}} \{ \|s - s_m\|_{\Sigma^{-1}}^2 + 2Z(\Sigma^{-\frac{1}{2}}s) + -2Z(\Sigma^{-\frac{1}{2}}(s_m - s)) - V_m + \text{pen}(m) \} \end{aligned}$$

En posant $d_m := \|\tilde{s} - s\|$, $U_m := \frac{Z(\Sigma^{-\frac{1}{2}}(s_m - s))}{d_m}$, et en utilisant le fait que $\|\tilde{s} - s\|_{\Sigma^{-1}} = V_{\tilde{m}} + d_m^2$, on obtient pour tout $\alpha \in (0, 1)$

$$\begin{aligned} (1 - \alpha)\|\tilde{s} - s\|_{\Sigma^{-1}}^2 &= (2 - \alpha)V_{\tilde{m}} - \alpha d_m^2 + 2d_m U_{\tilde{m}} - \text{pen}(\tilde{m}) + \\ &\quad \inf_{m \in \mathcal{M}} \{ d_m^2 - 2d_m U_m - V_m + \text{pen}(m) \} \end{aligned}$$

Posant $Q_m := (2 - \alpha)V_m - \alpha d_m^2 + 2d_m U_{\tilde{m}} - \text{pen}(m)$ et $R_m := d_m^2 - 2d_m U_m - V_m + \text{pen}(m)$, on a donc

$$\|\tilde{s} - s\|_{\Sigma^{-1}}^2 = \frac{1}{(1 - \alpha)} \left(Q_{\tilde{m}} + \inf_{m \in \mathcal{M}} R_m \right)$$

Puisque \tilde{m} peut être n'importe quel élément de \mathcal{M} , alors on a besoin de contrôler uniformément Q_m sur \mathcal{M} . Pour cela, posant $A_m := V_m + \frac{2}{2-\alpha}d_m U_m$, et $x_m := L_m D_m + \xi$ pour un certain $\xi > 0$, et définissant l'évènement

$$\Omega_{\xi, m} \triangleq \left\{ A_m < D_m + \frac{\alpha}{2-\alpha}d_m^2 + 2\sqrt{D_m x_m} + \frac{2x_m}{\alpha(2-\alpha)} \right\}$$

et l'évènement simultané $\Omega_\xi \triangleq \bigcap_{m \in \mathcal{M}} \Omega_{\xi, m}$, et notant $\Omega_{\xi, m}^c$ et Ω_ξ leurs évènements complémentaires respectifs. Compte tenu du fait que

$$\mathbb{E}[Z(\phi_\lambda)Z(s - s_m)] = \langle \phi_\lambda, s - s_m \rangle_{\Sigma^{-1}} = 0$$

alors les variables aléatoires ζ_m et $Z(s - s_m)$ sont indépendantes et ainsi sont V_m et U_m qui ont respectivement comme distribution un $\chi^2(D_m)$ et une $\mathbb{N}(0, 1)$. En appliquant alors le

lemme 2 avec $\rho := \frac{2}{2-\alpha}d_m$ et $b := \frac{2}{\alpha(2-\alpha)}$, on trouve que $\mathbb{P}(\Omega_{\xi,m}^c) \leq \exp(-x_m)$, d'où

$$\mathbb{P}(\Omega_{\xi}^c) \leq \sum_{m \in \mathcal{M}} \exp(-L_m D_m - \xi) \equiv \Delta \exp(-\xi) \quad (3.39)$$

Utilisant maintenant l'inégalité (3.15) appliquée à $2\sqrt{D_m(L_m D_m + \xi)} \equiv 2\sqrt{D_m}\sqrt{D_m L_m + \xi}$ avec $a := \sqrt{D_m}$ et $b := \sqrt{D_m L_m + \xi}$, on trouve que pour tout $\theta > 0$

$$A_m \leq (1 + \alpha)D_m + \frac{\alpha d_m^2}{2 - \alpha} + \left(\frac{1}{\theta} + \frac{2}{\alpha(2 - \alpha)}\right)L_m D_m + \left(\frac{1}{\theta} + \frac{2}{\alpha(2 - \alpha)}\right)\xi$$

d'où

$$\begin{aligned} (2 - \alpha)A_m &\leq \left[(2 - \alpha)(1 + \theta)D_m + \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)L_m D_m\right] + \alpha d_m^2 + \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)\xi \\ &\leq \text{pen}(m) + \alpha d_m^2 + \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)\xi \end{aligned}$$

d'où sauf sur $\Omega_{\xi,m}^c$, on a pour tout $m \in \mathcal{M}$

$$Q_m \leq \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)\xi$$

et en particulier, pour $m := \tilde{m}$, sauf sur $\Omega_{\xi,m}^c$

$$Q_{\tilde{m}} \leq \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)\xi$$

d'où, compte tenu de (3.39), on a

$$\mathbb{P}\left[Q_{\tilde{m}} > \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)\xi\right] \leq \Delta \exp(-\xi)$$

et en utilisant le fait que pour toute variable aléatoire V , on a si $\int_v \mathbb{P}(V > v)dv < \infty$, alors $\mathbb{E}(V) = \int_v \mathbb{P}(V > v)dv$, on trouve que

$$\mathbb{E}(Q_m) \leq \left(\frac{2}{\alpha} + \frac{2 - \alpha}{\theta}\right)\Delta$$

et comme $\mathbb{E}\left[\inf_{m \in \mathcal{M}} R_m\right] \leq \inf_{m \in \mathcal{M}} \mathbb{E}(R_m) = \inf_{m \in \mathcal{M}} (d_m^2 + \text{pen}(m) - D_m)$ et compte tenu de (3.39), on déduit alors le résultat du théorème.

3.4 Stratégie de choix des poids et des constantes dans la fonction de pénalité

Nous décrivons dans cette sous-section une stratégie optimale (dans le sens de la borne supérieure du risque) de choix des poids et des constantes dans la fonction de pénalité

$\text{pen}(m)$ des deux critères pénalisés que nous avons proposés dans la section précédente. Nous évaluons la borne supérieure du risque de chacun des deux critères pénalisés induite par une telle stratégie, et nous verrons que le deuxième critère que nous avons proposé a la borne supérieure du risque qui est plus petite pour la stratégie proposée.

3.4.1 Choix des poids L_m et des constantes K et θ dans la fonction de pénalité du premier critère

Comme nous avons utilisé des inégalités serrées pour dériver une borne supérieure du risque de l'estimateur pénalisé, cette stratégie tente alors de contrôler le risque de l'estimateur pénalisé à travers le contrôle de la borne supérieure suivante de son risque :

$$\mathcal{B}' = \left(\frac{K}{K-1}\right)^2 \inf_{m \in \mathcal{M}} \left\{ \left[\sqrt{\|s - s_m\|_{\Sigma^{-1}}^2 + \left(\frac{K-1}{K}\right) \left((K(1+\theta) - 1) D_m + \text{pen}(m) + K \left(2 + \frac{1}{\theta}\right) \Delta \right)} + \frac{\|s - s_m\|_{\Sigma^{-1}}}{K} \right]^2 \right\} \quad (3.40)$$

Choix de θ et L_m

Proposition 1 *Si la fonction de pénalité est choisie comme suit*

$$\text{pen}(m) \equiv K D_m \left((1 + \theta) + \left(2 + \frac{1}{\theta}\right) L_m \right)$$

où les poids L_m sont donnés par la formule

$$L_m \equiv 2\theta^2 - \frac{\Delta}{D_m}$$

avec $\Delta > 0$, et $\theta > 0$ étant la plus petite valeur vérifiant

$$2\theta^2 \geq \Delta$$

$$\Delta \exp(-\Delta) = \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-2\theta^2 D)$$

où $\text{Card}_{\mathcal{M}}(D)$ représente le nombre de modèles $\{\mathcal{S}_m \in \mathcal{M}\}$ ayant la dimension D , et D_{max} est la dimension maximale autorisée pour un modèle.

On a alors

$$\begin{aligned} & \mathbb{E}\|s - \tilde{s}\|_{\Sigma^{-1}}^2 \\ & \leq \left(\frac{K}{K-1}\right)^2 \inf_{m \in \mathcal{M}} \left\{ \left[\sqrt{\|s - s_m\|_{\Sigma^{-1}}^2 + (K-1) \left((2 + 4\theta + 4\theta^2) - \frac{1}{K} \right) D_m} + \frac{\|s - s_m\|_{\Sigma^{-1}}}{K} \right]^2 \right\} \end{aligned} \quad (3.41)$$

Preuve : Pour démontrer cette proposition, fixant le paramètre K à une valeur donnée et cherchant à trouver les expressions des poids L_m et la valeur de θ minimisant la borne \mathcal{B} indépendamment de m . On choisit alors de minimiser B_m d'abord par rapport à θ . Cela revient donc à minimiser par rapport à θ l'expression

$$\mathcal{B}'_m \triangleq 2D_m(1 + \theta) + \left(2 + \frac{1}{\theta}\right)L_m D_m + \left(2 + \frac{1}{\theta}\right)\Delta - \frac{D_m}{K} \quad (3.42)$$

En annulant donc la dérivée rapport à θ de cette dernière expression, on trouve que la valeur de θ optimale doit vérifier l'égalité suivante :

$$\frac{\partial \mathcal{B}'_m}{\partial \theta} = 2D_m - \frac{L_m D_m}{\theta^2} - \frac{\Delta}{\theta^2} = 0$$

d'où

$$L_m = 2\theta^2 - \frac{\Delta}{D_m} \quad (3.43)$$

On vérifie que

$$\frac{\partial^2 \mathcal{B}'_m}{\partial \theta^2} = \frac{4L_m D_m + 2\Delta}{\theta^3} > 0$$

ce qui prouve que θ qui vérifie l'expression (3.43) est bien un minimum de \mathcal{B}'_m . Comme l'expression des poids L_m (3.43) doit être vraie quel que soit $m \in \mathcal{M}$, il faut alors qu'il existe un réel $\delta > 0$ tel que les poids $L_m, \forall m \in \mathcal{M}$ s'expriment comme suit :

$$L_m \equiv 2\delta^2 - \frac{\Delta}{D_m} \quad (3.44)$$

et par identification, on trouve que θ vérifie $\theta \equiv \delta$. Et compte tenu de l'identité

$$\Delta \triangleq \sum_{m \in \mathcal{M}} \exp(-L_m D_m)$$

θ doit alors vérifier l'égalité

$$\Delta \exp(-\Delta) = \sum_{m \in \mathcal{M}} \exp(-2\theta^2 D_m) \quad (3.45)$$

et en regroupant dans le terme de la somme tous les modèles qui ont la même dimension, on trouve que θ doit vérifier l'égalité

$$\Delta \exp(-\Delta) = \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-2\theta^2 D) \quad (3.46)$$

Et compte tenu du fait que $L_m \geq 0, \forall m \in \mathcal{M}$, on a la contrainte suivante sur les valeurs de θ et de Δ :

$$2\theta^2 \geq \Delta \quad (3.47)$$

Remplaçant L_m par son expression dans (3.42), on trouve après calcul que

$$\mathcal{B}'_m = \left((2 + 4\theta + 4\theta^2) - \frac{1}{K} \right) D_m \quad (3.48)$$

par conséquent, pour minimiser \mathcal{B} , il faut choisir θ comme la plus petite valeur vérifiant l'égalité (3.46) et la contrainte (3.47)⁹. Il s'ensuit que

$$\mathcal{B} = \left(\frac{K}{K-1} \right)^2 \inf_{m \in \mathcal{M}} \left\{ \left[\sqrt{\|s - s_m\|_{\Sigma^{-1}}^2 + (K-1) \left((2 + 4\theta + 4\theta^2) - \frac{1}{K} \right) D_m} + \frac{\|s - s_m\|_{\Sigma^{-1}}}{K} \right]^2 \right\}$$

Choix de K

Reécrivant d'abord la borne supérieure du risque (3.40) comme suit :

$$\mathcal{B} = \inf_{m \in \mathcal{M}} \left\{ \left[\sqrt{\left(\frac{K}{K-1} \right)^2 \|s - s_m\|_{\Sigma^{-1}}^2 + \left(\frac{K^2}{K-1} \right) \left((2 + 4\theta + 4\theta^2) - \frac{1}{K} \right) D_m} + \frac{\|s - s_m\|_{\Sigma^{-1}}}{K-1} \right]^2 \right\} \quad (3.49)$$

Notant d'abord par K^* la valeur de K réalisant le minimum de l'expression :

$$\left(\frac{K^2}{K-1} \right) \left((2 + 4\theta + 4\theta^2) - \frac{1}{K} \right)$$

Maintenant, comme les deux expressions $\left(\frac{K}{K-1} \right)^2$ et $\frac{1}{K-1}$ diminuent quand K augmente, on déduit alors d'après la formule (3.49) que quand K augmente, les termes de biais ¹⁰ diminuent. Pour ce qui est du terme de variance ¹¹, en démarrant de $K = 1^+$, il diminue

⁹On reviendra plus tard à la technique pour calculer numériquement la valeur de θ

¹⁰i.e. $\left(\frac{K}{K-1} \right)^2 \|s - s_m\|_{\Sigma^{-1}}^2$ et $\frac{\|s - s_m\|_{\Sigma^{-1}}}{K-1}$

¹¹i.e. $\left(\frac{K^2}{K-1} \right) \left((2 + 4\theta + 4\theta^2) - \frac{1}{K} \right) D_m$

jusqu'à atteindre sa valeur minimale pour $K = K^*$, puis il augmente en allant de $K = K^{*+}$ vers l'infini. Et comme le coefficient de variance est plus important que le coefficient du biais, alors pour réaliser le meilleur compromis biais-variance, il faut prendre comme valeur de K dans la fonction de pénalité $K = K^*$ minimisant le coefficient du terme de variance. On peut vérifier facilement que $K^* \approx 2$ indépendamment du problème de sélection de modèle considéré et de sa famille de modèles $\{\mathcal{S}_m, m \in \mathcal{M}\}$. En Effet, l'égalité (3.46) implique que $\theta \geq \frac{1}{\sqrt{2}}$ (atteinte pour une famille de modèles de cardinalité 1, i.e. qui contient un seul modèle), ce qui fait que $\frac{1}{K}$ est presque négligeable devant $(2 + 4\theta + 4\theta^2)$, ce qui revient donc à minimiser une expression presque égale à $\left(\frac{K^2}{K-1}\right)$, et cette dernière atteint son minimum pour $K = 2$.

3.4.2 Choix des poids L_m et des constantes α et θ dans la fonction de pénalité du second critère

Comme on l'a fait pour le premier critère, le but est de trouver les expressions des poids L_m et des constantes α et θ dans la fonction de pénalité du second critère pénalisé que nous proposons qui minimisent la borne supérieure du risque de l'estimateur pénalisé correspondante.

Choix des poids L_m et de la constante θ

Proposition 2 *On considère le processus Y défini dans (3.1) et la famille de modèles $\{m \in \mathcal{M}\}$ auxquels on associe des poids positifs L_m donnés par la formule*

$$L_m \equiv \theta^2 - \frac{\Delta}{D_m}$$

où θ est la plus petite valeur positive vérifiant

$$\theta^2 \geq \Delta$$

$$\Delta \exp(-\Delta) = \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-\theta^2 D)$$

et soit réel $\alpha \in (0, 1)$, il s'ensuit que si la fonction de pénalité d'un modèle $\mathcal{S}_m \in \mathcal{M}$ est définie comme suit :

$$\text{pen}(m) \geq \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m$$

alors l'estimateur pénalisé \tilde{s} vérifie que :

$$\mathbb{E}\|s - \tilde{s}\|_{\Sigma^{-1}}^2 \leq \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}} \left\{ \frac{1}{1-\alpha} \|s - s_m\|_{\Sigma^{-1}}^2 + \left(\frac{2\theta^2}{\alpha(1-\alpha)} + \frac{2\theta}{1-\alpha} + 2\theta + 1 \right) D_m \right\}$$

■ **Preuve :** Considérant alors l'expression

$$\begin{aligned} \mathcal{B}_m &= \frac{1}{1-\alpha} \left[\|s - s_m\|_{\Sigma^{-1}}^2 + \text{pen}(m) - D_m + \left(\frac{2-\alpha}{\theta} + \frac{2}{\alpha} \right) \Delta \right] \\ &= \frac{1}{1-\alpha} \left[\|s - s_m\|_{\Sigma^{-1}}^2 + \left((2-\alpha)(1+\theta) - 1 + \left(\frac{2-\alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m + \left(\frac{2-\alpha}{\theta} + \frac{2}{\alpha} \right) \Delta \right] \end{aligned}$$

On choisit de minimiser \mathcal{B}_m d'abord par rapport à θ pour tout $m \in \mathcal{M}$. En annulant alors la dérivée de \mathcal{B}_m par rapport à θ , on trouve que la valeur optimale de θ correspondante existe quel que soit $m \in \mathcal{M}$ si et seulement si les poids s'expriment comme suit :

$$L_m \equiv \theta^2 - \frac{\Delta}{D_m}$$

et que θ est la plus petite valeur positive vérifiant

$$\begin{aligned} \theta^2 &\geq \Delta \\ \Delta \exp(-\Delta) &= \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-\theta^2 D) \end{aligned}$$

En remplaçant les expressions des poids L_m et de θ optimales dans \mathcal{B}_m , on trouve alors

$$\mathcal{B}_m = \left[\frac{1}{1-\alpha} \|s - s_m\|_{\Sigma^{-1}}^2 + \left(\frac{2\theta^2}{\alpha(1-\alpha)} + \frac{2\theta}{1-\alpha} + 2\theta + 1 \right) D_m \right] \quad (3.50)$$

Choix de α

En remarquant maintenant que le poids du terme de variance $\frac{2\theta^2}{\alpha(1-\alpha)} + \frac{2\theta}{1-\alpha} + 2\theta + 1$ est toujours plus grand que le poids du terme de biais $\frac{1}{1-\alpha}$, alors pour réaliser le meilleur compromis biais-variance, il faut équilibrer au maximum ces poids, et ceci revient donc à minimiser par rapport à α l'expression

$$\frac{2\theta^2}{\alpha(1-\alpha)} + \frac{2\theta}{1-\alpha} + 2\theta + 1$$

On trouve alors après calcul qu'une telle valeur de α s'écrit

$$\alpha = \frac{\theta}{\sqrt{\theta^2 + \theta} + \theta}$$

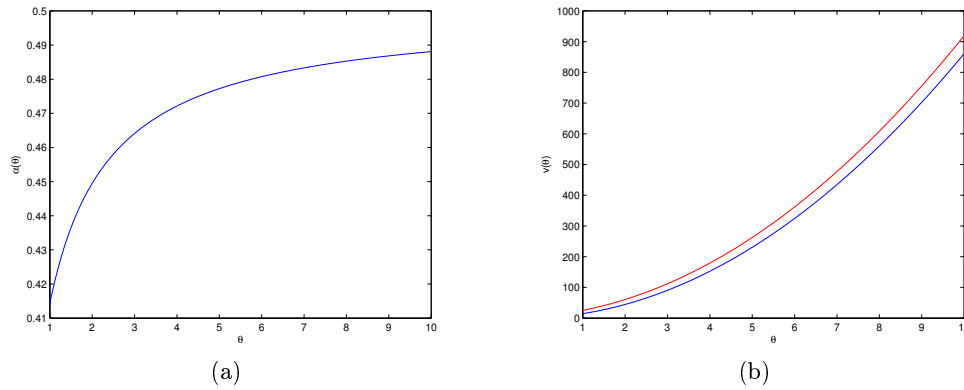


FIG. 3.1: (a) Les valeurs optimales de α en fonction des valeurs optimales de θ dans le second critère ; (b) Poids du terme de variance en fonction de θ . En rouge, le poids du terme de variance correspondant au premier critère, et en bleu, le poids du terme de variance correspondant au second critère

3.5 Comparaison des deux critères de sélection de modèles proposés

A priori, il est difficile de comparer les deux critères de sélection de modèle que nous avons proposés dans ce chapitre indépendamment de la stratégie de choix des paramètres de la fonction de pénalité (les poids et les autres constantes) utilisée. Toutefois, pour la stratégie que nous avons présentée dans ce chapitre, on remarque facilement la supériorité du second critère proposé sur le premier. En effet, d’abord, remarquant que le poids du terme de biais induit par le premier critère est supérieur à 4 tandis qu’il est compris entre $\sqrt{2} - 1$ et $\frac{1}{2}$ pour le second critère. Même chose pour le terme de variance, en effet, en effectuant dans le premier critère le changement de variable $\theta := \sqrt{2}\theta$, on se retrouve alors avec un même θ pour les deux critères, donc on pourra comparer les poids des termes de variance pour les deux critères. Comme le montre la figure FIG.3.8, le poids du terme de variance induit par le second critère est toujours plus petit que celui induit par le premier critère quelle que soit la valeur de θ considérée ; ce qui signifie que le second critère garantit une borne supérieure du risque qui est plus petite que celle du premier par une telle stratégie.

3.5.1 Calcul de θ et Δ

Considérant alors le second critère que nous avons proposé dans ce chapitre. Les poids sont alors donnés par l'expression

$$L_m \equiv \theta^2 - \frac{\Delta}{D_m}$$

et la fonction de pénalité est donnée par

$$\text{pen}(m) \equiv \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m$$

où $\theta > 0$ et $\Delta > 0$ vérifient que θ est la plus petite valeur satisfaisant aux contraintes

$$\begin{aligned} \theta^2 &\geq \Delta \\ \Delta \exp(-\Delta) &= \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-\theta^2 D) \end{aligned}$$

On remarque que la fonction

$$h(x) = x \exp(-x), x > 0$$

est unimodale concave et atteint son maximum $\exp(-1)$ pour $x^* = 1$ de sorte qu'elle est croissante dans l'intervalle $[0, 1]$ et décroissante dans l'intervalle $[1, +\infty[$, et elle vérifie la propriété

$$\forall x, 0 < x < 1, \exists! y > 1, h(x) \equiv h(y)$$

et

$$\forall x, 1 < x < \infty, \exists! y < 1, h(x) \equiv h(y)$$

et comme l'on cherche la plus petite valeur de θ et que $\theta^2 > \Delta$, alors on peut restreindre la recherche de la valeur de Δ sur l'intervalle $]0, 1]$. On remarque aussi que la fonction $g(\cdot)$ donnée par

$$g(\theta) = \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-\theta^2 D), \theta > 0$$

est décroissante dans l'intervalle $]0, +\infty[$. L'idée alors pour trouver les valeurs optimales de θ et de Δ est de parcourir avec un certain pas les valeurs de θ depuis une valeur minimale (1 par exemple) jusqu'à une valeur maximale ¹², et à chaque fois que θ vérifie que $g(\theta) \leq \exp(-1)$,

¹²4 devrait être une valeur largement suffisante.

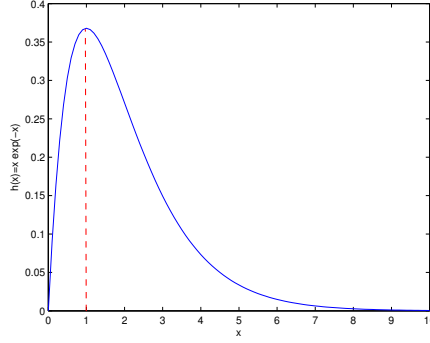


FIG. 3.2: Profil de la fonction $h(x) = x \exp(-x)$

on calcule alors la valeur de Δ correspondante. Celle-ci elle peut être calculée rapidement en minimisant la fonction unidimensionnelle unimodale suivante sur l'intervalle $]0, 1]$:

$$q(x) = (x \exp(-x) - g(\theta))^2, 0 < x \leq 1$$

Une telle minimisation peut se faire rapidement en utilisant l'algorithme de la recherche dorée (Golden Search). La première valeur de θ trouvée vérifiant $\theta^2 \geq \Delta$ est alors la valeur optimale recherchée.

3.6 La Sélection simultanée de modèles dans un cadre gaussien

Dans la section précédente, nous avons supposé que les données pour lesquelles on cherchait à choisir un modèle $\mathcal{S}_{\tilde{m}}$ pris dans une librairie de modèles $\{\mathcal{S}_m, m \in \mathcal{M}\}$ étaient toutes issues d'un même processus gaussien. Nous avons alors proposé deux critères pénalisés permettant le choix d'un tel modèle $\mathcal{S}_{\tilde{m}}$ tout en contrôlant la borne supérieure de l'estimateur pénalisé $\tilde{s} = \hat{s}_{\tilde{m}}$ correspondant.

Dans cette section, nous allons supposer que nous possédons L jeux de données $D_l, l = 1, \dots, L$ correspondant à L processus indépendants $Y_l, l = 1, \dots, L$, et chacun de ces processus satisfait au cadre gaussien suivant :

$$Y_i^l = s_i^l + \epsilon_i^l, i = 1, \dots, n^l; l = 1, \dots, L \tag{3.51}$$

où $\{s^l(\cdots), l = 1, \dots, L\}$ sont des fonctions (vecteurs) déterministes qui sont inconnues, et pour tout $l = 1, \dots, L$, on a $\epsilon_i^l, i^l = 1, \dots, n^l$ suivent un processus gaussien de moyenne nulle et de matrice de variances-covariances Σ_l supposée connue.

Le problème est alors d'estimer de façon simultanée; i.e. pour tous les $l = 1, \dots, L$ en même temps, les fonctions (vecteurs) déterministes $s^l(\cdot)$ à partir d'une réalisation de chacun des processus $Y^l, l = 1, \dots, L$ tout en contrôlant uniformément pour tout $l \in \{1, 2, \dots, L\}$ la statistique $L\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2$, ce qui revient donc à contrôler (à un terme de normalisation additif près comme on le verra plus loin dans cette section) la statistique

$$\sup_{l=1}^L \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 \quad (3.52)$$

où \tilde{s}^l est l'estimateur pénalisé de s^l , pour tout $l = 1, \dots, L$. On introduit alors pour chaque processus Y^l une famille de modèles linéaires $\{\mathcal{S}_m, m \in \mathcal{M}_l\}$ parmi laquelle sera choisi le modèle $\mathcal{S}_{\hat{m}^l}$. Pour choisir de tels modèles $\mathcal{S}_{\hat{m}^l}, l = 1, \dots, L$, on utilise un critère pénalisé et on définit \tilde{s}^l par $\tilde{s}^l \triangleq \hat{s}_{\hat{m}^l}$ l'estimateur pénalisé de s^l , pour tout $l = 1, \dots, L$. La théorie de l'inférence simultanée ¹³ nous fournit alors un cadre statistique qui permet de contrôler le sur-ajustement des données de façon uniforme (simultanée) pour tout $l = 1, \dots, L$.

En effet, lorsque le volume de données traité est trop grand, et comme le taux de sur-ajustement des données n'est pas nul pour un critère de sélection de modèle simple, alors en utilisant un tel critère, on peut se retrouver avec un nombre de sur-ajustements (ou de fausses détections de motifs ¹⁴) dans les données qui est trop grand. Un tel sur-ajustement des données est d'autant plus désastreux que le nombre de motifs existant réellement dans les données est petit car ça peut complètement fausser l'analyse par sélection de modèle.

Au fait, un tel constat signifie en des termes pratiques que le rapport signal/bruit (SNR) des données doit augmenter avec leur taille pour qu'une analyse par sélection de modèle soit efficace. Par exemple, dans un problème de segmentation de $L \gg 1$ signaux unidimensionnels simultanément en utilisant la sélection de modèle (*c.f.* la section 3.7), sup-

¹³La notion d'inférence simultanée [Bonf1925] connote généralement en statistique la correction de la "p-value" d'un test d'hypothèse dans le cas de tests multiples d'hypothèse qui est nécessaire pour éviter d'avoir trop de faux positifs[Bonf1925,Benj1985], et à notre connaissance, une telle notion n'a pas été utilisée auparavant en théorie de sélection de modèle pour adresser le cas de la sélection de plusieurs modèles simultanément pour plusieurs jeux de données différents, donc nous proposons dans ce qui suit la version "sélection de modèle" d'une telle notion.

¹⁴Par analogie aux fausses alarmes dans un problème de détection.

posant qu'on sache d'avance qu'il y a un très petit nombre p inconnu de segments intéressants tous signaux confondus, comme chaque signal a une certaine probabilité d'être sur-segmenté, alors sur l'ensemble de tous les signaux, on risque d'obtenir en utilisant une analyse par sélection de modèle simple, i.e. réalisée sur chaque signal indépendamment des autres, un nombre $\gg p$ de segments significatifs. Ceci est bien évidemment déconcertant pour nous puisque l'on ne saurait pas différencier les segments intéressants détectés des autres détectés à tort même lorsque le rapport signal/bruit des données est fort.

C'est alors cette idée de tenir compte de la nature simultanée de l'analyse par sélection de modèle, afin de ne détecter que les motifs "saillants" dans les données, que nous essayons de défendre et d'adresser dans cette section. Le théorème 3 suivant nous fournit alors un cadre statistique qui nous permet de corriger la fonction de pénalité pour atteindre un tel objectif. Comme nous avons prouvé dans la section 3.5 la supériorité du second critère de sélection de modèle sur le premier critère par la stratégie de choix des paramètres dans la fonction de pénalité que nous avons présentée dans la section précédente, nous nous contentons alors de donner la version "sélection simultanée de modèles" du second critère, mais le même raisonnement peut être utilisé pour étendre le premier critère de sélection de modèle au cas de la sélection simultanée de modèles.

Théorème 3 *On considère les processus Y^l définis dans (3.51), ainsi que leurs familles de modèles respectives $\mathcal{M}_l, l = 1, \dots, L$, et on associe aux modèles $\{\mathcal{S}_m, m \in \bigcup_{l=1, \dots, L} \mathcal{M}_l\}$ des poids positifs L_m vérifiant*

$$\Delta^* \triangleq \sum_{l=1}^L \sum_{m \in \mathcal{M}^l} \exp(-L_m D_m) < \infty \quad (3.53)$$

et soit deux réels $\theta > 0$ et $\alpha \in (0, 1)$.

Il s'ensuit que si la fonction de pénalité d'un modèle $\mathcal{S}_m, m \in \bigcup_{l=1, \dots, L} \mathcal{M}_l$ est définie comme suit :

$$\text{pen}(m) \geq \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m$$

alors, pour tout $m \in \bigcup_{l=1}^L \mathcal{M}_l$, il existe une variable aléatoire $r_m \leq 0$ p.s. et vérifiant que $\mathbb{E}[r_m] = -D_m$ telle qu'on ait

$$\mathbb{E} \left[\sup_{l=1}^L \left\{ \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1 - \alpha} \left[\inf_{m \in \mathcal{M}^l} \{ \|s^l - s_m^l\|_{\Sigma^{-1}}^2 + \text{pen}(m) + r_m \} \right] \right\} \right] \leq \left(\frac{2}{\alpha(1 - \alpha)} + \frac{2 - \alpha}{\theta(1 - \alpha)} \right) \Delta^*$$

■ **Preuve :** La preuve de ce théorème est quelque peu similaire à celle du théorème 2. En effet, en procédant comme le théorème 2, et en posant $\Delta_l \triangleq \sum_{m \in \mathcal{M}^l} \exp(-L_m D_m)$, on a pour tout $l = 1, \dots, L$

$$\mathbb{P}\left[Q_{\tilde{m}_l} > \left(\frac{2}{\alpha} + \frac{2-\alpha}{\theta}\right)\xi\right] \leq \Delta_l \exp(-\xi)$$

En définissant maintenant l'évènement

$$\zeta_l = \left\{Q_{\tilde{m}_l} \leq \left(\frac{2}{\alpha} + \frac{2-\alpha}{\theta}\right)\xi\right\}$$

et l'évènement simultané $\zeta = \bigcap_{l=1}^L \zeta_l$, ainsi que leurs évènements complémentaires respectifs ζ_l^c et ζ^c , on a

$$\mathbb{P}\left[\zeta^c\right] \leq \sum_{l=1}^L \mathbb{P}\left[\zeta_l^c\right] = \sum_{l=1}^L \Delta_l \exp(-\xi) = \Delta^* \exp(-\xi)$$

Il s'ensuit qu'en définissant R_m comme dans la preuve du théorème 2

$$\mathbb{P}\left[\bigcap_{l=1}^L \left\{\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 \leq \frac{1}{1-\alpha} \left(\inf_{m \in \mathcal{M}^l} \{R_m\} + \left(\frac{2}{\alpha} + \frac{2-\alpha}{\theta}\right)\xi\right)\right\}\right] \geq 1 - \Delta^* \exp(-\xi)$$

ou de façon équivalente

$$\mathbb{P}\left[\bigcap_{l=1}^L \left\{\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \left(\inf_{m \in \mathcal{M}^l} \{R_m\}\right) \leq \left(\frac{2}{\alpha} + \frac{2-\alpha}{\theta}\right)\xi\right\}\right] \geq 1 - \Delta^* \exp(-\xi)$$

d'où

$$\mathbb{P}\left[\sup_{l=1}^L \left\{\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}^l} \{R_m\}\right\} \leq \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)}\right)\xi\right] \geq 1 - \Delta^* \exp(-\xi)$$

ou encore

$$\mathbb{P}\left[\sup_{l=1}^L \left\{\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}^l} \{R_m\}\right\} \geq \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)}\right)\xi\right] \leq \Delta^* \exp(-\xi)$$

et en utilisant le fait que pour toute variable aléatoire V , on a si $\int_v \mathbb{P}(V > v) dv < \infty$, alors $\mathbb{E}(V) = \int_v \mathbb{P}(V > v) dv$, et en passant à l'espérance, on trouve que

$$\mathbb{E}\left[\sup_{l=1}^L \left\{\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}^l} \{R_m\}\right\}\right] \leq \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)}\right)\Delta^*$$

et en posant $r_m := R_m - d_m^2 - \text{pen}(m)$, qui vérifie que $\mathbb{E}[r_m] = -D_m$ et $r_m \leq 0$ p.s. (voir la preuve du théorème 2), on obtient

$$\mathbb{E} \left[\sup_{l=1}^L \left\{ \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}^l} \left\{ \|s^l - s_m^l\|^2 + \text{pen}(m) + r_m \right\} \right\} \right] \leq \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)} \right) \Delta^*$$

■

Le résultat du théorème 3 nous permet alors de contrôler (à un terme additif près) uniformément (i.e. pour tout $l = 1, \dots, L$) la statistique $\|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2$ par le biais d'un choix approprié de la fonction de pénalité qui est de la forme

$$\text{pen}(m) \geq \left((2-\alpha)(1+\theta) + \left(\frac{2-\alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m \quad (3.54)$$

3.6.1 Choix des paramètres L_m , θ et α dans la fonction de pénalité dans le cas de la sélection simultanée de modèles

Choix de L_m et θ

La proposition 3 suivante qui permet le choix des poids L_m , et de θ étant la version ‘simultanée’ de la proposition 2 :

Proposition 3 *On considère les processus Y^l définis dans (3.51), ainsi que leurs familles de modèles respectives $\mathcal{M}_l, l = 1, \dots, L$, et on associe aux modèles $\{\mathcal{S}_m, m \in \bigcup_{l=1, \dots, L} \mathcal{M}_l\}$ des poids positifs L_m donnés par*

$$L_m \equiv \theta^2 - \frac{\Delta^*}{D_m}$$

avec $\Delta^* > 0$, et $\theta > 0$ étant la plus petite valeur vérifiant

$$\theta^2 \geq \Delta^*$$

$$\Delta^* \exp(-\Delta^*) = \sum_{l=1}^L \sum_{D=1}^{D_{max}^l} \text{Card}_{\mathcal{M}_l}(D) \exp(-\theta^2 D)$$

où $\text{Card}_{\mathcal{M}_l}(D)$ représente le nombre de modèles $\{\mathcal{S}_m \in \mathcal{M}_l\}$ ayant la dimension D , et D_{max}^l est la dimension maximale autorisée pour un modèle $\mathcal{S}_m \in \mathcal{M}_l$. Soit maintenant un réel $\alpha \in (0, 1)$, il s'ensuit que si la fonction de pénalité d'un modèle $\mathcal{S}_m, m \in \bigcup_{l=1, \dots, L} \mathcal{M}_l$ est définie comme suit :

$$\text{pen}(m) = \left((2-\alpha)(1+\theta) + \left(\frac{2-\alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m$$

alors il existe une variable aléatoire $r_m \leq 0$ p.s. et vérifiant que $\mathbb{E}[r_m] = -D_m$ telle qu'on ait

$$\mathbb{E} \left[\sup_{l=1}^L \left\{ \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \left[\inf_{m \in \mathcal{M}^l} \left\{ \|s^l - s_m^l\|_{\Sigma^{-1}}^2 + \left(\frac{2\theta^2}{\alpha(1-\alpha)} + \frac{2\theta}{1-\alpha} + 2\theta + 2 \right) D_m + r_m \right\} \right] \right\} \right] \leq 0$$

■ Pour prouver la proposition 3, on utilise le théorème. Ce dernier nous garantit qu'avec un choix d'une fonction de pénalité de la forme (3.54), on a

$$\mathbb{E} \left[\sup_{l=1}^L \left\{ \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}^l} \left\{ \|s^l - s_m^l\|_{\Sigma^{-1}}^2 + \text{pen}(m) + r_m \right\} \right\} \right] \leq \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)} \right) \Delta^*$$

ou de façon équivalente

$$\mathbb{E} \left[\sup_{l=1}^L \left\{ \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 - \frac{1}{1-\alpha} \inf_{m \in \mathcal{M}^l} \left\{ \|s^l - s_m^l\|_{\Sigma^{-1}}^2 + \text{pen}(m) + \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)} \right) \Delta^* + r_m \right\} \right\} \right] \leq 0$$

et comme on garantirait que $\sup_{l=1}^L \left\{ \|s^l - \tilde{s}^l\|_{\Sigma^{-1}}^2 \right\}$ est d'autant plus petit que l'expression

$$\text{pen}(m) + \left(\frac{2}{\alpha(1-\alpha)} + \frac{2-\alpha}{\theta(1-\alpha)} \right) \Delta^*$$

l'est aussi, alors on tente de minimiser cette dernière pour tout $m \in \bigcup_{l=1}^L \mathcal{M}^l$ par rapport à θ , L_m et α . En procédant alors de façon similaire à la proposition 2 pour minimiser cette dernière expression, on obtient le résultat de la proposition 3.

Choix de α

Comme dans le cas de la sélection simple de modèle, la constante α dans la fonction de pénalité correspondant au critère de sélection simultanée de modèles est choisie pour réaliser un meilleur compromis biais-variance, elle est donc donnée par la formule

$$\alpha = \frac{\theta}{\sqrt{\theta^2 + \theta} + \theta}$$

3.7 Expériences numériques

Dans le but de démontrer la faisabilité de notre approche en pratique, nous proposons de traiter dans cette section un problème classique en sélection de modèle qui est celui de la détection des points de changement dans la moyenne d'un signal gaussien unidimensionnel connu aussi comme le problème de sélection d'histogramme. Nous nous restreignons

ici au second critère de sélection de modèle que nous avons proposé dans sa version simple et simultané qui a démontré sa supériorité sur le premier par la stratégie de choix des poids et des constantes dans la fonction de pénalité que nous avons présentée plus haut.

Le problème étant donc le suivant. On considère un processus unidimensionnel $Y = (Y_t)_{t=1, \dots, T}$, où l'indice t représente le temps, satisfaisant au cadre gaussien suivant :

$$Y_i = s_i + \epsilon_i, i = 1, \dots, N$$

où s est la moyenne (partie déterministe) du processus (vecteur) Y , et $\epsilon_i, i = 1, \dots, N$ sont des gaussiennes centrées d'écart-type σ . On suppose que la moyenne du signal Y peut subir à des instants inconnus τ_1, \dots, τ_k pendant l'intervalle de temps $[1, N]$ k changements (k étant inconnu également), sachant que dans l'intervalle de temps $]\tau_k, \tau_{k+1}]$ entre deux points de changement successifs τ_k et τ_{k+1} , la moyenne du signal reste constante dans cet intervalle. Le problème est alors d'estimer le nombre et la localisation des points de changement dans la moyenne du processus Y à partir d'une réalisation $\{y_i, i = 1, \dots, N\}$ de ce dernier.

3.7.1 Formulation du problème d'estimation des points de ruptures dans un signal 1D comme un problème de sélection de modèle

On peut alors formuler un tel problème comme un problème de sélection de modèle. Un modèle serait alors tout sous-espace de l'espace linéaire des fonctions constantes par morceaux dans l'intervalle $[1, N]$. On note un tel espace (ou famille de modèles) $\{\mathcal{S}_m, m \in \mathcal{M}\}$, où m désigne une subdivision (partition) de l'intervalle de temps $[1, N]$. On symbolise une subdivision temporelle m par rapport à ses points de changement successifs comme suit : $m = (\tau_k)_{k=1, \dots, |m|}$. La dimension D_m d'un modèle \mathcal{S}_m est donc donnée par le nombre de segments de sa subdivision spatio-temporelle m , autrement-dit $D_m = |m|$, et on suppose que $1 \leq D_m \leq D_{max}$. Par convention, on considère que quel que soit $m \in \mathcal{M}$, on a $\tau_1 \equiv 0$ et $\tau_{|m|} \equiv T$. On écrit alors un modèle \mathcal{S}_m comme suit :

$$\mathcal{S}_m(\cdot) = \sum_{k=1}^{|m|} U_k \mathbb{I}_{(\tau_{k-1}, \tau_k]}(\cdot)$$

où $\{U_k, k = 1, \dots, |m|-1\}$ sont des coefficients réels. Comme on considère le cas homoscedastique, alors l'estimateur de projection \hat{s}_m correspondant au modèle \mathcal{S}_m est celui minimisant

le critère quadratique sur \mathcal{S}_m , i.e.

$$\hat{s}_m = \arg \min_{t \in \mathcal{S}_m} \sum_{i=1}^N (y_i - t_i)^2$$

et s'exprime donc pour tout $t = 1, 2, \dots, T$ comme suit :

$$\hat{s}_m(i) = \sum_{k=1}^{|m|} \bar{y}_k \mathbb{I}_{(\tau_{k-1}, \tau_k]}(i)$$

avec

$$\bar{y}_k = \frac{1}{\tau_k - \tau_{k-1}} \sum_{i=\tau_{k-1}+1}^{\tau_k} y_i$$

et le critère pénalisé, en utilisant le second critère pénalisé (SCP) qu'on a présenté plus haut, s'écrit comme suit :

$$\text{Crit}(\hat{s}_m) = \gamma(\hat{s}_m) + \text{pen}(m) \quad (3.55)$$

où

$$\text{pen}(m) \triangleq \text{pen}(D_m) \equiv \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m$$

avec :

- Si on considère un seul processus Y à la fois (sélection de modèle simple) :

$$L_m \equiv \theta^2 - \frac{\Delta}{D_m}, \theta \text{ est la plus petite valeur positive vérifiant}$$

$$\theta^2 \geq \Delta$$

$$\Delta \exp(-\Delta) = \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-\theta^2 D)$$

$$\text{et } \alpha = \frac{\theta}{\sqrt{\theta^2 + \theta + \theta}}.$$

- Si on considère L processus Y^l simultanément (sélection simultanée de modèles) :

$$L_m \equiv \theta^2 - \frac{\Delta^*}{D_m}$$

avec $\Delta^* > 0$, et $\theta > 0$ étant la plus petite valeur vérifiant

$$\theta^2 \geq \Delta^*$$

$$\Delta^* \exp(-\Delta^*) = \sum_{l=1}^L \sum_{D=1}^{D_{max}^l} \text{Card}_{\mathcal{M}_l}(D) \exp(-\theta^2 D)$$

$$\text{et } \alpha = \frac{\theta}{\sqrt{\theta^2 + \theta + \theta}}.$$

Etant donné que la pénalité ne dépend que de la dimension d'un modèle, alors, pour minimiser (3.55), on calcule pour $D = 1, \dots, D_{max}$ les estimateurs $\hat{s}_D = \arg \min_{\hat{s}_m, |m|=D} \|y - \hat{s}_m\|^2$ ¹⁵, puis on minimise sur $D = 1, \dots, D_{max}$ la quantité $\gamma(\hat{s}_D) + \text{pen}(D)$. Pour ce qui est des quantités $\text{Card}_{\mathcal{M}}(D)$, celles-ci peuvent se calculer par le biais de la formule récursive suivante :

- $c_{N,d} = 0$; si $d > 1$
- $c_{k,1} = 1$, quel que soit $k = 1, \dots, N$
- $c_{k,d} = c_{k+1,d} + c_{k+1,d-1}$
- $\text{Card}_{\mathcal{M}}(D) = c_{1,D}$

3.7.2 Procédure de simulation des signaux 1D

Nous fixons N à une valeur donnée et nous simulons un signal 1D gaussien de taille N comme suit :

- La dimension (ou le nombre de segments) du signal suit une Poisson $\mathcal{P}(10)$.
- Pour une dimension D donnée d'un signal, la localisation de ses $D - 1$ points de rupture suit une loi uniforme $U[1, N]$.
- La moyenne du signal dans chacun de ses D segments suit loi gaussienne de moyenne $\mu = 2$ et d'écart-type $\sigma = 2$. On fait attention à ce que la différence entre deux niveaux successifs soit significative (on la prend supérieure ou égale à 2) pour qu'on puisse interpréter proprement nos résultats.

3.7.3 Résultats obtenus et leur interprétation

On simule à la fois une liste de taille L de tels signaux unidimensionnels $Y^l, l = 1, \dots, L$, et on procède à la détection de rupture par sélection de modèle comme on l'a décrit dans la sous-section précédente en utilisant le deuxième critère de sélection de modèle que nous avons présenté plus haut, d'abord, dans sa version "sélection de modèle simple", puis dans sa version "sélection simultanée de modèles". On réalise une telle expérience pour différentes valeurs de L ($L = 100, 200, 1000$) et pour chaque expérience, on calcule les quantités

¹⁵ces calculs de $\hat{s}_D, D = 1, \dots, D_{max}$ peuvent se faire en un temps polynomial en utilisant la programmation dynamique [Kay1998]

empiriques suivantes :

- La dimension moyenne \bar{D}^L donnée par

$$\bar{D}^L = \frac{1}{L} \sum_{l=1}^L \tilde{D}_l^L$$

- L'indice de sur-ajustement moyen D_x^{L+} :

$$D_x^{L+} = \frac{1}{L} \sum_{l=1}^L (\tilde{D}_l^L - D_l^L) \mathbb{I}_{(0, \infty]}(\tilde{D}_l^L - D_l^L)$$

- Le risque empirique moyen $\bar{\mathcal{R}}^L$:

$$\bar{\mathcal{R}}^L = \frac{1}{L} \sum_{l=1}^L \frac{\|\tilde{s}^l - s^l\|^2}{N\sigma^2}$$

Voici les résultats qu'on a obtenus :

- $L = 100$

| Nom du critère | \bar{D}^L | D_x^{L+} | $\bar{\mathcal{R}}^L$ |
|----------------|-------------|------------|-----------------------|
| <i>BIC</i> | 11.880 | 1.520 | 0.327 |
| <i>SCP</i> | 10.40 | 0.365 | 0.310 |
| <i>SCPS</i> | 9.170 | 0.060 | 0.327 |

- $L = 500$

| Nom du critère | \bar{D}^L | D_x^{L+} | $\bar{\mathcal{R}}^L$ |
|----------------|-------------|------------|-----------------------|
| <i>BIC</i> | 12.498 | 1.632 | 0.370 |
| <i>SCP</i> | 10.658 | 0.380 | 0.321 |
| <i>SCPS</i> | 8.814 | 0.004 | 0.372 |

- $L = 1000$

| Nom du critère | \bar{D}^L | D_x^{L+} | $\bar{\mathcal{R}}^L$ |
|----------------|-------------|------------|-----------------------|
| <i>BIC</i> | 12.370 | 1.473 | 0.361 |
| <i>SCP</i> | 10.717 | 0.401 | 0.321 |
| <i>SCPS</i> | 8.652 | 0.001 | 0.384 |

- $L = 5000$

| Nom du critère | \bar{D}^L | D_x^{L+} | $\bar{\mathcal{R}}^L$ |
|----------------|-------------|------------|-----------------------|
| <i>BIC</i> | 12.294 | 1.540 | 0.358 |
| <i>SCP</i> | 10.5846 | 0.377 | 0.315 |
| <i>SCPS</i> | 8.114 | 0.001 | 0.426 |

On remarque d'abord que le critère *BIC* a tendance à sur-estimer la dimension d'un signal tandis que notre critère de sélection de modèle simple *SCP* l'estime efficacement et réalise un sur-ajustement moyen qui est environ quatre fois plus petit que celui que réalise le *BIC*.

Notre critère de sélection simultanée de modèles *SCPS* permet alors de corriger une telle sur-estimation de la dimension d'un modèle, et les tableaux ci-dessus montrent que quelle que soit le nombre de signaux analysés simultanément, l'indice de sur-ajustement moyen est toujours stabilisé pas loin de zéro. Une telle correction est nécessaire pour se débarrasser des nombreuses fausses détections inhérentes au caractère simultané de l'analyse.

Seulement dans le but de pouvoir apprécier visuellement la performance du critère de sélection de modèle sur le problème d'estimation de fonctions constantes par morceaux, nous montrons sur la figure 3.3 quelques résultats.

3.8 Conclusion

Dans ce chapitre, nous avons étudié la sélection de modèle dans un cadre gaussien basée sur la pénalisation d'un critère de type moindres carrés. Nous avons alors proposé deux nouveaux critères garantissant une borne supérieure du risque qui est serrée pour la sélection de modèle dans un cadre gaussien simple, i.e. lorsque les données sont issues d'un unique processus gaussien. Nous avons ensuite proposé une stratégie optimale (dans le sens de la borne supérieure du risque) de choix des poids et des constantes dans la fonction de pénalité, ainsi que des méthodes numériques efficaces permettant leur calcul. Puis, nous avons étendu un tel résultat en utilisant la théorie de l'inférence simultanée au cas de la sélection simultanée de modèles; i.e. lorsque l'on dispose de plusieurs jeux de données issues de plusieurs processus gaussiens indépendants et que l'on cherche à ajuster de façon simultanée un modèle à chacun de ces jeux données choisis, chaque modèle étant choisi parmi une librairie de modèles qui peut être la même ou différente pour les différents jeux de données, et ce, dans le but de contrôler le nombre de sur-ajustements des données indépendamment du nombre de processus ayant généré les données. Un tel résultat de sélection simultanée de modèles va nous permettre dans le chapitre 3 prochain d'effectuer une segmentation spatio-temporelle globale efficace d'une séquence de vidéo-microscopie à partir de plusieurs segmentations spatio-temporelles locales tout en contrôlant de façon automatique le nombre de fausses alarmes.

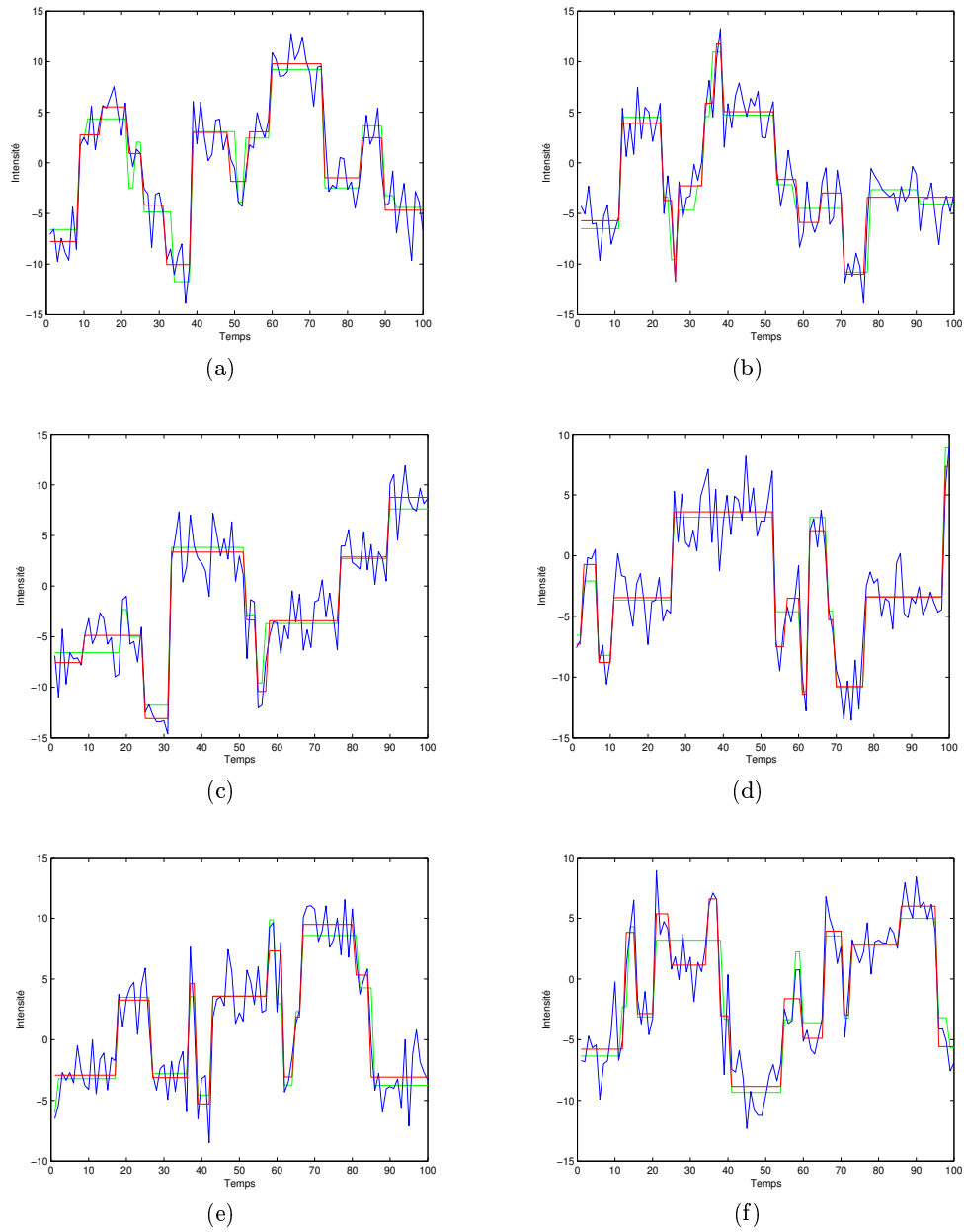


FIG. 3.3: Quelques résultats d'estimation par sélection de modèle de fonctions constantes par morceaux en utilisant le second critère pénalisé (SCP) que nous avons proposé. En bleu le signal original, en vert sa partie déterministe, et en rouge son estimateur pénalisé.

Chapitre 4

Segmentation spatio-temporelle par sélection de modèle d'une séquence de vidéo-microscopie. Application à la détection automatique de vésicules

Dans ce chapitre, nous présentons une approche de segmentation spatio-temporelle d'une séquence de vidéo-microscopie de fluorescence $3D + T$ basée sur l'approche de sélection de modèle décrite dans le chapitre 3 précédent, et nous l'appliquons à la détection automatique de vésicules avec contrôle automatique du nombre de fausses alarmes.

En effet, pour segmenter tout le volume spatio-temporel d'une séquence de vidéo-microscopie, notre approche procède d'abord par diviser le volume spatial en plusieurs blocs de taille fixe, ensuite elle réalise des segmentations spatio-temporelles dans chaque bloc par ajustement de modèles spatio-temporels constants par morceaux. Pour ce faire, la théorie de sélection de modèle est utilisée, et pour contrôler automatiquement le nombre de fausses segmentations spatio-temporelles dans tout le volume spatio-temporel de la séquence, notre approche utilise le concept de sélection simultanée de modèles décrit dans le chapitre 3 précédent. Nous résolvons numériquement le problème de manière efficace par le biais de la programmation dynamique.

Nous appliquons ensuite l'approche développée dans ce chapitre pour réaliser une détection automatique des vésicules (objets biologiques contrastés en mouvement) dans une séquence de vidéo-microscopie. Pour ce faire, nous réalisons d'abord la correction d'intensité, la stabilisation de la variance et la soustraction du comme c'est décrit en détails dans le chapitre 2. Ensuite, nous réalisons la segmentation spatio-temporelle de la séquence qui est détaillée dans ce chapitre. Les vésicules en mouvement dans la séquence sont alors détectés dans l'espace spatio-temporel comme des régions spatio-temporelles subissant un changement significatif dans le niveau de leur intensité. La projection ensuite de tous les voxels détectés comme significatifs dans le volume spatial à chaque instant et leur regroupement perceptuel permettra de segmenter les vésicules qui sont uniquement en mouvement dans la séquence.

Nous démontrons l'intérêt d'une telle approche aussi bien sur des données synthétiques que sur des données image réelles issues de la vidéo-microscopie de fluorescence $3D + T$ champ-large ¹.

4.1 Approche de sélection de modèle pour la détection d'évènements en vidéo-microscopie de fluorescence $3D + T$

Nous nous intéressons dans ce projet à l'étude de la dynamique due au trafic membranaire de vésicules transportant des macro-molécules, à savoir des protéines de type RAB ², entre les différents compartiments cellulaires de la cellule (Golgi, Réticulum endoplasmique (ER), etc...). Il est alors important de pouvoir détecter ce trafic et de le quantifier dans la séquence de vidéo-microscopie avant d'envisager une analyse descriptive du contenu dynamique du processus d'échange intra-cellulaire étudié.

Pour ce faire, notre idée de base pour la détection spatio-temporelle de mouvement

¹Séquences d'images réalisées par nos collaborateurs (équipe de Jean Salamero) de l'Institut Curie de Paris.

²On pense que de telles protéines sont responsables de la plupart des échanges intra-cellulaires entre l'appareil de Golgi et le réticulum endoplasmique (ER). Donc la compréhension de ce processus d'échange de la matière entre les différents compartiments cellulaires passe par la compréhension de la dynamique des protéines impliquées dans le trafic membranaire. Pour plus de détails biologiques sur le sujet, le lecteur intéressé pourra se référer à la page web de nos collaborateurs dans le projet de l'Institut Curie de Paris (équipe de J. Salamero) et les références qui y sont citées :
http://www.curie.fr/recherche/themes/equipe_publi.cfm/lang/_fr/id_equipe/296/pre/1.htm

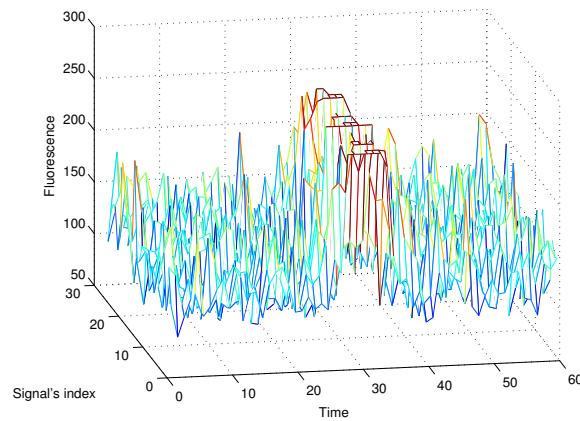


FIG. 4.1: Un paquet de 27 signaux voisins dans une séquence de vidéo-microscopie de fluorescence $3D+T$. La figure montre un changement significatif dans la fluorescence de certains des 27 voxels entre l'instant 30 et l'instant 40 dû au trafic de vésicules dans la séquence.

en vidéo-microscopie de fluorescence est qu'un objet (un spot fluorescent) dans la scène de vidéo-microscopie, lorsqu'il traverse une petite région du volume $3D$, il fait passer le signal³ de cette région d'un niveau de signal bas (qui correspond au signal d'autofluorescence dans cette région) à un niveau signal haut (qui correspond au signal de l'objet contrasté en mouvement dans la séquence traversant cette région) (cf. figure 4.1). Mais au lieu d'analyser les signaux individuellement, notre approche procède de manière robuste en analysant plusieurs signaux voisins (i.e. provenant d'une même région de l'image) en même temps. La force d'une telle approche réside dans sa prise en compte à la fois de la cohérence spatiale et de la cohérence temporelle du signal. Cette façon procéder est certainement plus robuste au bruit dans l'image qu'une autre approche ne prenant en compte qu'une seule des dimensions (temporelle ou spatiale) de l'images, et constitue un défi par rapport au très faible rapport signal-bruit (SNR) des séquences d'images issues de la vidéo-microscopie de fluorescence. En effet, on peut réaliser les deux expériences suivantes :

1. **Expérience 1** : On crée une image avec un fond très bruité, et on y place un spot fluorescent avec un rapport signal bruit assez faible, de sorte que si on présente cette image à une personne, celle-ci aura du mal à isoler le spot du fond de l'image.

³On entend par signal le profil temporel de la fluorescence d'un voxel donné de l'image $3D$

2. **Expérience 2** : Cette fois, on simule un mouvement de l'objet fluorescent dans le fond de l'image, alors sans changer son rapport signal bruit, si on présente cette fois la vidéo résultante à la même personne que pour l'expérience 1, cette dernière peut facilement reconnaître l'objet mobile dans le fond de l'image.

Cette expérience, bien que simple, montre que la prise en compte à la fois de la cohérence spatiale et de la cohérence temporelle du signal dans l'analyse d'une séquence d'images permet en quelque sorte d' "augmenter" le rapport signal bruit de la séquence. Une telle démarche est d'un grand intérêt en analyse de séquences de vidéo-microscopie, car elle permet de défier leur rapport signal bruit très faible, et par là, rajouter de la robustesse aussi bien à l'étape de détection de spots fluorescents en mouvement dans ces séquences, qu'à l'étape l'analyse quantitative de leur dynamique.

Dans la suite de cette section, nous allons montrer comment un tel problème de détection spatio-temporelle de passages de spots fluorescents dans des régions de l'image préalablement choisies peut se ramener à un problème de sélection simultanée de modèles dans l'espace spatio-temporel. L'utilisation alors de nos résultats du chapitre précédent permettra de résoudre efficacement le problème en évitant le problème de sur-ajustement qui se poserait avec une approche classique. Nous présenterons par la suite l'algorithme dynamique que nous avons développé pour résoudre le problème numériquement, ainsi que les résultats que nous avons obtenus aussi bien sur des données synthétiques que sur des données image issues de la vidéo-microscopie de fluorescence $3D + T$ qui concernent le trafic intra-cellulaire de vésicules transportant des protéines de la famille RAB entre les différents compartiments cellulaires.

4.1.1 Formulation du problème dans un cadre Gaussien

Nous commençons d'abord par diviser l'image $3D$ en de petits blocs de taille M (par exemple $M = 3 \times 3 \times 2$), et pour chacun de ces blocs, nous observons le signal de ses M voxels simultanément. Notre but est alors de détecter le passage de spots fluorescents dans ces blocs éventuellement à plusieurs instants de la séquence. De tels passages de spots fluorescents dans un bloc de l'image correspondent donc à des changements significatifs que subissent simultanément plusieurs signaux parmi les M signaux du bloc considéré.

Pour cela, on considère alors le groupe des M signaux unidimensionnels $Y_i, i = 1, \dots, M$ d'un bloc donné de l'image qui sont supposés être des vecteurs aléatoires Gaussiens à N composantes temporelles : $Y_i = (Y_{i,t})_{(t=1, \dots, N)}$ quel que soit $i = 1, \dots, M$. De tels vecteurs aléatoires satisfont au cadre Gaussien suivant :

$$Y_i(t) = S_i(t) + \epsilon_i(t), \quad i = 1, \dots, M, \quad t = 1, \dots, N \quad (4.1)$$

où $S_i(t)$ est la partie déterministe (valeur moyenne) du signal Y_i à l'instant t et $\epsilon_i(t)$ est un bruit blanc Gaussien de moyenne nulle et de variance $\sigma_i(t)^2$ supposée connue. Ce modèle suppose donc une matrice de variances-covariances Σ du processus Y qui est diagonale.

Construction de modèles de S , calcul des estimateurs de projection de S

On suppose que la partie déterministe $S = \{S_i(t), i = 1, \dots, M; t = 1, \dots, N\}$ des signaux $Y_i, i = 1 \dots, M$ peut subir des changements aléatoires temporellement et spatialement. Ceci signifie qu'il existe H intervalles de temps aléatoires (H étant aléatoire), disjoints $[t_1^h, t_2^h] \subseteq [1, N], h = 1, \dots, H$ et dont la réunion donne l'intervalle $[1, N]$, et tels que dans chaque intervalle de temps $[t_1^h, t_2^h]$, on ait :

- Un premier groupe de signaux dont les indices aléatoires sont $i_{h,1}^{(1)}, i_{h,2}^{(1)}, \dots, i_{h,k}^{(1)} \in [1, \dots, M]$ subissent simultanément un même changement (aléatoire) dans leur moyenne, i.e. :

$$S_{h,i_{h,1}^{(1)}}(t) \equiv S_{h,i_{h,2}^{(1)}}(t) \equiv \dots \equiv S_{h,i_{h,k}^{(1)}}(t) \equiv S_{h,1}, \forall t \in [t_1^h, t_2^h]$$

- Un deuxième groupe de signaux dont les indices aléatoires sont $i_{h,1}^{(2)}, i_{h,2}^{(2)}, \dots, i_{h,k}^{(2)}$ subissent simultanément un même changement m' (aléatoire) dans leur moyenne, i.e. :

$$S_{h,i_{h,1}^{(2)}}(t) \equiv S_{h,i_{h,2}^{(2)}}(t) \equiv \dots \equiv S_{h,i_{h,k}^{(2)}}(t) \equiv S_{h,2}, \forall t \in [t_1^h, t_2^h]$$

- ...

- Un $n_h \equiv |m_h|$ -ième (n_h étant aléatoire également) groupe de signaux dont les indices sont $i_{h,1}^{(n_h)}, i_{h,2}^{(n_h)}, \dots, i_{h,k}^{(n_h)} \in [1, \dots, M]$ subissent un changement (aléatoire) dans leur moyenne, i.e. :

$$S_{h,i_{h,1}^{(n_h)}}(t) \equiv S_{h,i_{h,2}^{(n_h)}}(t) \equiv \dots \equiv S_{h,i_{h,k}^{(n_h)}}(t) \equiv S_{h,(n_h)}, \forall t \in [t_1^h, t_2^h]$$

A la lumière de ce qu'on vient de dire à propos de la partie déterministe S du vecteur spatio-temporel aléatoire Y , il correspond à S une configuration spatio-temporelle m qui consiste en une partition temporelle de H intervalles de temps $\tau_h, h = 1, \dots, H$, et pour chaque intervalle de temps τ_h , il correspond une partition spatiale consistant en $|m_h|$ ensembles d'indices spatiaux $\{m_h = \bigcup_{j=1}^{|m_h|} m_{h,j}\}$, de sorte qu'on peut écrire le modèle de S qui est donc un modèle constant par morceaux dans l'espace spatio-temporel ⁴ comme suit : $\forall i = 1 \dots, M, \forall t = 1, \dots, N$,

$$S_i(t) = \sum_{h=1}^H \sum_{j=1}^{|m_h|} S_{h,j} \mathbb{1}_{m_{h,j}}(i) \mathbb{1}_{\tau_h}(t) \quad (4.2)$$

où $S_{h,j}$ désigne le niveau du signal commun à tous les indices due j -ième ensemble d'indices du h -ième intervalle de temps de la configuration spatio-temporelle m de S .

On considère que le nombre d'ensembles d'indices spatiaux correspondant à chaque intervalle de temps de la partition spatio-temporelle de S est borné par un nombre positif p , i.e. :

$$|\tau_h| \leq p, \forall h = 1, \dots, H$$

Même si l'approche qu'on va décrire dans la suite reste valable pour p quelconque (moyennant un peu plus de complexité algorithmique qui est polynomiale en p), pour les besoins de notre étude, nous nous limitons au cas où $p \in \{1, 2\}$ car on s'intéresse à deux états : "changement dans la moyenne du signal", et "pas de changement dans la moyenne du signal".

Il est clair d'après l'écriture (4.2) qu'un modèle \mathcal{S}_m de S est un modèle linéaire (ou espace linéaire) indexé par une configuration spatio-temporelle $m \in \mathcal{M}$, où \mathcal{M} désigne l'ensemble de toutes les configurations spatio-temporelles possibles. On note

$$\mathcal{S} = \left\{ \mathcal{S}_m, m \in \mathcal{M} \right\}$$

la famille des modèles de S .

Un modèle linéaire de S étant alors caractérisé par sa base linéaire

$$\left\{ \mathbb{1}_{m_{h,j}}(\cdot) \mathbb{1}_{\tau_h}(\cdot), h = 1, \dots, H, j = 1, \dots, |m_h| \right\}$$

⁴Un modèle constant par morceaux n'est peut être pas un modèle idéal pour la détection de passages de spots fluorescents en vidéo-microscopie de fluorescence, mais c'est tout de même un très bon modèle approximatif, et il engendre une complexité calculatoire très raisonnable. Bien évidemment, mis à part l'algorithme dynamique de calcul de l'estimateur pénalisé que nous présenterons plus tard dans ce chapitre, toute la théorie développée ici est valable pour tout autre modèle spatio-temporel linéaire comme les ondelettes (spatio-temporelles).

Il s'ensuit que la dimension d'un modèle \mathcal{S}_m se définit comme suit :

$$D_m = \sum_{h=1}^H |m_h| \triangleq |m|$$

On définit maintenant la quantité $\gamma(\cdot)$ suivante pour tout vecteur $U \in \mathcal{S}$

$$\gamma(U) = \sum_{i=1}^M \sum_{t=1}^N \frac{(U_i(t) - Y_i(t))^2}{\sigma_{i,t}^2}$$

ou de façon équivalente (si on veut se conformer à la définition de la quantité $\gamma(\cdot)$ du chapitre 3) la quantité

$$\gamma'(U) = \sum_{i=1}^M \sum_{t=1}^N \frac{U_i(t)^2}{\sigma_{i,t}^2} - 2 \sum_{i=1}^M \sum_{t=1}^N \frac{U_i(t)Y_i(t)}{\sigma_{i,t}^2}$$

et on définit l'estimateur de projection de S par rapport à un modèle \mathcal{S}_m comme suit :

$$\hat{S}_m = \arg \min_{U \in \mathcal{S}_m} \gamma(U)$$

on obtient alors l'expression \hat{S}_m (qui n'est autre que son estimateur moyenne sur la configuration m) : $\forall i = 1 \dots, M, \forall t = 1, \dots, N$,

$$(\hat{S}_m)_i(t) = \sum_{h=1}^H \sum_{j \in m_h} \bar{Y}_{h,j} \mathbb{I}_{\tau_{h,j}}(i) \mathbb{I}_{\tau_h}(t)$$

où

$$\bar{Y}_{h,j} = \frac{1}{\sum_{t \in \tau_h} \sum_{i \in m_{h,j}} \frac{1}{\sigma_{i,t}^2}} \sum_{t \in \tau_h} \sum_{i \in m_{h,j}} \frac{Y_i(t)}{\sigma_{i,t}^2}$$

Un critère pénalisé pour l'estimation de S

Nous utilisons donc notre résultat du chapitre ?? pour estimer la meilleure configuration \tilde{m} , et par là, la partie déterministe S du signal spatio-temporel⁵ Y . Ce résultat du chapitre 2 nous dit que pour estimer efficacement la partie déterministe S du signal spatio-temporel observé correspondant à chaque bloc de l'image, il faut à la fois tenir compte de la richesse de la famille de modèles $\{\mathcal{S}_m, m \in \mathcal{M}\}$ de S , ainsi que de la nature multi-processus des données observées. Comme nous avons divisé l'image en L blocs, et que nous observons le signal spatio-temporel dans chacun de ces L blocs, nous avons alors la réalisation

⁵Ici et dans toute la suite, nous entendons par signal spatio-temporel un paquet de signaux voisins unidimensionnels

d'un nombre L de processus spatio-temporels de taille $M \times N$. Il est donc important de tenir compte de cette nature multi-sources des données observées pour éviter le problème de sur-ajustement inhérent à cette multiplicité de la source des données. Comme les L signaux spatio-temporels sont tous supposés de la même taille ($M \times N$), et qu'ils ont tous la même famille de modèles $\{\mathcal{S}_m, m \in \mathcal{M}\}$ ⁶, alors pour chaque bloc de l'image, nous estimons la partie déterministe S du signal spatio-temporel corerspondant en minimisant le critère quadratique pénalisé suivant

$$\text{Crit}(\hat{S}_m) = \gamma(\hat{S}_m) + \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m \quad (4.3)$$

où les poids L_m sont donnés par la formule

$$L_m \equiv \theta^2 - \frac{\Delta}{D_m}$$

avec θ et Δ sont les plus petites valeurs vérifiant

$$\theta > 0$$

$$\Delta \in (0, 1)$$

$$\Delta \exp(-\Delta) = L \sum_{D=1}^{D_{max}} \text{Card}_{\mathcal{M}}(D) \exp(-\theta^2 D)$$

et L le nombre de blocs de l'image, et $\alpha = \frac{\theta}{\sqrt{\theta^2 + \theta} + \theta}$.

Un algorithme dynamique pour le calcul de l'estimateur pénalisé \tilde{S} de S

Notre but est de détecter le(s) passage(s) d'objets fluorescents dans chaque bloc de l'image. Pour atteindre un tel objectif dans le cadre de sélection de modèle que nous venons d'établir, il est alors important de pouvoir introduire cette information *a priori* (qui

⁶Il faut souligner que cette dernière affirmation n'est pas toujours vraie. En effet, si nous voulons par exemple rajouter de l'*a priori* (au vu des données par exemple) sur la classe du modèle d'un signal spatio-temporel de chaque bloc, il se peut alors qu'on se retrouve avec des familles de modèles $\{\mathcal{S}_m, m \in \mathcal{M}_l, l = 1, \dots, L\}$ qui sont toutes différentes. C'est par exemple le cas ici comme nous allons le voir après la présentation de notre algorithme dynamique de calcul numérique des estimateurs pénalisés $\{\tilde{S}_l, l = 1, \dots, L\}$. Mais comme l'idée de l'algorithme garantit un nombre de modèles pour chaque dimension D d'un modèle qui est le même quel que soit le bloc considéré, alors pour simplifier l'écriture, on peut considérer une famille virtuelle de modèles $\{\mathcal{S}_m, m \in \mathcal{M}\}$, et écrire que le nombre de modèles ayant la dimension D pour chaque processus spatio-temporel est égal à $\text{Card}_{\mathcal{M}}(D)$, il s'ensuit qu'au total (i.e. pour tous les blocs de l'image confondus), on a une richesse de $L \text{Card}_{\mathcal{M}}(D)$ de modèles ayant la dimension D quel que soit $D = D_{min}, \dots, D_{max}$.

consiste en quels motifs on cherche dans les données) dans la famille de modèles candidats \mathcal{S} . Cela revient donc à choisir de façon appropriée l'ensemble des configurations spatio-temporelles \mathcal{M} qui sont susceptibles d'être la configuration de la partie déterministe du signal $Y = \{Y_i(t), i = 1 \dots, M, t = 1, \dots, N\}$. Une telle introduction d'une information *a priori* possède les trois avantages suivants :

1. Ecarte les modèles qui ne correspondent pas aux données pour obtenir un estimateur de S qu'on saura interpréter et exploiter pour les étapes d'analyse suivantes.
2. Limite la famille des modèles candidats \mathcal{S} , et par là, réduit le risque de l'estimateur pénalisé, ce qui se traduit donc par une estimation plus efficace de S .
3. Permet de concevoir un algorithme efficace de minimisation du critère pénalisé. En effet, si l'on ne prend pas compte de cette information *a priori*, il est impossible dans notre cas de trouver un algorithme ayant une complexité algorithmique raisonnable (complexité polynomiale par exemple) pour résoudre notre problème de sélection de modèle.

L'introduction de cette information *a priori* dans la famille de modèles \mathcal{S} est fondée sur l'idée suivante : "Si on suppose que l'intensité de l'objet qui traverse une région (ou bloc) de l'image est constante pour tous les voxels de l'objet, alors à chaque instant du passage, l'objet divise les voxels du bloc en deux classes : une première classe caractérisée par une intensité moyenne haute dont les voxels appartiennent à l'objet, et une deuxième classe caractérisée par une intensité moyenne basse dont les voxels appartiennent au fond de l'image correspondant au bloc."

On peut alors inclure cette information *a priori* dans la famille de modèles \mathcal{S} en introduisant une fonction "rank" qui ordonne les voxels d'un bloc à un instant donné (ou pendant une plage de temps donnée) d'une configuration spatio-temporelle de S par rapport à leur intensité moyenne à cet instant (respectivement dans la plage de temps considérée). La fonction "rank" est alors indexée par l'ensemble des intervalles de temps τ inclus dans l'intervalle $[1, N]$, de sorte que $\text{rank}_\tau(i)$ désigne le rang du voxel i dans l'intervalle de temps τ . On a alors

$$\begin{aligned} \text{rank} : \{ \tau \subseteq [1, N] \} \times \{ 1, \dots, M \} &\longrightarrow \{ 1, \dots, M \} \\ (\tau, i) &\longmapsto \text{rank}_\tau(i) \end{aligned}$$

et

$$\begin{aligned} \text{rank}_\tau(i) = k \Leftrightarrow & \quad \exists j_i^1 \neq j_i^2 \neq \dots \neq j_i^M \in \{1, \dots, M\} \quad \text{tq.} \\ & \frac{1}{|\tau|} \sum_{t \in \tau} S_i(t) \geq \frac{1}{|\tau|} \sum_{t \in \tau} S_j(t), \forall j \in \{j_i^1, \dots, j_i^{k-1}\} \\ & \text{et } \frac{1}{|\tau|} \sum_{t \in \tau} S_i(t) \leq \frac{1}{|\tau|} \sum_{t \in \tau} S_j(t), \forall j \in \{j_i^{k+1}, \dots, j_i^M\} \end{aligned}$$

Avec l'introduction de cette fonction "rank", il devient alors possible via son utilisation de définir une subdivision spatio-temporelle m seulement à partir de sa configuration temporelle, i.e. à partir de sa partition $(\tau_h, h = 1, \dots, H)$ de l'intervalle $[1, \dots, N]$, en y ajoutant des nombres r_h compris dans $[0, M]$ et qui dictent à quel rang on fait la séparation entre voxels de la classe C_h^1 qui appartiennent donc à l'objet, et voxels de la classe C_h^2 qui appartiennent donc au fond de l'image, i.e.

$$\text{rank}_{\tau_h}(i) \leq r_h \Leftrightarrow i \in C_h^1$$

et

$$\text{rank}_{\tau_h}(i) > r_h \Leftrightarrow i \in C_h^2$$

On écrit alors l'ensemble des subdivisions spatio-temporelles \mathcal{M} de S comme suit :

$$\begin{aligned} \mathcal{M} = \left\{ \{(\tau_h, r_h), h = 1, \dots, H\}; \tau_h \subseteq [1, N], \tau_h \cap \tau_{h'} = \emptyset, \text{if } h \neq h', \cup_{h=1}^H \tau_h = [1, N]; \right. \\ \left. r_h \in [1, \dots, M], \forall h \in \{1, \dots, H\}; H \in \{1, \dots, N\} \right\} \quad (4.4) \end{aligned}$$

et l'estimateur de projection associé à une subdivision spatio-temporelle $m = \{(\tau_h, r_h), h = 1, \dots, H\}$ comme suit

$$(\hat{S}_m)_i(t) = \sum_{h=1}^H \left[\bar{Y}_{h,1} \mathbb{I}(\text{rank}_{\tau_h}(i) \leq r_h) + \bar{Y}_{h,2} \mathbb{I}(\text{rank}_{\tau_h}(i) > r_h) \right] \mathbb{I}_{\tau_h}(t)$$

où $\bar{Y}_{h,1}$ s'écrit

$$\bar{Y}_{h,1} = \frac{1}{\sum_{t \in \tau_h} \sum_{i, \text{rank}_{\tau_h}(i) \leq r_h} \frac{1}{\sigma_{i,t}^2}} \sum_{t \in \tau_h} \sum_{i, \text{rank}_{\tau_h}(i) \leq r_h} \frac{Y_i(t)}{\sigma_{i,t}^2}$$

et $\bar{Y}_{h,2}$ (s'il existe) s'écrit :

$$\bar{Y}_{h,2} = \frac{1}{\sum_{t \in \tau_h} \sum_{i, \text{rank}_{\tau_h}(i) > r_h} \frac{1}{\sigma_{i,t}^2}} \sum_{t \in \tau_h} \sum_{i, \text{rank}_{\tau_h}(i) > r_h} \frac{Y_i(t)}{\sigma_{i,t}^2}$$

spatio-temporelle La question qui se pose maintenant est comment connaître la fonction rank puisque elle dépend de la moyenne du signal, donc de S qui est inconnu. La spécificité de notre problème dû au fait que les objets contrastés présentent une moyenne d'intensité supérieure à celle du fond implique donc l'utilisation de la moyenne empirique du signal comme suit :

$$\begin{aligned} \widehat{\text{rank}}_\tau(i) = k \Leftrightarrow & \quad \exists j_i^1 \neq j_i^2 \neq \dots \neq j_i^M \in \{1, \dots, M\} \quad \text{tq.} \\ & \frac{1}{|\tau|} \sum_{t \in \tau} Y_i(t) \geq \frac{1}{|\tau|} \sum_{t \in \tau} Y_j(t), \forall j \in \{j_i^1, \dots, j_i^{k-1}\} \\ & \text{et } \frac{1}{|\tau|} \sum_{t \in \tau} Y_i(t) \leq \frac{1}{|\tau|} \sum_{t \in \tau} Y_j(t), \forall j \in \{j_i^{k+1}, \dots, j_i^M\} \end{aligned}$$

Nous sommes maintenant en mesure de décrire notre algorithme dynamique de minimisation du critère pénalisé suivant :

$$\text{Crit}(\hat{S}_m) = \sum_{i=1}^M \sum_{t=1}^N \frac{((\hat{S}_m)_i(t) - Y_i(t))^2}{\sigma_{i,t}^2} + \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_m \right) D_m \quad (4.5)$$

Remarquant d'abord que $\text{Crit}(\hat{S}_m)$ est composé de deux termes additifs, le premier terme ne dépendant que des données $Y_i(t)$ et non de la dimension D_m de \hat{S}_m , et le second terme ne dépendant que de la dimension D_m de \hat{S}_m et non des données $Y_i(t)$. On définit alors l'estimateur \hat{S}_D pour tout $D \in [D_{min}, \dots, D_{max}]$ comme suit :

$$\hat{S}_D = \arg \min_{m \in \mathcal{M}, D_m \equiv D} \gamma(\hat{S}_m), \quad \forall D \in [D_{min}, \dots, D_{max}]$$

On en déduit que l'estimateur pénalisé \tilde{S} vérifie

$$\tilde{S} \triangleq \arg \min_{m \in \mathcal{M}, D_m \equiv D} \text{Crit}(\hat{S}_m) = \arg \min_{D=D_{min}, \dots, D_{max}} \left[\gamma(\hat{S}_D) + \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_D \right) D \right]$$

Notre algorithme procède donc d'abord par le calcul pour chaque dimension D ($\forall D \in [D_{min}, \dots, D_{max}]$) de l'estimateur \hat{S}_D , sa fonction "coût" $\gamma(\hat{S}_D)$ ainsi que la quantité $\text{Card}_{\mathcal{M}}(D)$ qui représente le nombre de modèles (ou de façon équivalente, le nombre de configurations $m \in \mathcal{M}$) ayant la dimension D , et ce, en utilisant la programmation dynamique, puis par la minimisation par rapport à D de l'expression

$$\text{Crit}(\hat{S}_D) \triangleq \left[\gamma(\hat{S}_D) + \text{Big} \left((2 - \alpha)(1 + \theta) + \left(\frac{2 - \alpha}{\theta} + \frac{2}{\alpha} \right) L_D \right) D \right], \quad \forall D \in [D_{min}, \dots, D_{max}]$$

Avant de décrire l'algorithme dynamique qui permet le calcul de l'estimateur \hat{S}_D , de sa fonction "coût" $\gamma(\hat{S}_D)$ ainsi que la quantité $\text{Card}_{\mathcal{M}}(D)$, fixant d'abord quelques notations qui vont nous aider à décrire l'algorithme :

- On définit la fonction "coût" $\gamma_{\tau}(i, j)$, pour tout $\tau \subseteq [1, N]$ et pour tout $1 \leq i \leq j \leq M$ comme la somme des différences au carré des données $Y_k(t), t \in \tau; k, i \leq \widehat{\text{rank}}_{\tau}(k) \leq j$ par rapport à la moyenne des données $\bar{Y}_{\tau}(i, j)$ dans l'intervalle spatio-temporel ordonné $\tau \times [i, j]$ comme suit :

$$\bar{Y}_{\tau}(i, j) = \frac{1}{\sum_{\substack{t \in \tau \\ k, i \leq \widehat{\text{rank}}_{\tau}(k) \leq j}} \frac{1}{\sigma_k(t)^2}} \sum_{\substack{t \in \tau \\ k, i \leq \widehat{\text{rank}}_{\tau}(k) \leq j}} \frac{Y_k(t)}{\sigma_k(t)^2}$$

$$\gamma_{\tau}(i, j) = \sum_{\substack{t \in \tau \\ k, i \leq \widehat{\text{rank}}_{\tau}(k) \leq j}} \frac{(Y_k(t) - \bar{Y}_{\tau}(i, j))^2}{\sigma_k(t)^2}$$

ainsi, avec la convention $\gamma_{\tau_h}(i, j) = 0$ si $i > j$, la quantité $\gamma(\hat{S}_m)$ définie plus haut dans cette section, où la subdivision spatio-temporelle m est donnée par $m = \{(\tau_h, r_h), h = 1, \dots, H\}$, s'écrit

$$\gamma(\hat{S}_m) = \sum_{h=1}^H [\gamma_{\tau_h}(1, r_h) + \gamma_{\tau_h}(r_h + 1, M)]$$

Nous revenons plus tard à l'algorithme pour calculer efficacement les quantités $\gamma_{\tau}(i, j), \forall \tau \subseteq [1, N], \forall 1 \leq i \leq j \leq M$.

- Soit $\tilde{\gamma}_t(D)$ le coût minimal sur toutes les subdivisions spatio-temporelles de cardinalité (de dimension) D du rectangle spatio-temporel $[t, N] \times \{1, \dots, M\}$. Avec cette définition, il est clair que $\gamma(\hat{S}_D) = \tilde{\gamma}_1(D), \forall D \in [D_{\min}, \dots, D_{\max}]$.
- Soit $c_t(D)$ le nombre de subdivisions spatio-temporelles de cardinalité (de dimension) D du rectangle spatio-temporel $[t, N] \times \{1, \dots, M\}$. Avec cette définition, il est clair aussi que $\text{Card}_{\mathcal{M}}(D) = c_1(D), \forall D \in [D_{\min}, \dots, D_{\max}]$.
- Soit w et W les tailles minimale et maximale (respectivement) d'un intervalle temporel d'une subdivision spatio-temporelle. Par défaut, $w = 1$ et $W = N$, mais l'utilisateur peut spécifier d'autres valeurs de w supérieures à 1 et des valeurs de W inférieures à 1 par exemple pour ajouter un *a priori* sur la taille temporelle d'une rupture, ou seulement pour accélérer le processus de calcul.

- Soit u la taille spatiale minimale d'une rupture spatio-temporelle. Par défaut, u est évidemment égal à 1, mais l'utilisateur peut spécifier une autre valeur de u supérieure à 1 pour ajouter un *a priori* sur la taille spatiale minimale d'une rupture spatio-temporelle, par exemple pour éviter la détection de ruptures trop petites et qui ne sont pas utiles. taille minimale d'un ensemble d'indices spatiaux d'une subdivision spatio-temporelle.

On a alors les formules récursives suivantes de calcul des quantités $\tilde{\gamma}_t(D)$ et $c_t(D)$:

$$\tilde{\gamma}_t(D) = \min_{\substack{\min(N-w, t+W-1) \geq s \geq t+w, \\ u \leq i \leq M-1}} \left\{ \gamma_{t,s}(1, i) + \gamma_{t,s}(i+1, M) + \tilde{\gamma}_{s+1}(D-2), \right. \\ \left. \gamma_{t,s}(1, M) + \tilde{\gamma}_{s+1}(D-1) \right\}, \quad N-W+1 \leq t \leq N-w+1, 1 \leq D \leq D_{max}$$

$$c_t(D) = \sum_{\min(N-w+1, t+W-1) \geq s \geq t+w} c_s(D-1) + (M-u) \times c_s(D-2), \\ N-W+1 \leq t \leq N-w+1, 1 \leq D \leq D_{max}$$

Notant bien que le calcul des quantités $\tilde{\gamma}_t(D)$ et $c_t(D)$ avec ces formules peut se faire en parallèle et cela résulte en une complexité algorithmique polynomiale en $O(D_{max}NWM)$.

4.2 Détection des vésicules dans une séquence de vidéo-microscopie

4.2.1 Etat de l'art

Il existe de nombreuses méthodes de détection de structures plus ou moins particulières ; pour la détection de vésicules, on peut mentionner quatre types d'approches : une approche générale dite de centroides, une approche basée sur l'ajustement de modèles, une approche basée sur l'analyse des coefficients d'ondelettes et une approche basée sur la sélection de modèle.

Méthode de centroides

Cette approche repose sur une binarisation préalable de l'image. Ceci effectué, les parties connexes sont extraites et dans la version la plus simple de cette famille d'algorithmes, le centroïde de chaque partie connexe est calculé [Cheezum et al. (2001), Carter et al. (2005)].

Une variante de cet algorithme [Crocker et al. (1996), Thompson et al. (2002)] consiste à pondérer (dans un schéma itératif destiné à améliorer la précision de la localisation) les points selon la formule (en 2D) :

$$x_0^k = \frac{\sum_{i,j} x_i I_{i,j} N_{i,j}^{k-1}}{\sum_{i,j} I_{i,j} N_{i,j}^{k-1}}$$

$$y_0^k = \frac{\sum_{i,j} y_j I_{i,j} N_{i,j}^{k-1}}{\sum_{i,j} I_{i,j} N_{i,j}^{k-1}}$$

avec $N_{i,j}^{k-1} = \int_{x_i}^{x_{i+1}} \int_{y_j}^{y_{j+1}} \exp\left(-\frac{(x-x_0^{k-1})^2 + (y-y_0^{k-1})^2}{2R^2}\right) dx dy$ et $x_0^0 = \frac{\sum_{i,j} x_i I_{i,j}}{\sum_{i,j} I_{i,j}}$, $y_0^0 = \frac{\sum_{i,j} y_j I_{i,j}}{\sum_{i,j} I_{i,j}}$

Ajustement de modèles d'intensité

On peut distinguer deux situations : celle des objets d'une taille inférieure à la limite de diffraction de la lumière qu'on peut cependant observer en microscopie de fluorescence, celle d'objets au delà de cette limite de diffraction. Dans le premier cas, les algorithmes de détection utilisent la fonction d'étalement de l'instrument ou PSF⁷. Cette fonction est souvent approximée par une Gaussienne [Thompson (2002), Bonneau et al. (2005), Zhang et al. (2007)] ou estimée conjointement à la détection [Aguet et al. (2005)]. La détection d'une particule unique peut alors être réalisée avec une précision de l'ordre de quelques nm et des particules même très proches peuvent être séparées alors qu'elles sont distantes de quelques dizaines de nm seulement [Thomann et al. (2002)]; le facteur limitant est alors surtout le bruit. Dans le cas d'objets dont la forme et le profil d'intensité sont connus, certains algorithmes de détection sont basés sur une modélisation du profil comme une somme de fonctions de différentes nature (gaussiennes par exemple, ou encore lorentziennes), selon les applications [Rink et al. (2005), Zhang et al. (2006)].

Parfois, il n'y pas de fonction analytique évidente pour représenter le profil d'intensité des objets à détecter et on préfère avoir recours à un patron (ou template) d'intensité défini directement à partir des données. Ce patron correspond à une distribution d'intensités sur un domaine généralement rectangulaire et un coefficient de corrélation est localement calculé avec ce patron. Une précision un peu meilleure de la position de l'objet peut être

⁷ pour Point Spread Function.

obtenue par modélisation locale autour de l'optimum de la fonction de corrélation. Un des principaux défauts de cette approche réside dans le fait que des pixels du fond sont aussi généralement présents dans le patron et que ces pixels sont traités de manière identique aux pixels de l'objet à détecter. Par conséquent plus le patron contient de pixels du fond et plus on a de chances de faire des fausses détections. Cho et al. (2005) ont proposé une méthode baptisée d'attention sélective pour remédier à ce problème et réduire le nombre de fausses détections ou fausses alarmes.

Analyse multi-échelle L'analyse multi-échelle et en particulier les décompositions en ondelettes ont été très popularisées notamment après la découverte de l'algorithme de décomposition rapide (FWT⁸) par Stéphane Mallat (1989). Les décompositions en ondelettes qui semblent le plus adaptées à la segmentation d'images utilisent une décomposition dite *à trous* [Holschneider et al. (1989)]. Contrairement aux décompositions par le FWT, il n'y a pas d'étapes de décimation suite à la projection du signal à chaque échelle dyadique. De ce fait, la décomposition est invariante par translation au prix d'une redondance des coefficients. Ainsi deux images contenant chacune un même objet mais avec un léger décalage spatial donneront lieu à deux jeux de coefficients identiques mais légèrement décalés en index. Les méthodes de segmentation d'objets basées sur l'algorithme *à trous* consistent alors [Racine et al. (2006)] : a) à choisir une fonction d'échelle ou une ondelette (e.g B-spline [Starck et al. (1998)], b) à choisir des seuils à appliquer aux coefficients d'ondelettes dans les différents plans, c) à combiner les images binaires obtenues dans les différents plans pour faire ressortir les objets d'intérêt. Les seuils sur les coefficients d'ondelettes sont estimés à partir du premier plan d'ondelette et d'une observation de Starck (2002) dans le cas d'un bruit gaussien, selon laquelle le rapport $\frac{\sigma_c^j}{\sigma_\epsilon}$ de l'écart type du bruit dans le plan j d'ondelettes et de l'écart type du bruit est constant et dépend seulement de l'ondelette qui a servi à la décomposition.

Une approche un peu différente avait aussi été mise en œuvre par J.-C. Olivo Marin pour détecter des protéines marquées [Olivo Marin et al. (2002)]. Elle reposait aussi sur l'algorithme à trous et un seuillage des coefficients dans chaque plan. Cependant la

⁸Fast Wavelet Transform

décision d'étiquetage de l'objet reposait en outre sur une image de corrélation inter-échelles des coefficients.

Sélection de modèles L'approche proposée par [Boulanger et al. (2006)] et inspirée des travaux de [Huet (2002)] suppose que les images ont été au préalable débruitées et débarassées du fond. Le signal corrigé est alors modélisé par

$$r_{\mathbf{x}} = m_{\mathbf{x}} + e_{\mathbf{x}}, \quad e \sim \mathcal{N}(0, \Sigma)$$

où r désigne le vecteur des intensités corrigées au pixel $\mathbf{x} = (x, y, z)$ pour tous les instants observés et $e_{\mathbf{x}}$ un bruit résiduel gaussien i.i.d. centré de variance $\Sigma = \tau_{\mathbf{x}}^2 I_n$; n désigne le nombre d'instant. Le vecteur $r_{\mathbf{x}}$ est alors ordonné et la détection est formulée comme la sélection parmi tous les modèles gaussiens précédents d'un modèle maximisant une log-vraisemblance pénalisée par la dimension du modèle ou son nombre de composantes non nulles. A un modèle correspond donc un seuil d'intensité en deça duquel on considère que le pixel n'est pas allumé aux instants pour lesquels l'intensité corrigée est inférieure à ce seuil. Différentes pénalités sont comparées par Boulanger (2006) qui opte pour une pénalité de la forme :

$$pen_{MKR}(k) = n(c_1 \log \frac{n}{k} + c_2) \frac{k}{n - k}$$

Il s'agit donc d'une méthode de seuillage adaptatif spatialement, avec un calcul de seuil automatique qui est le même pour tous les instants et un même pixel. L'hypothèse faite derrière cette modélisation consiste à supposer qu'un pixel donné quelconque de la séquence corrigée peut ne pas être 'allumé'. En pratique certains pixels de l'image ne sont jamais 'allumés' (e.g. pixels en dehors de la cellule) et la même méthode de sélection de modèle, basée cette fois sur le vecteur des seuils, permet de sélectionner automatiquement un seuil minimum. La carte des seuils révèle aussi le support de l'activité dans l'image mais aussi la persistance de l'activité en un pixel si on colorie cette carte proportionnellement au nombre d'instant au dessus du seuil en chaque voxel. Cette méthode de segmentation est relativement rapide, elle dépend beaucoup de la méthode de correction et n'exploite pas la cohérence spatio-temporelle locale. Boulanger et al ont validé cette méthode sur des données synthétiques et réelles.

4.2.2 Détection en se basant sur notre segmentation spatio-temporelle

On distingue les trois étapes suivantes :

Pré-traitement

Cette étape a été discutée en détails dans le chapitre 2 et son but était alors de produire un volume spatio-temporel pour lequel :

- On a corrigé la perte de fluorescence temporelle due entre autres au phénomène de photo-blanchiment.
- On a stabilisé la variance spatio-temporellement dans la séquence dans le but de ‘Gaussianiser’ le signal.
- On a supprimé la partie déterministe quasi-stable pour avoir un fond de moyenne nulle et n’en garder que l’information dynamique pertinente. Cela permettra de s’affranchir du délicat de choix du seuil au dessus duquel on considère qu’un voxel appartient à un vésicule en mouvement.

Segmentation spatio-temporelle

Après l’étape de pré-traitement, on applique la méthode de segmentation spatio-temporelle décrite dans la section précédente. Les parties spatio-temporelles résultant d’une telle segmentation spatio-temporelle et dont la moyenne d’intensité dépasse un certain seuil pré-spécifié⁹ noté s_0 . Ainsi, on aura binarisé tout le volume spatio-temporel¹⁰.

Groupement perceptuel

Il s’ensuit alors qu’un groupement perceptuel (en utilisant l’algorithme des composantes connexes) des voxels étiquetés “en mouvement” à chaque instant de la séquence permet de détecter les vésicules en mouvement dans la séquence en tant que composantes connexes. Et pour éliminer les éventuelles fausses détections, on élimine toutes les composantes connexes de taille inférieure par exemple à 5 voxels.

⁹Théoriquement, $s_0 \equiv 0$ mais en pratique, il est important que le seuil s_0 soit légèrement supérieur à zéro.

¹⁰i.e. étiqueté tous ses voxels par “en mouvement” ou “statique”

4.3 Résultats expérimentaux

4.3.1 Résultats sur des signaux spatio-temporels synthétiques

Afin d'évaluer les performances de notre approche, nous l'avons d'abord expérimentée sur des données synthétiques. Nous avons alors créé un premier programme pour simuler des signaux bidimensionnels (une dimension spatiale et une dimension temporelle) dans lesquels nous créons un nombre aléatoire d'interruptions bidimensionnelles aléatoires dans le signal (*cf.* figure 4.3) de manière similaire à celle qui se produirait pour un paquet de signaux voisins d'une séquence de vidéo-microscopie lors d'un passage d'un spot fluorescent. Pour une taille spatiale (M) et une taille temporelle (N) fixées, notre algorithme de simulation procède comme suit pour simuler un nombre n de signaux bidimensionnels :

- Nous simulons d'abord un fond nul et nous lui superposons un bruit blanc additif Gaussien de variance égale à 1.
- On se fixe une taille temporelle maximale (qu'on note H) et une taille spatiale minimale (qu'on note w) d'une rupture spatio-temporelle. On pose $\tau_2(0) = 1$ comme le temps correspondant à la zéro-ième rupture dans le signal bidimensionnel, et on simule les autres instants de début de rupture $\tau_1(h)$ et de fin de rupture d'un signal $i = 1, \dots, n$ comme suit :

$$\tau_1(h) = \text{Uniforme}[\tau_2(h-1), N], \quad h = 1, 2, \dots$$

et quant à l'instant final du h -ième intervalle de rupture temporelle $\tau_2(h)$, il est simulée comme suit :

$$\tau_2(h) = \text{Uniforme}[\tau_1(h), \min(\tau_1(h) + H - 1, N)] \quad i = 1, 2, \dots$$

et pour chaque intervalle de rupture $[\tau_1 h, \tau_2(h)]$, on simule l'indice r_h de rupture dans la dimension spatiale comme suit :

$$r_h = \text{Uniforme}[w, M - w], \quad h = 1, 2, \dots$$

alors, les r_h -ièmes premiers signaux ¹¹ verront leur valeur $Y_{h,1}$ modifiée dans l'intervalle $[\tau_1 h, \tau_2(h)]$ comme suit. D'abord, on tire le niveau (valeur moyenne)

¹¹Notons que cette manière de simuler les ruptures spatiales quoique pratique est sans perte de généralité puisque notre algorithme dynamique ignore cet ordre et utilise son "propre" ordre grâce à la fonction "rank"

du signal $S_{h,1}$ comme suit :

$$S_{h,1} = \text{Uniforme}[1.5, 3]$$

et pour chacun des r_h premiers signaux, on modifie leur valeur selon la formule

$$Y_i(t) = S_{h,1} + \epsilon_i(t), \quad \forall t \in [\tau_1(h), \tau_2(h)], \quad \text{et} \quad \forall i = 1, \dots, r_h$$

où $\epsilon_i(t)$, $\forall t \in [\tau_1(h), \tau_2(h)]$, et $\forall i = 1, \dots, r_h$ sont des Gaussiennes standards.

On arrête ce processus de simulation d'interruptions dans le signal i dès que la condition $N - \tau_2(h) \leq 3$ est vérifiée.

Une fois les n signaux spatio-temporels simulés (*cf.* figure 4.3 pour quelques exemples de tels signaux), on leur fait alors tourner dessus notre algorithme dynamique de recherche de ruptures spatio-temporelles. Pour apprécier les performances de notre algorithme et de notre critère de sélection de modèle appliqué au problème de recherche de ruptures spatio-temporelles multiples dans un signal spatio-temporel, nous avons alors combiné le même algorithme dynamique avec les différents critères pénalisés suivants : Mallow's C_p , BIC et le second critère pénalisé que nous avons proposé dans le chapitre précédent. Nous définissons alors les deux quantités suivantes pour comparer les performances de ces critères :

- Le risque moyen défini par

$$\overline{\mathcal{R}} = \frac{1}{n} \sum_{i=1}^n \frac{\|S^i - \tilde{S}^i\|^2}{NM}$$

où S^i et \tilde{S}^i désignent respectivement la partie déterministe et l'estimateur pénalisé du i -ème signal spatio-temporel simulé,

- Le sur-écart moyen entre la dimension effective du i -ième signal spatio-temporel simulé et la dimension de l'estimateur pénalisé correspondant

$$\overline{\mathcal{D}^+} = \frac{1}{n^+} \sum_{i=1}^n (\tilde{D}^i - D^i) \mathbb{1}_{(0, \infty)}(\tilde{D}^i - D^i)$$

où $n^+ = \sum_{i=1}^n \mathbb{1}_{(0, \infty)}(\tilde{D}^i - D^i)$,

- Le sous-écart moyen entre la dimension effective du signal simulé et la dimension de l'estimateur pénalisé correspondant

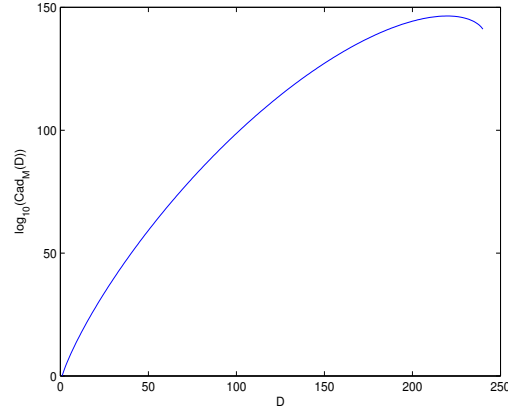


FIG. 4.2: Le Logarithme en base 10 de $\text{Card}_{\mathcal{M}}(D)$ pour différentes valeurs de D et pour $M = 18$ et $N = 120$. Ce graphique présente un pic à $D = 220$ avec $\text{Card}_{\mathcal{M}}(220) \approx 3.02141 \times 10^{146}$.

$$\overline{\mathcal{D}}^- = \frac{1}{n^-} \sum_{i=1}^n (\tilde{D}^i - D^i) \mathbb{I}_{(-\infty, 0)}(D^i - \tilde{D}^i)$$

où $n^- = \sum_i^n \mathbb{I}_{(-\infty, 0)}(\tilde{D}^i - D^i)$, D^i et \tilde{D}^i désignent respectivement la dimension réelle et la dimension de l'estimateur pénalisé du i -ième signal spatio-temporel simulé.

Les valeurs de $\overline{\mathcal{R}}$, $\overline{\mathcal{D}}^+$ et $\overline{\mathcal{D}}^-$ obtenus sur une liste de $n = 1000$ signaux spatio-temporels simulés sont représentées dans le tableau suivant :

| Nom du critère | $\overline{\mathcal{R}}$ | $\overline{\mathcal{D}}^+$ | $\overline{\mathcal{D}}^-$ |
|-------------------|--------------------------|----------------------------|----------------------------|
| Mallows' C_p CP | 0.7507 | 180.71 | 0 |
| BIC | 0.3461 | 49.36 | 0 |
| SCP* | 0.0939 | 0.1 | 0.0909 |
| SCPS* | 0.0996 | 0 | 0.09 |

où

- SCP* est le second critère de sélection de modèle que nous avons présenté dans le chapitre précédent pour ajuster un modèle à un seul jeu de données.
- SCPS* est la version sélection simultanée de modèles de SCP* que nous avons proposé dans le chapitre précédent.

Ce tableau montre clairement que mis à part notre critère, tous les autres ont tendance à sur-ajuster les données, ce qui prouve encore une fois la parcimonie du critère de sélection de modèle que nous avons proposé ainsi que l'efficacité de l'algorithme dynamique que nous

avons développé pour le problème de détection de ruptures spatio-temporelles multiples dans un signal spatio-temporel. On remarque également que la prise en compte de la correction dans le critère pénalisé due au fait qu'on cherche à sélectionner un modèle pour plusieurs jeux de données simultanément, permet d'en améliorer les performances en évitant le phénomène de sur-ajustement discuté dans le chapitre précédent.

4.3.2 Résultats sur des séquences d'images

On a réalisé des expériences sur des séquences d'images $3D + T$ synthétiques et réelles issues de la vidéo-microscopie de fluorescence. On simule dans les séquences synthétiques le mouvement de sondes fluorescentes le long d'un réseau créé de façon aléatoire dans la séquence. On simule un tel réseau en tirant aléatoirement n nœuds dans le volume $3D$ et connectant ces derniers de façon aléatoire en utilisant q arêtes. On a réalisé l'expérience avec deux types de profil d'intensité des sondes fluorescentes ; le profil constant et le profil gaussien. Quelques résultats de détection de sondes en mouvement dans de telles séquences simulées sont représentés en figures 4.4 et 4.5.

Le but bien évidemment c'est de pouvoir appliquer la méthode à la détection de vésicules dans des séquences de vidéo-microscopie réelles. Quelques résultats sur des séquences de vidéo-microscopie sont alors représentés en figure 4.6. Faute de la réalité terrain en vidéo-microscopie de fluorescence $3D+T$, la validation de nos résultats s'est faite de façon visuelle en superposant par exemple les détections aux images $3D$ originales.

4.4 Conclusion

Nous avons présenté une approche de segmentation spatio-temporelle d'une séquence de vidéo-microscopie de fluorescence $3D + T$ en utilisant la théorie de sélection de modèle présentée dans le chapitre 3 et nous avons développé un algorithme dynamique qui permet d'implémenter efficacement une telle approche. L'approche procède d'un point de vue "bottom-up" en ce sens que pour réaliser une segmentation temporelle globale d'une séquence de vidéo-microscopie, elle réalise d'abord des segmentations spatio-temporelles locales dans plusieurs blocs du volume $3D$ de la séquence, avant de les agréger. L'utilisation

du concept de sélection simultanée de modèles présenté dans le chapitre 3 permet alors de contrôler de façon automatique le nombre de fausses segmentations spatio-temporelles.

La méthode a l'avantage d'être complètement automatique et elle est facile à implémenter sur machine, toutefois, sa limite réside surtout dans le choix d'un modèle spatio-temporel constant par morceaux qu'on essaie d'ajuster au profil d'intensité spatio-temporel correspondant à chaque bloc du volume $3D$ de la séquence. En effet, même si un tel modèle peut constituer un bon compromis coût de calcul/faculté de modélisation, nous pensons qu'il présente beaucoup plus un modèle "pédagogique" qui nous a servi à démontrer la faisabilité de notre approche de segmentation spatio-temporelle d'une séquence de vidéo-microscopie, qu'un modèle standard qu'on pourrait utiliser pour la détection à grande échelle de vésicules en vidéo-microscopie. Heureusement, notre approche n'exclut pas l'utilisation d'autres modèles spatio-temporels plus réalistes. De tels modèles doivent reposer sur des considérations théoriques intéressantes certes, mais dont on peut prouver en pratique par le biais d'une étude expérimentale à grande échelle sur des données image issues de la vidéo-microscopie de fluorescence la validité. Il faudrait signaler aussi que la tendance actuellement en vidéo-microscopie c'est l'utilisation de modèles de PSF spatiaux super-résolution pour réaliser une détection sous-pixelle de sondes fluorescentes en vidéo-microscopie de fluorescence, et par là, accroître la précision de l'analyse des données de vidéo-microscopie. Nous pensons alors qu'une bonne voie de recherche serait axée sur l'utilisation de modèles spatio-temporels super-résolution réaliser une détection sous-pixelle des vésicules, tout en défiant le faible rapport signal à bruit des données images issues de la vidéo-microscopie de fluorescence puisqu'on pourra exploiter, avec de tels modèles spatio-temporels, à la fois la cohérence spatiale et la cohérence temporelle du signal.

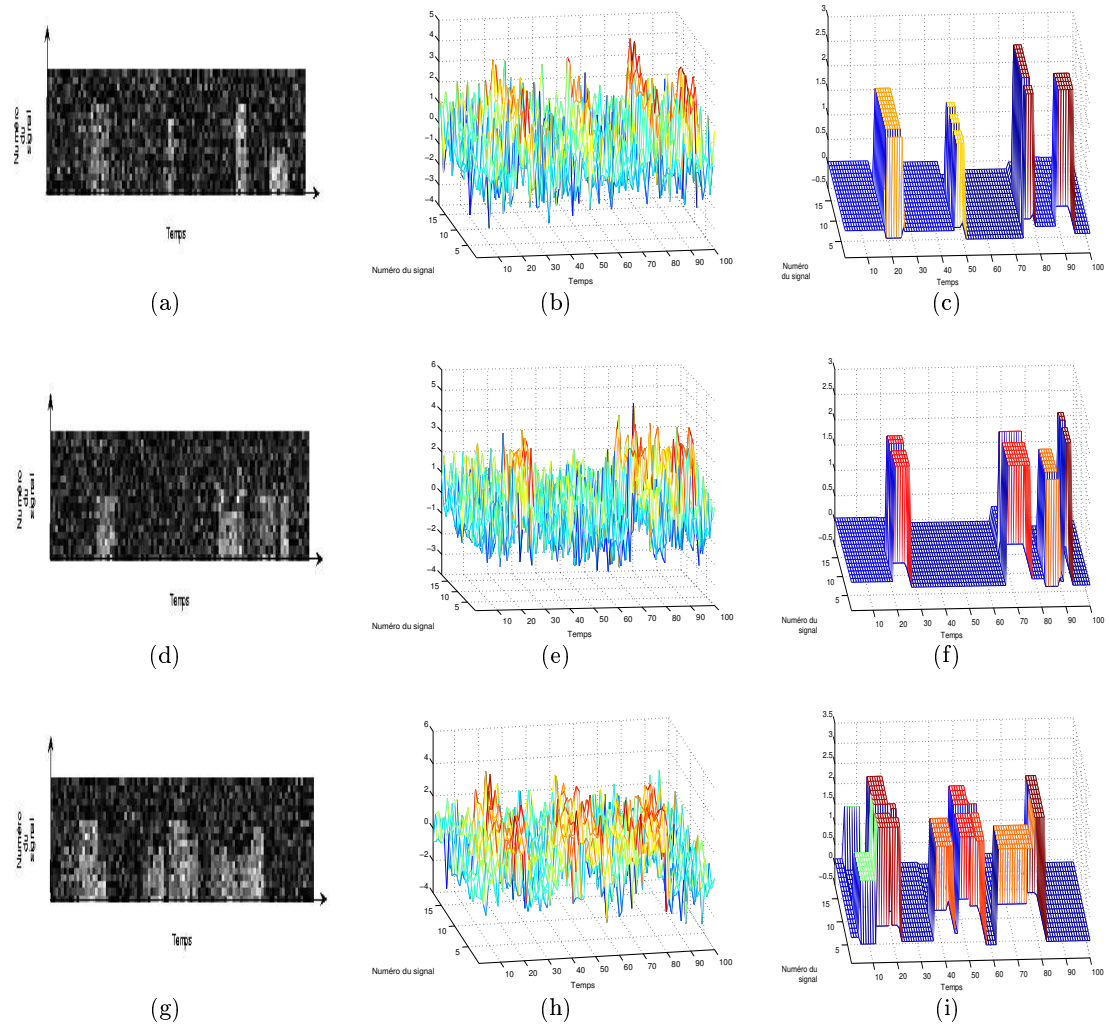


FIG. 4.3: Quelques exemples de résultats de détection de ruptures spatio-temporelles dans un paquet de signaux spatio-temporels (a)-(d)-(g) Représentation "image" d'un signal spatio-temporel; (b)-(e)-(h) Représentation topographique d'un signal spatio-temporel; (c)-(f)-(i) Résultat de la détection de ruptures spatio-temporelles dans le signal.

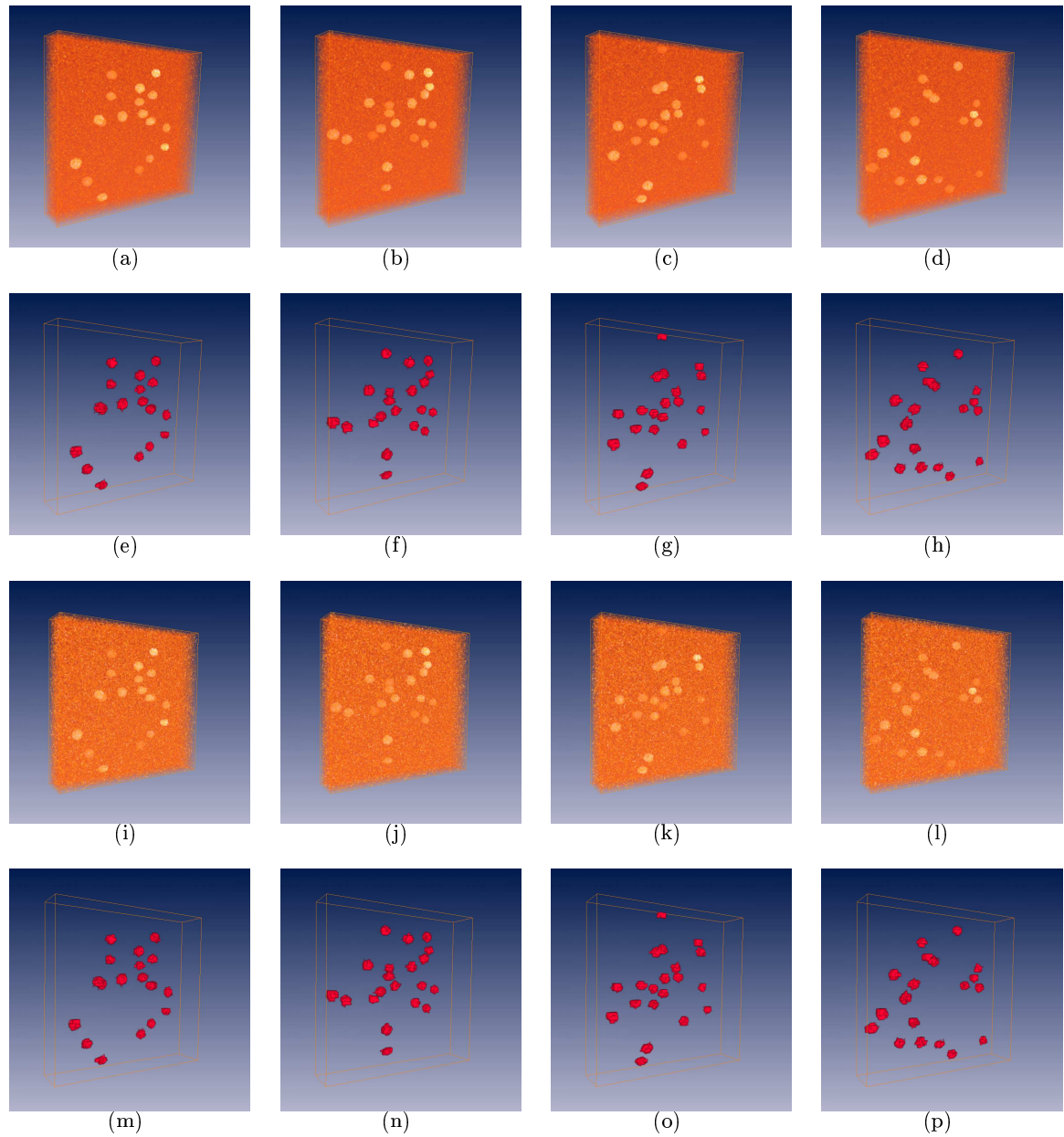


FIG. 4.4: Détection spatio-temporelle de spots fluorescents mobiles ayant un profil d'intensité constant dans une séquence simulée en utilisant l'approche développée dans ce chapitre : (a)-(b)-(c)-(d) avec $\sigma = 1$; (i)-(j)-(k)-(l) avec $\sigma = 1.5$.

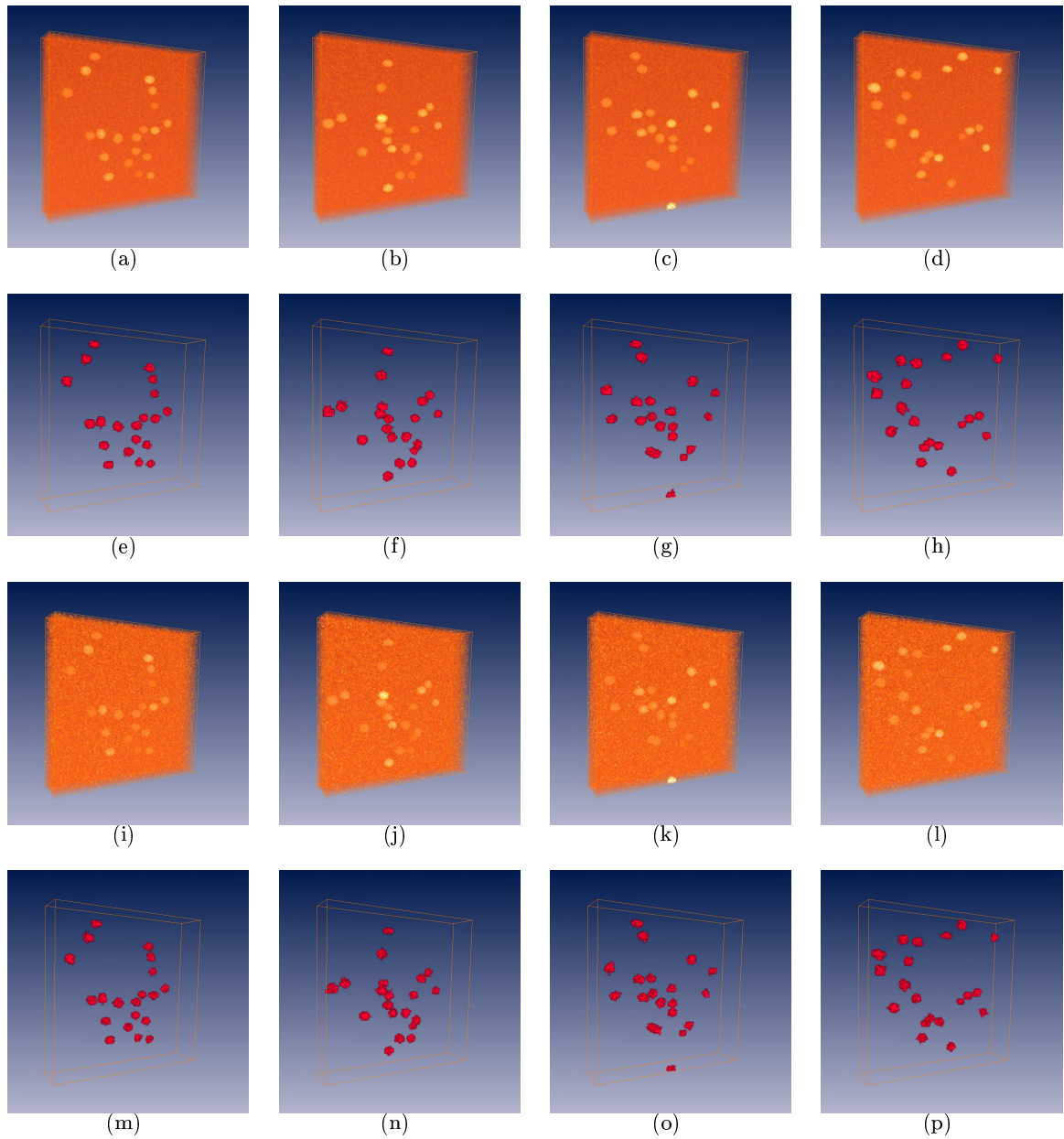


FIG. 4.5: Détection spatio-temporelle de spots fluorescents mobiles ayant un profil d'intensité gaussien dans une séquence simulée en utilisant l'approche développée dans ce chapitre : (a)-(b)-(c)-(d) avec $\sigma = 1$; (i)-(j)-(k)-(l) avec $\sigma = 1.5$.

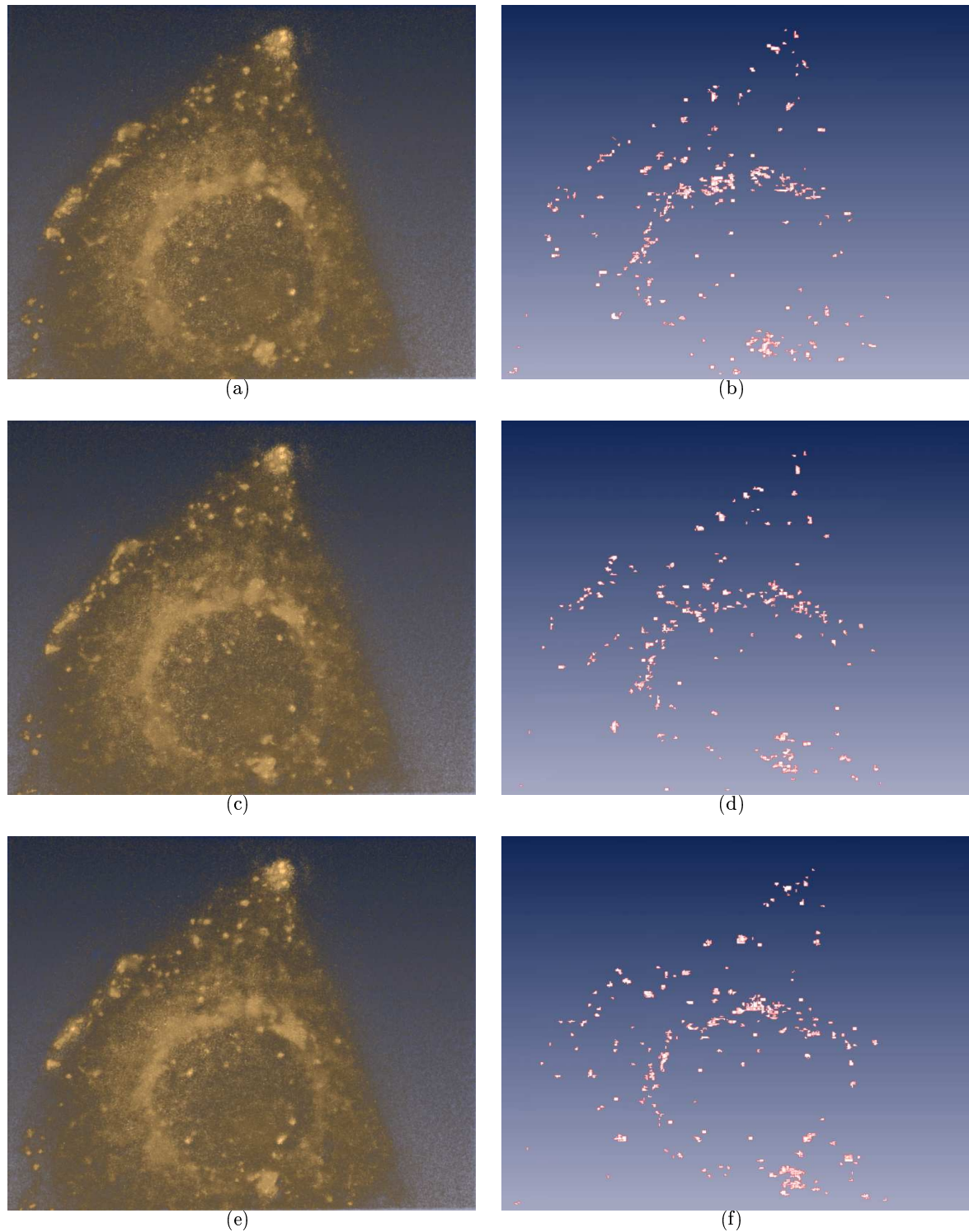


FIG. 4.6: Détection spatio-temporelle de vésicules dans une séquence de vidéo-microscopie de taille $395 \times 345 \times 10 \times \text{voxels} \times 120\text{sec}$ ($\approx 320 \text{ M0}$). On note les amas vésiculaires qui sont très nombreux dans la séquence surtout autour des régions quasi-stables de la séquence (Golgi, ER, etc. ...).

Chapitre 5

Estimation de la dynamique de vésicules en vidéo-microscopie

Une analyse précise des dynamiques vésiculaires observées lors du trafic membranaire est compliquée par le grand nombre de vésicules en mouvement (plusieurs centaines), le faible rapport signal à bruit, la complexité des événements qui se produisent au cours du mouvement avec en particulier des phénomènes de fusion-fission des amas vésiculaires détectés. Cette analyse est aussi compliquée par le mouvement d'ensemble de la cellule et la dynamique du réseau de microtubules qui supporte une part importante du trafic. En pratique, il faudrait donc aussi estimer le mouvement global de la cellule [Zimmer et al. (2006)] ainsi que celui des microtubules. Toutefois des techniques d'adhésion des cellules (micro-patterning) [They et al. (2005)] permettent de contrôler en partie ces mouvements et nous ne nous intéresserons pas ici à ces mouvements.

On peut distinguer au moins trois types de démarches pour l'analyse de dynamiques spatiales. La démarche usuelle consiste à analyser les dynamiques individuelles de tous les objets présents. Ces dynamiques individuelles sont alors le résultat de procédures plus ou moins complexes de 'tracking' multi-cibles [Genovesio et al. (2004), Racine et al. (2006)]. On peut cependant s'affranchir de cette tâche difficile et souvent semi-automatique au prix d'une analyse plus grossière des dynamiques. Une idée particulièrement intéressante développée par Boulanger [Boulanger et al. (2006)] consiste à rechercher une représentation de la dynamique

sous la forme d'un graphe dans lequel les nœuds seraient des localisations particulières dans la cellule et une arête entre deux nœuds matérialiserait un échange ou un flux de vésicules entre ces nœuds. Toute la difficulté de l'analyse repose alors sur l'estimation de ce graphe dont les arêtes seraient évaluées par ces flux. Cette approche est inspirée de la tomographie de réseau utilisée pour analyser des réseaux de télécommunication [Vardi et al. (1996)]. Enfin, une démarche encore plus globale consiste à estimer un champ de vitesses dans tout l'espace et analyser ce champ *a posteriori*. Ce champ de vitesses est alors estimé par l'intermédiaire du flot optique [Miura (2004)]. L'approche que nous développons ici se situe entre ces deux dernières approches.

On s'intéresse donc à l'estimation de caractéristiques globales de la dynamique : distribution du module de la vitesse, distribution des directions du mouvement. Bien sûr, ces caractéristiques peuvent être indexées par le temps avec un pas d'échantillonnage temporel plus ou moins long. Dans le cadre le plus précis, on estimera ces caractéristiques globales à chaque instant alors que dans le cadre le plus grossier ces caractéristiques seront calculées en agrégeant toute la séquence temporelle. Ces caractéristiques peuvent aussi être régionalisées bien entendu. L'information restituée diffèrera de celle qui pourrait être déduite d'un calcul de flot optique pour plusieurs raisons : a) le grain est ici plus grossier, les techniques de flot optiques restituent un champ dense de vecteurs, b) l'estimation est faite à partir d'une segmentation spatio-temporelle dans laquelle on a pris soin d'intégrer de la cohérence et spatiale et temporelle, c) du fait de l'apprentissage, on apprendra le lien entre des distributions locales de luminance et la vitesse.

La méthode que nous présentons dans ce chapitre procède différemment du tracking pour l'analyse globale de la dynamique des vésicules impliqués dans le trafic cellulaire. Nous formulons le problème dans un cadre statistique, et nous dérivons alors une relation entre la densité de probabilité du déplacement des vésicules et une certaine statistique calculée sur la séquence. Les paramètres de cette relation sont appris à partir de séquences synthétiques une fois pour toutes.

Dans ce chapitre, nous traitons le cas général d'apprentissage (estimation) de la distribution de probabilité d'une variable aléatoire X (qui comporte une partie observable et une partie cachée) à partir d'une variable aléatoire observable Z liée à X par une fonction

stochastique inconnue.

Nous présenterons , dans une première section introductive les principales approches mentionnées ci-dessus pour l'analyse de dynamique pour fixer dans une seconde partie le cadre général de notre approche avant de détailler dans la section suivante les méthodes d'estimation que nous mettons en œuvre et présenter dans une quatrième section l'application à l'estimation des distributions de vitesse. Nous construirons pour cela une v.a. à partir du processus ponctuel des centres des amas vésiculaires (plus généralement des objets). Enfin dans une dernière section nous présenterons quelques expériences numériques de validation de notre approche.

5.1 Bref état de l'art sur l'estimation de dynamiques

Comme on l'a dit en introduction de ce chapitre, les travaux relatifs à l'estimation de dynamiques à partir de modalités d'imagerie sont extrêmement nombreux, et ce bref tour que nous proposons à pour seule ambition de situer notre travail. Il aurait sans doute été possible de restreindre ce bref tour d'horizon aux modalités de microscopie mais cela n'a pas été notre parti car nous pensons qu'un travail de mise à l'épreuve des techniques les plus récentes d'analyse de dynamiques développés dans d'autres contextes tels la surveillance de différents types de trafic (aérien, routier) restent à faire et qu'il valait donc mieux être plus général au niveau des modalités pour surtout présenter quelques concepts importants. Par ailleurs, le lecteur intéressé peut aussi consulter [Meijering et al. (2006)] pour un point de vue sur l'état de l'art et les défis à relever pour le suivi d'objets en microscopie.

Dans cette section nous abordons brièvement les techniques de flot optique et de tracking. Nous avons mis un focus un peu plus important sur les méthodes de type 'TBF' encore sous-utilisées¹ en imagerie de fluorescence.

5.1.1 Flot optique

Ces techniques visent à l'estimation, à chaque instant, d'un champ de taux de déplacements ou flot optique plus ou moins dense dans l'image. Depuis les travaux pionniers

¹peut-être parce qu'à adapter ?

de Horn et Schunk en 1981, beaucoup de travaux ont eu pour objectif de rendre plus robustes et plus précises ces estimations [Liu et al. (2002), Papenberg et al. (2006)]. D'une manière assez générale, si $x = (x_1, \dots, x_{n-1}, x_n)$ désigne la position d'un pixel, avec éventuellement $x_n \equiv t$, $u(x)$ le flot optique et $I(x)$ l'image au même point. L'estimation de u s'effectue en minimisant une quantité de la forme [Weickert et al. (2006)] :

$$E(u) = \int_{\Omega} \left[M(D^k I, u) + \alpha S(\nabla I, \nabla u) \right] dx$$

où Ω est selon les méthodes le domaine spatial ou un domaine spatio-temporel, $M(D^k I, u)$, un terme d'attache aux données qui s'appuie sur les dérivées de I , et $S(\nabla I, \nabla u)$, un terme de régularisation du flot optique. On peut trouver dans [Weickert et al. (2006)] plusieurs exemples de terme d'attache aux données et de régularisation. On peut mentionner également des estimations du flot optique dans le cas de grands déplacements [Alvarez et al. (2000)]. On trouve quelques utilisations de ces techniques pour la microscopie cde fluorescence mais visiblement assez peu [Miura (2004)].

Une analyse statistique du flot optique dans certaines catégories d'images est aussi développée dans [Roth (2007)].

5.1.2 Suivi individuel des objets en mouvement ou 'tracking'

Le suivi ou 'tracking' d'objets en vision par ordinateur a suscité depuis plusieurs décennies de nombreux travaux. C'est sans doute dans le domaine de la surveillance militaire, du trafic routier, ou plus généralement des lieux publics que les travaux sont les plus nombreux. Les principales difficultés à surmonter avaient trait aux contraintes de temps réel, aux occlusions des objets à suivre, aux instabilités lumineuses des objets à suivre et du fond. Dans le domaine médical, de nombreux travaux ont été consacrés au suivi temporel du coeur. C'est une grande diversité de contextes et de modalités d'imagerie qui ont concouru à une impressionnante productivité dans le domaine du tracking. Beaucoup des techniques mises en œuvre dans ces contextes ont aussi été appliquées pour le suivi de structures en biologie tissulaire [Zimmer et al. (2006), Li et al. (2006)] et cellulaire [Maillet et al. (2002), Genovesio et al. (2006), Arhel et al. (2006), Racine et al (2007a, 2006b), Ji et al. (2005), Meijering

et al. (2006)], avec une demande de plus en plus importante. Mais là encore les modalités d'acquisition d'images à cette échelle sont diverses et complémentaires. La différence la plus importante par rapport à la plupart des autres contextes me semble être le caractère partiel et largement mal connu des observations effectuées. Lorsqu'on observe une scène de trafic routier, la plupart des objets constituant la scène sont connus de l'opérateur aussi bien d'un point de vue physique que d'un point de vue fonctionnel. L'échelle choisie pour l'observation est bonne et peu d'évènements échappent à la perception. En biologie cellulaire, il n'en est rien de tout cela. On est contraint de se limiter à l'observation de quelques structures parmi toutes les structures présentes. Une cellule de mammifère contient environ 10 milliards de protéines de 10000 sortes et même si des progrès considérables ont été effectués, il demeure encore difficile d'observer simultanément plus d'une dizaine de sortes différentes pour diverses raisons ; en pratique les observations effectuées concernent la plupart du temps un ou deux marqueurs d'une ou deux protéines. La seconde difficulté concerne l'échelle. L'échelle micrométrique de la microscopie de fluorescence est insuffisante pour analyser de manière individualisée des vésicules et les fissions ou fusions de structures sont difficiles à analyser. Enfin la troisième difficulté vient du caractère tridimensionnel du milieu étudié. En résumé, le caractère incomplet, multidimensionnel (3D+T+C) avec 3 dimensions spatiales, une dimension temporelle et éventuellement plusieurs dimensions spectrales rendent l'analyse et l'interprétation de dynamiques en biologie cellulaire particulièrement complexe et sujette à des aller-retour entre biologistes, microscopistes, traiteurs d'images et modélisateurs. Avant de défendre un point de vue relativement nouveau dans ce domaine selon lequel, le tracking individuel des structures même s'il est souhaitable n'est pas totalement incontournable à l'analyse de dynamiques nous allons passer en revue les principales méthodes de tracking employées pour le suivi de structures particulières. La plupart des algorithmes employés² reposent sur une étape de détection des structures suivie d'une étape d'association temporelle. Ces deux étapes sont généralement précédées de différents prétraitements (*c.f.* chapitre 2). Nous avons déjà évoqué quelques méthodes de détection d'objets 'particulaires' au chapitre 4 et nous focaliserons donc ici sur quelques grandes familles de méthodes d'association temporelle en nous autorisant cependant une brève incursion autour de méthodes TBD (Track

²Certains algorithmes mêlent cependant détection et tracking [Tonissen (1996)].

Before Detect) qui mêlent détection et tracking.

Méthodes de type 'TAF'

Nous prenons la liberté de regrouper sous cette appellation, les méthodes d'association mises en œuvre après détection des objets d'intérêt.

Soit $E_i(t)$, $i = 1, \dots, n(t)$, $t = 1, \dots, T$ les objets détectés aux différents instants de la séquence temporelle. La tâche d'association consiste à partitionner $\cup_{i=1}^n \text{cup}_{i=1}^n(t)$ en K sous-ensembles disjoints T_k , $k = 1, \dots, K$ constitués des positions successives des différents objets. Chaque sous-ensemble T_k renseigne alors sur la trajectoire d'un seul objet en mouvement.

Depuis plusieurs décennies, de nombreux algorithmes de tracking ont vu le jour et il serait illusoire de vouloir restituer, même de manière condensée, tous ces algorithmes. On peut cependant distinguer quelques grands principes et catégories d'algorithmes.

On suppose que chaque objet, détecté à un instant donné, peut être décrit à l'aide d'un certain nombre de variables. Ces variables peuvent être des variables de position bien sûr mais aussi éventuellement de forme, luminance, dynamique (vitesse, accélération), trajectoire passée.

Les autres hypothèses à considérer concernent : l'incertitude sur ces variables (sans remise en cause de la détection même), la possibilité de fausses détections ou de non-détections, la densité spatiale des objets, la complexité numérique tolérable, la disponibilité de toute la séquence (tracking hors ligne ou en ligne).

L'appariement des objets au temps courant est généralement obtenu en minimisant un critère faisant intervenir les variables observées au temps courant et le plus souvent au seul temps précédent. Plusieurs approches de ce type ont été proposées ces dernières 20 années (au moins) [Sethi (1987), Chetverikov (1998), Sbalzarini (2005)] et mises en œuvre, y compris dans des contextes biologiques [Sbalzarini (2005), Rink (2005)]. Lorsque le nombre d'objets est grand, le problème de minimisation est essentiellement de nature combinatoire et des heuristiques sont nécessaires pour réduire l'espace à considérer. L'heuristique la plus simple consiste pour chaque objet, à l'instant précédent à considérer son plus proche voisin, généralement en limitant en outre la fenêtre locale de recherche parce qu'on a des *a priori*

sur la dynamique. Des algorithmes de programmation dynamique sont mis en œuvre sur ces problèmes d'optimisation [Tonissen (1996), Sage (2003)]. Les difficultés commencent lorsqu'il y a incertitude sur la position et/ou encore lorsque, dans un voisinage de petite taille, plusieurs objets se croisent, éventuellement disparaissent ou fusionnent ou se divisent ou bien encore lorsque les déplacements sont grands par rapport à l'échantillonnage temporel et enfin lorsqu'il y a des fausses détections. L'algorithme de référence par rapport au traitement de l'incertitude de la mesure est dû à Kalman (1960). Cet algorithme est aussi le point de départ d'algorithmes prenant en compte des fausses détections³ [Bar-Shalom (1975)] puis la proximité des objets⁴ [Fortmann (1983)]. Dans le cas d'objets présentant un rapport signal sur bruit faible, Blanding *et al* [Blanding (2007)] proposent plusieurs améliorations d'un algorithme de type ML-PDA⁵ autorisant le tracking de plusieurs cibles. Certains algorithmes [Cox et al. (1996), Streit et al. (2002)] maintiennent aussi en parallèle plusieurs hypothèses ou trajectoires possibles et reposent sur des modélisations bayésiennes ; ces algorithmes sont limités par le nombre de cibles.

Dans le cas du trafic intra-cellulaire, les méthodes les plus employées reposent sur des heuristiques pour la minimisation limitant le nombre d'associations possibles de façon à pouvoir toutes les évaluer⁶. Des dynamiques relatives à différents moteurs moléculaires ont été appréhendées par ce type d'algorithme [Gelles (1988), Yildiz (2003), Kural (2005)]. Dès que le rapport signal à bruit devient faible et que la densité des objets devient grande, ces méthodes ne fonctionnent plus très bien et l'utilisation de plusieurs instants peut parfois améliorer le tracking.

Méthodes de type TBD

Les méthodes TBD (Track Before Detect) ont connu un développement important ces dernières années, sans doute en raison de la forte modélisation sous-jacente et de leur puissance potentielle. Ces méthodes trouvent cependant leur limite dès que le nombre de cibles ou objets devient grand. Nous présentons ici deux types d'approche que nous consi-

³Algorithme PDA (Probabilistic Data Association).

⁴Algorithme JPDA (Joint Probabilistic Data Association).

⁵Maximum likelihood-Probabilistic Data Association

⁶Algorithmes glouton (greedy algorithms).

dérons relevant de cette famille de méthodes.

Méthode des kimogrammes Dans cette approche illustrée, sur la figure ?? (c.f. [Racine et al. (2006)]), avant de détecter et suivre les objets on commence par isoler les différentes parties du support du mouvement. En ce sens, nous considérons qu'il s'agit de méthodes de type TBD.

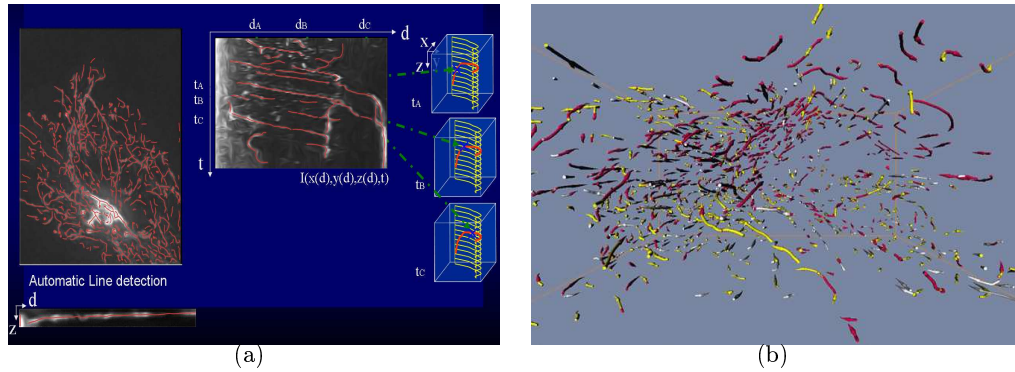


FIG. 5.1: Analyse de trajectoires 4D des vésicules par la méthode des kymogrammes. (a) Projections et kymogrammes ; (b) Trajectoires détectées et distribution des vitesses.

Dans le cas de mouvements s'effectuant essentiellement sur des supports curvilignes, on peut projeter les images correspondant aux différents instants sur une même image. Ainsi, si les objets sont clairs dans l'image et si I_t désigne l'image de l'instant t dans la séquence, l'image projetée est simplement $I^P = \max_{t=1, \dots, T} I_t$, soit $I^P(x) = \max_{t=1, \dots, T} I_t(x)$ pour toute position x dans l'image. Sur une telle image, on peut distinguer des structures curvilignes qu'on extrait par différentes méthodes [Steger (1998)]. Chacune de ces structures prises séparément supporte en principe le mouvement d'un ou plusieurs objets. On construit donc autant d'images ou kimogrammes que de structures en représentant, dans un système de coordonnées à deux dimensions (s, t) où s désigne l'abscisse curviligne sur la structure et t le temps, les valeurs $K(s, t) = I(x(s), t)$ où $x(s)$ désigne la position du point d'abscisse curviligne s dans l'image I . L'analyse de cette image révèle généralement elle aussi des structures curvilignes qui correspondent à des trajectoires individuelles. Cette approche a été automatisée et mise en œuvre sur des données 4D de trafic membranaire [Racine et al. (2007a)]. Elle permet d'extraire bon nombre de trajectoires partielles mais présente des

limites dès que : soit la densité locale des objets est trop grande, soit les mouvements sont trop rapides, soit les mouvements sont non directionnels.

Un exemple de méthode TBD probabiliste On peut se rendre compte de l'esprit de ces méthodes et de la complexité de leur mise en œuvre au travers de l'algorithme proposé par Ng *et al* [Li (2005b)].

Soit $\phi_t \triangleq \{x_t, K_t\}$ la variable d'intérêt. Ici $x_t = [x_{1,t}^T, \dots, x_{K_t,t}^T]^T$ est le vecteur d'état des K_t cibles présentes à l'instant t . On pose alors une équation d'évolution sur ce vecteur d'état et sur la variable K_t :

$$\begin{aligned} x_{k,t} &= f(x_{k,t-1}) + v_{k,t}, \quad k \in \{1, \dots, K_t\} \\ K_t &= K_{t-1} + \epsilon_{K_t} \end{aligned}$$

où f traduit l'*a priori* sur le mouvement des cibles et $v_{k,t}$ un bruit alors que ϵ_{K_t} est une v.a. discrète associée aux apparitions ou disparitions de cibles. Dans la modélisation de Ng *et al* une seule cible peut apparaître ou disparaître à chaque pas de temps.

Côté observations, on suppose qu'on observe un vecteur $y_t = [y_{1,t}^T, \dots, y_{M_t,t}^T]^T$ de M_t vecteurs de mesures indépendants à l'instant t . Ces mesures peuvent provenir de véritables cibles ou être des fausses alarmes. Une cible ne peut donner lieu qu'à un seul vecteur de mesures et certaines cibles peuvent ne pas être détectées. On pose aussi une équation d'état entre ces mesures et les variables d'état, soit :

$$y_{m,t} = g(x_{k,t}) + w_{m,t}, \quad m \in \{1, \dots, M_t\}$$

si le vecteur de mesure $y_{m,t}$ provient de la cible No k . Dans ce modèle $w_{m,t}$ est un bruit. Lorsque la mesure provient d'une fausse alarme, on suppose qu'elle est distribuée suivant une loi uniforme dans la région inspectée.

Pour parachever cette modélisation, il faut encore un vecteur de variables rendant compte de l'association entre mesures et cibles. Soit donc $\lambda_t = (\alpha_t, N_{C_t}, N_{D_t})$, une variable aléatoire avec $\alpha_t \in \{0, 1\}^{M_t}$ et $\alpha_{m,t} = k$ si la mesure $y_{m,t}$ provient de la cible No k alors que $\alpha_{m,t} = 0$, si la mesure ne provient d'aucune mesure. Les variables N_{C_t} et N_{D_t} désignent quant à elles le nombre de fausses alarmes et le nombre de cibles détectées.

Avec ces hypothèses la vraisemblance de y_t s'écrit :

$$p(y_t|\phi_t, \lambda_t) = V^{-N_{C_t}} \prod_{l \in \mathcal{I}_D} p(y_{l,t}|x_{\alpha_{l,t},t})$$

où \mathcal{I}_D est l'ensemble des indices des observations correspondant aux N_{D_t} cibles détectées.

Enfin , un *a priori* est mis sur $\lambda_t|K_t$:

$$p(\lambda_t|K_t) = p(\alpha_t|K_t, N_{C_t}, N_{D_t})p(N_{C_t})p(N_{D_t}|K_t)$$

où $p(\alpha_t|K_t, N_{C_t}, N_{D_t}) = \left[\frac{M_t}{N_{D_t}} \frac{K_t!}{(K_t - N_{D_t})!} \right]^{-1}$ est l'*a priori* sur le nombre d'hypothèses valides alors que N_{C_t} et $N_{D_t}|K_t$ suivent respectivement des lois de Poisson et binomiale.

Ceci fait, la distribution *a posteriori* de ϕ_t est fournie par la relation récursive

$$p(\phi_t|y_{1:t}) \propto p(y_t|\phi_t) \int p(\phi_t|\phi_{t-1})p(\phi_{t-1}|y_{1:t-1})d\phi_{t-1}$$

initialisée par une distribution $p(\phi_0)$ à choisir.

L'inférence à partir de la distribution *a posteriori* est faite avec des techniques d'échantillonnage d'importance séquentiels.

Cette méthode estime donc de manière récursive conjointement le vecteur d'état des cibles et l'association des cibles. D'autres hypothèses sont faites pour réduire la complexité numérique du calcul. Ce type d'algorithme se compare favorablement par rapport à l'algorithme JPDAF. Cette approche peut cependant être difficilement envisagée sur un grand nombre de cibles.

Face à la difficulté de mise en œuvre de toutes ces méthodes pour l'analyse de dynamique vésiculaires, notre point de vue a donc été de s'en écarter, tout en espérant *in fine* produire des statistiques assez robustes pour la comparaison de dynamiques. Dans les sections suivantes nous développons le cadre général de construction et surtout d'estimation de ces statistiques.

5.2 Cadre général

Soit Z une variable aléatoire (v.a.) observable à partir de la segmentation spatio-temporelle de la séquence d'images. Ainsi Z peut être le temps de passage d'un objet sur un

site de l'image, ou bien encore le nombre de passages d'objets en un site, ou d'autres caractéristiques basées sur la segmentation spatio-temporelle de la séquence et la séquence brute. On verra plus loin l'utilisation de la densité locale d'objets. La variable Z pourra être mono ou multi-dimensionnelle selon qu'on agrègera (fusionnera) une ou plusieurs caractéristiques.

On notera $X \equiv (V, R)$, un couple de variables aléatoires pour lequel V sera la v.a. d'intérêt, pour laquelle on proposera un estimateur de la distribution, et R une variable aléatoire directement observable à partir de la séquence (e.g. volume des objets, etc. .).

On suppose que Z et X sont des variables dépendantes et qu'il existe une relation de la forme :

$$Z = \Phi(X, U) \quad (5.1)$$

où U est un résidu indépendant ou très peu corrélé à X et $\Phi(., .)$ une fonction stochastique de X et de U supposée inconnue.

On note alors :

- $f_{(x,z)}(x, z)$ la distribution de probabilité jointe du couple (X, Z) .
- $f_x(x)$ la distribution de probabilité de X
- $f_{z/x}(z)$ la distribution de probabilité conditionnelle de Z sachant $X = x$

En utilisant alors la formule de Bayes, on peut écrire

$$f_{(x,z)}(x, z) = f_{z/x}(z)f_x(x)$$

Selon la formule (5.1), le terme $f_{z/x}(z)$ est complètement déterminé par la fonction stochastique $\Phi(., .)$ indépendamment de la loi $f_x(x)$ de X . On définit alors le noyau $H(., \cdot)$ comme suit :

$$H(x, z) = f_{z/x}(x, z)$$

qui est donc indépendant de la distribution de X , et ainsi, on obtient l'équation intégrale de Fredholm inhomogène de première espèce suivante :

$$f_z(z) = \int_x H(x, z)f_x(x)dx \quad (5.2)$$

qu'on peut également écrire comme suit :

$$f_z(z) = \int_{v,r} H(v, r, z)f_{v,r}(v, r)dvdr \quad (5.3)$$

Les densités f_z et f_x sont donc liées par une équation intégrale linéaire en f_x . Par ailleurs, on peut tenter d'estimer le noyau H hors ligne pourvu qu'on dispose d'un simulateur de dynamiques. L'espace des dynamiques à échantillonner peut être réduit par la connaissance *a priori* que l'on a des dynamiques à analyser dans le cas des études qui nous intéressent. Dans le cadre du trafic membranaire, on peut exploiter le fait que beaucoup de déplacements s'effectuent sur un réseau, que les objets ont une distribution de taille relativement proche d'une expérience à l'autre.

Dans le cas où V et R sont indépendantes, on peut agréger ces variables pour la construction du noyau. En effet, on a alors $f_x(x) = f_v(v)f_r(r)$ et l'équation (5.2) devient :

$$f_z(z) = \int_{v,r} H(x, z) f_v(v) f_r(r) dv dr \quad (5.4)$$

En posant

$$H(v, z) := \int_r H(x, z) f_r(r) dr$$

l'équation intégrale (5.4) se réduit alors à

$$f_z(z) = \int_v H(v, z) f_v(v) dv \quad (5.5)$$

Nous ferons souvent usage de cette formulation par la suite pour simplifier la présentation.

5.2.1 Estimation du noyau H

Elle repose sur une intégration numérique de l'équation intégrale 5.2. Pour cela, on utilise les discrétisations suivantes par rapport à z , v et r respectivement : $\{z_l = l \times \Delta_z, l = 1, \dots, L\}$, $\{v_i = v_0 + i \times \Delta_v, i = 1, \dots, M\}$, et $\{r_j = r_0 + j \times \Delta_r, j = 1, \dots, N\}$ et on utilise une méthode d'intégration numérique naïve pour laquelle la fonction à intégrer est approximée par une fonction constante par morceaux⁷. La version discrétisée de (5.4) est alors :

$$f_z(l) = \sum_{i=1}^M \sum_{j=1}^N H_{i,j,l} \times \Delta_v \times \Delta_r \times f_{v,r}(i, j), \quad l = 1, \dots, L \quad (5.6)$$

où les inconnues sont donc les $H_{i,j,l}$ et $f_{v,r}(i, j)$. Par commodité, on notera aussi par la suite $f_{v,r}(i, j)$ la quantité $\Delta_v \times \Delta_r \times f_{v,r}(i, j)$.

⁷On pourrait choisir une autre méthode d'intégration numérique toutefois.

Cependant, parce qu'on peut simuler des dynamiques, on peut effectuer l'apprentissage du noyau pour la famille de dynamiques qui nous intéresse. Ainsi, en injectant dans nos simulations les *a priori* qu'on a sur ces dynamiques, on peut générer P simulations correspondant à différentes distributions de probabilité $f_{v,r}(\cdot, \cdot)$ représentatives de la famille de dynamiques à analyser. Ces P simulations sont alors reliées au travers d'un système linéaire de $P \times L$ équations à $M \times N \times L$ inconnues :

$$f_z^s(l) = \sum_{i=1}^M \sum_{l=1}^L H_{i,j,l} \times f_{v,r}^s(i, j) , \quad l = 1, \dots, L; s = 1, \dots, P \quad (5.7)$$

qui n'est autre que la version discrétisée des équations intégrales suivantes :

$$f_z^s(z) = \int_{v,r} H(x, z) f_{v,r}^s(v, r) dr dv, s = 1, \dots, P \quad (5.8)$$

L'estimation naïve du noyau consiste à produire $P = M \times N$ simulations $S_{i,j}$, $i = 1, \dots, M$, $j = 1, \dots, N$ pour lesquelles les distributions de vitesse sont monocolores, soit :

$$f_{v,r}(p, q) = \begin{cases} 1 & \text{si } p \equiv i \text{ et } q \equiv j, \\ 0 & \text{sinon} \end{cases}$$

pour la simulation $S_{i,j}$.

On estimera alors H directement par :

$$\hat{H}_{i,j,l} = f_z^{i,j}(l), \quad i = 1, \dots, M, \quad j = 1, \dots, N, \quad l = 1, \dots, L$$

Cette approche naïve présente cependant plusieurs inconvénients. Du fait des nombreux degrés de liberté disponibles (taille des objets, nombre d'objets, ...) pour simuler des dynamiques correspondant à une distribution de vitesses donnée, baser l'estimation de $H_{i,j}$ sur une seule simulation n'est pas très raisonnable même si on peut effectuer la simulation sur une durée assez longue. En pratique, il faudrait pouvoir apprécier la diversité des $H_{i,j}$ pour chaque (i, j) fixé. Ensuite, du point de vue de l'identifiabilité de $f_{v,r}$, il est important que H soit bien conditionné. Les facteurs qui peuvent influencer ce conditionnement sont multiples, ils peuvent avoir trait à la discrétisation (pas dans les différentes dimensions) mais aussi à la simulation. Pour améliorer le conditionnement, on peut essayer de conditionner la

variable Z par rapport à davantage de variables auxiliaires (i.e. augmenter le nombre de variables dans R), ou encore augmenter la dimension de Z et donc fusionner plusieurs sources d'informations pour l'estimation de la distribution de V . Dans cette thèse nous n'avons pas consacré beaucoup d'efforts à l'estimation du noyau, faute de temps bien évidemment, et seules expériences numériques reportées en section 5.5 permettent d'apprécier l'allure du noyau dans le cadre de variables Z reliées à des gradients temporels de densité locales d'objets.

Avant de passer à l'estimation de f_v , on peut se demander si l'estimation de f_v et de H sont obligatoirement séparées? ceci nous a renvoyé vers la séparation de sources en traitement du signal et de l'image.

Autre regard : séparation de sources

L'équation (5.7) peut être considérée sous plusieurs angles différents selon ce qu'on veut extraire.

En effet, cette équation peut être perçue comme modélisant un problème de séparation de sources. Les sources seraient constituées des $M \times N$ signaux $H_{i,j}$ de longueur L alors que les quantités $f_{v,r}(i,j)$ seraient vues comme des coefficients de mélange de ces sources. A une source correspondrait une distribution. Si on a de bonnes raisons de penser que les sources et f_z sont des signaux supportant des représentations parcimonieuses, une estimation simultanée des sources et des coefficients de mélange peut être envisagée à l'aide des algorithmes JADE, FastICA [Hyvarinen (2001)] ou MCA [Starck (2004)]. Dans le cas de l'utilisation simultanée de plusieurs caractéristiques (Z multi-dimensionnel) l'algorithme MMCA [Bobin (2005)] peut apporter plus de robustesse dans l'estimation.

Avant de clore cette section, nous réécrivons les systèmes d'équation utilisés pour l'estimation de H dans le cas où l'on peut supposer V et R indépendants.

Ceci reviendra donc à supprimer le conditionnement par rapport à R , soit par conséquent :

$$f_z(l) = \sum_{i=1}^M H_{i,l} \times f_v(i), \quad l = 1, \dots, L \quad (5.9)$$

et en effectuant $P = M$ simulations avec différentes distributions de probabilité de V , on obtient alors après discrétisation le système d'équations linéaires suivant :

$$f_z^s(l) = \sum_{i=1}^M H_{i,l} \times f_v^s(i) , \quad l = 1, \dots, L; \quad s = 1, \dots, P \quad (5.10)$$

Pour estimer les $H_{i,l}$ on considère P simulations S_i de dynamiques vérifiant :

$$f_v(p) = \begin{cases} 1 & \text{si } p \equiv i , \\ 0 & \text{sinon} \end{cases}$$

L'estimateur naïf des $H_{i,l}$ est alors :

$$\hat{H}_{i,l} = f_z^i(l), l = 1, \dots, L$$

Nous utiliserons plutôt ces équations dans les sections à suivre.

5.2.2 Estimation de la distribution de probabilité empirique $f_v(\cdot)$

On analyse une dynamique au travers de sa segmentation spatio-temporelle et d'un ensemble de caractéristiques (ou "features") ou réalisations de la v.a. Z . On sait que f_z peut être entachée d'erreurs dues au processus de segmentation compliqué par le faible rapport signal à bruit et les fausses alarmes. On sait aussi qu'un 'représentant' a été choisi pour H alors qu'il vaudrait sans doute mieux considérer H aléatoire. On modélise alors ces sources d'incertitude comme suit :

$$f_z(l) \triangleq f_z(l) + \delta_l \nu_l \quad l = 1, \dots, L \quad \text{avec } \nu_l \sim \mathcal{N}(0, 1)$$

et pour le noyau :

$$\hat{H}_{i,l} \triangleq H_{i,l} + \alpha_{i,l} \xi_{i,l} \quad \text{avec } \xi_{i,l} \sim \mathcal{N}(0, 1)$$

On supposera en outre que les variables ν_l , $\xi_{i,l}$, $i = 1, \dots, M$, $l = 1, \dots, L$ sont indépendantes. En supposant que $f_z = H f_v$, on en déduit donc que :

$$\hat{f}_z(l) = \sum_{i=1}^M \hat{H}_{i,l} \times f_v(i) + \sigma_l \epsilon_l , \quad l = 1, \dots, L; \quad (5.11)$$

avec

$$\epsilon_l = \left[\delta_l^s \nu_l^s - \sum_{i=1}^M \alpha_{i,l} f_v^s(i) \xi_{i,l} \right]$$

et

$$\sigma_l = \sqrt{\sum_{i=1}^M \alpha_{i,l}^2 f_v^s(i)^2 + \delta_l^{s2}}$$

Pour estimer f_v , on est donc amené à résoudre un problème inverse linéaire avec contraintes linéaires, soit :

$$\begin{aligned} \hat{f}_z(l) &= \sum_{i=1}^M \hat{H}_{i,l} \times f_v(i) + \sigma_l \epsilon_l, \quad l = 1, \dots, L; \\ \text{t.q.} \quad &\begin{cases} \sum_{i=1}^M f_i^v = 1 \\ f_i^v \geq 0, \quad i = 1, \dots, M \end{cases} \end{aligned} \quad (5.12)$$

On est donc en présence d'un problème linéaire inverse contraint dont la forme générale peut s'écrire :

$$\begin{aligned} Y &= A\theta + \epsilon \\ g(\theta) &\geq 0 \end{aligned} \quad (5.13)$$

où Y est le vecteur des variables observables du processus, A est une matrice $n \times p$ avec $n \geq p$, ϵ est un bruit qu'on suppose Gaussien ayant une matrice de variances-covariances Σ , θ est le paramètre que l'on cherche à estimer, et $g(\theta) \geq 0$ modélise les contraintes linéaires d'inégalité (*l'a priori*) que l'on a sur les valeurs réalisables de θ .

Une des classes de problèmes inverses contraints les plus étudiées dans la littérature notamment pour la restaration d'images en imagerie médicale sont "les problèmes inverses avec une contrainte de non-négativité de la solution". Ceux-ci se formulent comme suit :

$$\begin{aligned} Y &= A\theta + \epsilon \\ \theta &\geq 0 \end{aligned} \quad (5.14)$$

Le problème 5.12 peut aussi être réécrit sous la forme d'un problème inverse avec contraintes de positivité sur les variables, pour cela il suffit d'intégrer la contrainte d'égalité à H et f_z , et donc de définir

$$\hat{H}_{i,l}^* = \begin{cases} \hat{H}_{i,l}, 1 \leq i \leq M; 1 \leq l \leq N \\ 1, 1 \leq i \leq M; l \equiv N + 1; \end{cases}$$

et

$$\hat{f}_z^*(l) = \begin{cases} \hat{f}_z^s(l), 1 \leq l \leq L \\ 1, l \equiv L + 1 \end{cases}$$

On réécrit alors (5.12) comme suit :

$$\hat{f}_z^*(l) = \sum_{i=1}^M \hat{H}_{i,l}^* \times f_v(i) + \sigma_l \epsilon_l, \quad l = 1, \dots, L + 1; \\ \text{t.q. } f_i^v \geq 0, i = 1, \dots, M \quad (5.15)$$

où ϵ_{L+1} peut être considérée comme une Gaussienne standard d'écart type $\sigma_{L+1} \rightarrow 0$. En pratique, on peut prendre σ_{L+1} suffisamment petit.

5.3 Méthodes de résolution du problème linéaire inverse avec une contrainte de non-négativité de la solution (5.14)

Il existe deux catégories de méthodes pour résoudre le problème inverse avec une contrainte de non-négativité de la solution (5.14). La première catégorie résout (5.14) sans la contrainte de non-négativité puis annule toutes les composantes négatives de la solution. Une deuxième catégorie résout (5.14) en prenant en considération la contrainte de non-négativité de la solution. Nous allons présenter quelques approches possibles des deux catégories pour résoudre (5.14) et les appliquer au problème (5.15) .

5.3.1 Résolution par le critère des moindres carrés pondérés

Un estimateur au sens des moindres carrés pondérés minimise

$$J(\theta) \triangleq \|Y - A\theta\|_{\Sigma^{-1}}^2$$

sous les contraintes de positivité. Une solution approchée de ce problème est fournie par :

$$\hat{\theta}_{MCP} \triangleq \left[\left(A' \Sigma^{-1} A \right)^{-1} A' \Sigma^{-1} Y \right]^+$$

où $\left(A' \Sigma^{-1} A \right)^{-1}$ est une matrice inverse ou à défaut une pseudo-inverse de $A' \Sigma^{-1} A$. \square^+ désigne la projection sur du vecteur sur \mathbb{R}^{M+} .

D'un point de vue numérique, ce problème est plus souvent résolu par des méthodes qui évitent le recours à des matrices dont le conditionnement est le carré de celui de A . On pourra consulter Gill *et al* [Gill (1981)], p 181-2 pour plus de détails. On peut aussi noter que la solution de ce type de problème présente la particularité d'avoir la même 'sensibilité' autour de l'optimum dans les différentes directions actives [Bertsekas (1995)] ou encore $\frac{\partial J}{\partial \theta_i}(\theta^*) = \lambda^*$ pour $\theta_i > 0$ et $\frac{\partial J}{\partial \theta_i}(\theta^*) = \geq \lambda^*$ pour $\theta_i = 0$.

On pourrait également préférer de résoudre le problème quadratique suivant :

$$\begin{aligned} \|Y - A\theta\|_{\Sigma^{-1}}^2 &\rightarrow \text{Min} \\ \theta &\geq 0 \end{aligned}$$

En appliquant cette dernière méthode à notre problème pour estimer la densité de probabilité discrétisée $f_v(\cdot)$, cela revient donc à résoudre le programme quadratique suivant :

$$\begin{aligned} \sum_{l=1}^L \frac{(\hat{f}_l^z - \sum_{i=1}^M \hat{H}_{i,l} \times f_i^v)^2}{\sigma_l^2} &\rightarrow \text{Min} \\ \text{t.q.} \begin{cases} \sum_{i=1}^M f_i^v = 1 \\ f_i^v \geq 0, \quad i = 1, \dots, M \end{cases} & \end{aligned} \quad (5.16)$$

Toutefois, une telle méthode peut donner des solutions très biaisées même pour des petites intensités du bruit. C'est pourquoi une régularisation de la solution est généralement considérée pour résoudre efficacement un problème inverse de la forme (5.14). Le but d'une telle régularisation de la solution est d'introduire un *a priori* sur la solution attendue afin de la rendre moins sensible aux perturbations dues au bruit. Nous en présenterons dans la suite quelques méthodes utilisant une régularisation pour résoudre le problème inverse contraint.

5.3.2 Résolution par la méthode de troncature des valeurs singulières

Si on oublie la contrainte de non-négativité de la solution dans le problème linéaire inverse (5.14), et si la matrice A est de rang plein, i.e. elle n'a aucune valeur singulière nulle, alors l'estimateur des moindres carrés pondérés de θ est donné par la fomule

$$\hat{\theta}_{MCP} = \left(A' \Sigma^{-1} A \right)^{-1} A' \Sigma^{-1} Y$$

Cet estimateur $\hat{\theta}_{MCP}$ est sans biais mais possède une variance d'autant plus forte que le rapport $\frac{\nu_{max}}{\nu_{min}}$ entre sa plus grande valeur singulière et sa plus petite valeur singulière est grand. Généralement, la matrice A est mal-conditionnée ; ce qui signifie qu'il existe un rang p à partir duquel le rapport des valeurs singulières $\frac{\nu_{max}}{\nu_k} \rightarrow \infty$ pour tout $k \geq p$. Par conséquent, la matrice inverse ou même la matrice pseudo-inverse de $A' \Sigma^{-1} A$ n'ont pas de sens. La méthode de troncature des valeurs singulières procède alors comme suit. Soit (U, D, V) la décomposition en valeurs singulières de la matrice $A' \Sigma^{-1} A$, i.e.

$$A' \Sigma^{-1} A = U D \{ \nu_k \} V'$$

où $D \{ \nu_k \}$ dénote une matrice diagonale de taille (n, p) contenant sur sa diagonale les valeurs $\nu_k, k = 1, \dots, p$ ordonnées par ordre décroissant, $U(n, n)$ et $V(M, M)$ sont deux matrices orthonormales vérifiant $U' * U = I_n$ et $V' * V = I_p$, I_n et I_p étant deux matrices identités de taille n et p respectivement. Dans la méthode de troncature des valeurs singulières, on considère que les vecteurs propres associés aux valeurs singulières en deça d'un certain seuil ne doivent pas entrer dans l'estimateur. Ceci revient à définir

$$\hat{\theta}_{TVS} = V D_{\tau}(\mu_k) U' \Sigma^{-1} Y$$

Dans cette formule $D_{\tau}(\mu_k)$ désigne une matrice diagonale dont les coefficients diagonaux vérifient :

$$\mu_k = \begin{cases} \frac{1}{\nu_k}, & \text{si } \nu_k \geq \tau \\ 0, & \text{sinon} \end{cases}$$

On peut démontrer qu'une telle estimateur de θ réalise un certain compromis entre le biais et variance, mais qu'il est seulement sous-optimal [Fessler et al. (1995)]. Pour le problème contraint, on peut à nouveau choisir comme estimateur :

$$\hat{\theta}_{TVSP} \triangleq \left[\hat{\theta}_{TVS} \right]^+$$

Enfin, on prend comme solution de (5.14) le vecteur $\hat{\theta}$ après avoir mis à zéro toutes ses valeurs négatives.

5.3.3 Résolution par pondération des valeurs singulières

On écrit la décomposition en valeurs singulières de $\Sigma^{-\frac{1}{2}}A$:

$$\Sigma^{-\frac{1}{2}}A = UD\{\nu_k\}_nV'$$

On définit l'estimateur DVS (Décomposition en Valeurs Singulières) de θ par rapport à un vecteur de poids $w \equiv (w_k)_{k=1\dots,p}$ comme suit :

$$\hat{\theta}(w) = \sum_{k=1}^p w_k v_k u'_k \Sigma^{-\frac{1}{2}} Y = VD\{w_k\}U'\Sigma^{-\frac{1}{2}}Y$$

où u_k et v_k sont les k -ièmes vecteurs colonnes de U et de V respectivement. On démontre facilement que la matrice de variances-covariances de $\hat{\theta}(w)$ s'écrit :

$$\text{Cov}\{\hat{\theta}(w)\} = VD\{w_k^2\}V'$$

et que le biais pour une telle estimation de θ s'écrit :

$$b(\hat{\theta}(w)) = E(\hat{\theta}(w)) - \theta = -VD\{1 - w_k \nu_k\}V'\theta$$

et posant $x := V'\theta$ on obtient alors

$$\begin{aligned} \|b(\hat{\theta}(w))\|^2 &= \|D\{1 - w_k \nu_k\}V'\theta\|^2 \\ &= \theta'VD\{(1 - w_k \nu_k)^2\}V'\theta \\ &= \sum_{k=1}^p (1 - w_k \nu_k)^2 x_k^2 \end{aligned}$$

On remarque qu'en choisissant des w_k petits, on réduit la variance de $\hat{\theta}$ mais en même temps on augmente son biais, et vice-versa. Il est alors intéressant d'estimer θ en réalisant un bon compromis entre le biais et la variance. L'idée alors de l'auteur dans [Fessler et al. (1995)] est de minimiser la variance pour une certaine quantité de biais C . La trace de la matrice de variances-covariances associées à $\hat{\theta}(w)$ s'écrit

$$\text{tr}(\text{Cov}\{\hat{\theta}(w)\}) = \text{tr}(VD\{w_k^2\}V') = \text{tr}(D\{w_k^2\}) = \sum_{k=1}^p w_k^2$$

et comme le biais dépend de x , inconnu, on considère un vecteur $q \geq 0$ qui pondère chaque composante du terme du biais et vérifie $\sum_{k=1}^p q_k = 1$. L'estimateur proposé par Fessler et al. (1995) cherche à minimiser la variance tout en contrôlant le biais ; il s'obtient, en considérant la contrainte de non-négativité de la solution, comme solution du problème

$$\begin{aligned} & \sum_{k=1}^p w_k^2 \rightarrow \text{Min} \\ \text{t.q.} \quad & \begin{cases} \sum_{k=1}^p (1 - w_k \nu_k)^2 q_k^2 \leq C \\ VD\{w_k\} U' \Sigma^{-\frac{1}{2}} Y \geq 0 \end{cases} \end{aligned} \quad (5.17)$$

On peut également procéder de façon légèrement différente en se donnant un $\lambda > 0$ et en minimisant une somme pondérée du terme de variance et du terme de biais pour se ramener à un programme quadratique plus facile à résoudre que (5.17), soit :

$$\begin{aligned} & \sum_{k=1}^p w_k^2 + \lambda \sum_{k=1}^p (1 - w_k \nu_k)^2 q_k^2 \rightarrow \text{Min} \\ \text{t.q.} \quad & VD\{w_k\} U' \Sigma^{-\frac{1}{2}} Y \geq 0 \end{aligned} \quad (5.18)$$

Bien entendu, toute la difficulté des deux approches réside dans le choix du vecteur q et de la quantité de biais C pour (5.17), ou du vecteur q et du paramètre de compromis biais-variance λ pour (5.18).

5.3.4 Résolution par pénalisation du critère de moindres-carrés pondérés

Régularisation de Tikhonov

Une autre idée pour résoudre un problème inverse en utilisant un critère de moindres carrés pondérés et de rajouter un terme à sa fonction objectif pour le rendre strictement convexe. Contrairement au problème inverse original, un tel problème strictement convexe possède bien entendu une unique solution. La méthode de régularisation de Tikhonov est l'une des méthodes qui procèdent ainsi. L'application de la régularisation de Tikhonov au problème (5.14) revient donc à résoudre le problème quadratique strictement convexe suivant

pour une certaine matrice définie positive symétrique R et pour un certain $\lambda > 0$:

$$\begin{aligned} \|Y - A\hat{\theta}\|_{\Sigma^{-1}}^2 + \lambda\hat{\theta}'R\hat{\theta} &\rightarrow \text{Min} \\ \text{t.q. } \hat{\theta} &\geq 0 \end{aligned} \tag{5.19}$$

On montre [Fessler et al. (1995)] que pour un choix approprié du paramètre de régularisation λ et de la matrice R ($R = V\text{diag}\{q_k\}V'$ avec V orthogonale) que les solutions des problèmes (5.19) et (5.17) non contraints coïncident. Que ce soit pour cette méthode ou celles de la section précédente, le paramétrage reste à la discrétion de l'utilisateur !

Résolution par sélection de modèle

Le grand inconvénient des méthodes de régularisation que nous avons présentées précédemment est qu'elles dépendent de certains paramètres qui sont difficiles à ajuster de façon automatique. La théorie de la sélection de modèle permet quant à elle de s'affranchir dans une certaine mesure du délicat choix des paramètres de régularisation en imposant un certain modèle sur les données. Comme nous l'avons expliqué en détails dans le chapitre 3, le but d'une approche par sélection de modèle est de choisir parmi une librairie de modèles celui qui décrit au mieux des données observées tout en tenant compte de la présence du bruit.

Il est à noter qu'il y a ici une subtilité importante avec l'approche de sélection de modèle basée sur la pénalisation d'un critère de moindres carrés pondérés présentée dans le chapitre 3 qui tente d'estimer $A\theta$ plutôt que θ . Or, imposer un modèle sur $A\theta$ ne revient pas nécessairement à imposer un modèle sur θ et vice versa. Pour expliciter cette situation, on laisse tomber la contrainte de non-négativité de θ dans (5.14) et on tente d'estimer $A\theta$ en utilisant par exemple le second critère de sélection de modèle (*SCP*) que nous avons présenté dans le chapitre 3. Cela revient donc à minimiser sur tous les modèles de $A\theta$ un critère pénalisé de la forme

$$\|Y - \widehat{A\theta}\|_{\Sigma^{-1}}^2 + \rho D(\widehat{A\theta}) \tag{5.20}$$

où $D(\widehat{A\theta})$ désigne la dimension du modèle de $\widehat{A\theta}$ et ρ est une constante qui dépend uniquement de la famille de modèles de $(A\theta)$ et non pas ponctuellement de chaque modèle $(A\theta)$;

elle peut être calculée conformément à la proposition 3.2 (*c.f.* chapitre 3).

Une question intéressante qui se pose alors (on considère toujours le problème inverse (5.14) sans l'hypothèse de non-négativité de θ) est dans quelle mesure le vecteur θ peut être estimé en utilisant une approche par sélection de modèle qui tente par exemple de minimiser un critère pénalisé de la forme

$$\|Y - A\hat{\theta}\|_{\Sigma^{-1}}^2 + \eta D(\hat{\theta}) \quad (5.21)$$

où $D(\hat{\theta})$ est la dimension du modèle imposé à $\hat{\theta}$ et η est une constante qui dépend uniquement de la famille de modèles de θ et non pas ponctuellement de chaque modèle de θ qui est à définir. Une approche intéressante d'estimation de θ appelée procédure de sélection de variables canonique tentant alors de minimiser un critère comparable à (5.21) dans un cadre de régression multiple est décrite dans [Foster et al. (1993)]. La suite de cette sous-section est donc dédiée à la description de cette approche.

D'abord, pour être fidèle à l'approche originale décrite dans [Foster et al. (1993)], on considère le cas homoscédastique pour lequel $\Sigma \equiv \sigma^2 I_p$, où I_p désigne la matrice identité de taille $p \times p$. On définit l'estimateur des moindres carrés $\hat{\theta}_\gamma$ de θ par rapport au vecteur $\gamma = (\gamma_i)_{i=1, \dots, M}$ ⁸ indiquant les positions des composantes de $\hat{\theta}$ qui sont forcées à zéro, i.e. $\gamma_i = 0 \Rightarrow \hat{\theta}_\gamma = 0$, et $\hat{\theta}_\gamma \neq 0 \Rightarrow \gamma_i = 1$. La quantité $|\gamma|$ est alors définie comme le nombre de composantes non-nulles de γ . On définit maintenant le vecteur η par $\eta_i = \mathbb{1}(\theta_i = 0)$. L'estimateur moindres-carrés $\hat{\theta}_\gamma$ de θ par rapport au vecteur γ est donc donné par la formule

$$\hat{\theta}_\gamma = (D_\gamma X' X D_\gamma)^{-1} D_\gamma X' Y$$

où D_γ désigne la matrice diagonale vérifiant $D_\gamma(i, i) \equiv \gamma_i$, et A^{-1} désigne la matrice pseudo-inverse de A . Donc, en particulier, pour $\gamma \equiv \eta$, on a

$$\hat{\theta}_\eta = (D_\eta X' X D_\eta)^{-1} D_\eta X' Y$$

On définit maintenant le risque prédictif \mathcal{R} d'un estimateur $\hat{\theta}_\gamma$ de θ comme suit :

$$\mathcal{R}(\theta, \hat{\theta}_\gamma) \triangleq \mathbb{E} \left[\|X\hat{\theta}_\gamma - X\theta\|^2 \right]$$

⁸En termes de sélection de modèle, un tel vecteur est appelé "modèle".

et on définit l'inflation du risque prédictif RI de $\hat{\theta}$ comme suit :

$$RI(\gamma) \triangleq \sup_{\theta} \frac{\mathcal{R}(\theta, \hat{\theta}_{\gamma})}{\mathcal{R}(\theta, \hat{\theta}_{\eta})}$$

La quantité $RI(\gamma)$ mesure l'augmentation du risque d'estimation de $\hat{\theta}_{\gamma}$ par un critère de moindres carrés en utilisant comme vecteur qui indique les positions des composantes nulles de θ le vecteur γ au lieu du vrai vecteur η . Au fait, l'estimation de θ basée sur la minimisation de RI est plus sûre que l'estimation de θ basée sur \mathcal{R} . En effet, tandis qu'une petite valeur de RI correspond à une bonne performance de $\hat{\theta}_{\gamma}$, une petite valeur de \mathcal{R} ne correspond pas nécessairement à une bonne performance de $\hat{\theta}_{\gamma}$ puisque généralement pour les problèmes mal-posés, on peut toujours trouver un vecteur θ_1 très différent de θ mais vérifiant que la quantité $\|X\hat{\theta}_{\gamma} - X\theta\|^2$ est très petite. Les auteurs dans [Foster et al. (1994)] proposent alors d'estimer θ en utilisant un critère pénalisé de la forme :

$$\text{Crit}_{\mu}(\gamma) \triangleq \frac{\|Y - A\hat{\theta}_{\gamma}\|^2}{\sigma^2} + \|\gamma\|_0 \mu \quad (5.22)$$

où μ est une constante dont la valeur optimale vérifie

$$\mu = \arg \min_{\nu \geq 0} RI(\hat{\gamma}(\nu))$$

avec $\hat{\gamma}(\nu)$ est donnée par la formule

$$\hat{\gamma}(\nu) = \arg \min_{\gamma} \text{Crit}_{\nu}(\gamma)$$

Afin de démontrer la pertinence de leur procédure de sélection de variable canonique, les auteurs montrent que $RI(\hat{\theta}_{\gamma_{L_s}}) = M$ pour l'estimateur des moindres carrés $\hat{\theta}_{\gamma_{L_s}} = (X'X)^{-1}X'Y$. Ils montrent aussi que la valeur optimale de μ , lorsque $X'X$ est diagonale est $\mu \cong 2 \log M$ et que cette valeur de μ réalise une inflation du risque $RI(\mu) \cong 2 \log M$, significativement inférieure à la valeur M de l'inflation du risque pour $\mu = 0$ (i.e. correspondant à l'estimateur moindres carrés naïfs $\hat{\theta}_{\gamma_{L_s}}$). Pour une matrice $X'X$ quelconque, la valeur de μ minimisant RI n'est malheureusement pas disponible de façon explicite puisqu'elle dépend de la structure de corrélation de $X'X$, mais ils démontrent que la valeur $\mu = 2 \log M$ est asymptotiquement la plus sûre quand $M \rightarrow \infty$.

Estimation de f_v par la méthode de Foster & George On pose $\mu_M := 2 \log(M)$, et on propose d'utiliser cette approche pour estimer le vecteur f_i^v . En considérant les contraintes de non-négativité de f_i^v , ceci revient donc à résoudre le problème d'optimisation à variables mixtes (discrettes et continues) suivant :

$$\sum_{l=1}^L \frac{(\hat{f}_l^z - \sum_{i=1}^M \hat{H}_{i,l} \times f_i^v)^2}{\sigma_l^2} + \left(\sum_{i=1}^M \gamma_i \right) \mu_M \rightarrow \text{Min}$$

$$\text{t.q.} \begin{cases} \sum_{i=1}^M f_i^v = 1, \\ f_i^v \geq 0, & i = 1, \dots, M \\ f_i^v \leq \gamma_i, & i = 1, \dots, M \\ \gamma_i = 0 \text{ ou } \gamma_i = 1, & i = 1, \dots, M \end{cases} \quad (5.23)$$

La résolution de ce type de problème est onéreuse dès que le nombre de variables discrètes augmente. Il peut cependant être facilement transformé en un problème de programmation quadratique :

$$\sum_{l=1}^L \frac{(\hat{f}_l^z - \sum_{i=1}^M \hat{H}_{i,l} \times f_i^v)^2}{\sigma_l^2} + \lambda \sum_{i=1}^M \gamma_i (1 - \gamma_i) + \left(\sum_{i=1}^M \gamma_i \right) \mu_M \rightarrow \text{Min}$$

$$\text{t.q.} \begin{cases} \sum_{i=1}^M f_i^v = 1, \\ f_i^v \geq 0, & i = 1, \dots, M \\ f_i^v \leq \gamma_i, & i = 1, \dots, M \\ \gamma_i \leq 1, & i = 1, \dots, M \end{cases} \quad (5.24)$$

On peut en effet montrer que le problème 5.24 est équivalent au problème 5.25 pour vu que λ soit suffisamment grand.

En effet, le terme $\lambda \sum_{i=1}^M \gamma_i (1 - \gamma_i)$ dans la fonction objectif contraint les γ_i à prendre une valeur 0 ou 1 quand λ est suffisamment grand, et remarquant qu'on a la relation suivante pour la solution optimale (f_v^*, γ_i^*) de ce dernier programme quadratique :

$$f_i^{v*} \equiv 0 \Leftrightarrow \gamma_i^* \equiv 0$$

$$f_i^{v*} > 0 \Leftrightarrow \gamma_i^* \equiv 1$$

En effet, avec la contrainte $f_i^v \leq \gamma_i$, on a si $\gamma_i^* \equiv 0$ alors $f_i^{v*} \equiv 0$, et si $f_i^{v*} \equiv 0$, alors γ_i^* prend nécessairement la valeur 0, sinon, dans le cas contraire (i.e. s'il prenait la valeur 1),

il induirait une valeur de la fonction objectif qui est plus grande. Si maintenant $\gamma_i^* \equiv 1$, alors nécessairement $f_i^{v*} \equiv 0$ car sinon, la solution (f_v^*, γ^*) ne serait pas optimale étant donné que la même solution et auquel on met $\gamma_i^* \equiv 0$ aura une valeur de la fonction objectif qui est plus petite. Enfin, si $f_i^{v*} > 0$ alors compte tenu de la contrainte $f_i^v \leq \gamma_i$, on a nécessairement $\gamma_i^* \equiv 1$. Donc l'expression $\sum_{i=1}^M \gamma_i$ correspond à la dimension (nombre de valeurs non nulles) de la solution f_v^* , de sorte que le terme $(\sum_{i=1}^M \gamma_i)\mu_M$ dans la fonction objectif correspond au terme de pénalité en utilisant la procédure de sélection de variable canonique décrite plus haut.

Toutefois, un tel programme quadratique (5.24) n'est pas convexe et un minimum global n'est donc pas garanti. Une alternative pour réduire la complexité calculatoire de (5.23) serait alors d'utiliser l'algorithme de Branch & Bound qui utilise la stratégie "diviser pour mieux regner" pour le résoudre. Un tel algorithme construit (récursivement) une arborescence de programmes quadratiques en fixant à chaque branche de l'arborescence la valeur d'un certain γ_i à 0 ou à 1. Soit alors G^* la valeur minimale courante de la fonction objectif du programme (5.23). A l'itération 0 de l'algorithme, on prend soit $G^* = \infty$ ou de préférence

$$G^* = G_* + D_*\mu_M$$

où G_* et D_* sont respectivement la valeur de la fonction objectif et la dimension (nombre de composantes non nulles) de la solution du programme quadratique (5.16). Considérant maintenant un nœud quelconque de cette arborescence et appelant $\{\gamma_i^*, i = 1, \dots, n^*\}$ l'ensemble des indices des γ_i déjà fixés (à 0 ou à 1) en commençant par le nœud source de l'arborescence et en finissant par le nœud courant. On résout alors le programme quadratique suivant :

$$\sum_{l=1}^L \frac{(\hat{f}_l^z - \sum_{i=1}^M \hat{H}_{i,l} \times f_i^v)^2}{\sigma_l^2} \rightarrow \text{Min}$$

$$\text{t.q.} \begin{cases} \sum_{i=1}^M f_i^v = 1, \\ f_i^v \geq 0, & i = 1, \dots, M \\ f_i^v \leq \gamma_i^*, & i = 1, \dots, n^* \end{cases} \quad (5.25)$$

et soit G^{n^*} la valeur optimale de la fonction objectif trouvée. On réalise le test suivant :

$$G^{n^*} + \left(\sum_{i=1}^{n^*} \gamma_i^* \right) \mu_M \stackrel{?}{<} G^*$$

- Si un tel test n'est pas vérifié, alors il n'y a aucune chance d'améliorer la fonction objectif du programme mixte (5.23) dans les nœuds fils du nœud courant, on peut alors couper cette branche.
- Si un tel test est vérifié, alors :
 - Si $n^* \equiv M$ alors on met à jour la valeur G^* en posant

$$G^* := G^{n^*} + \left(\sum_{i=1}^{n^*} \gamma_i^* \right) \mu_M$$

- Si $n^* < M$, on crée alors deux nœuds fils du nœud courant. Pour le premier fils, on pose $\gamma_{n^*+1}^* := 0$ et on considère l'ensemble des γ_i fixés : $\{\gamma_i^*, i = 1, \dots, n^*+1\}$. Pour le second fils, on pose $\gamma_{n^*+1}^* := 1$ et on considère l'ensemble des γ_i fixés : $\{\gamma_i^*, i = 1, \dots, n^*+1\}$, et ainsi de suite.

Et ainsi, à la fin de l'algorithme, on aura trouvé le minimum global de (5.23). Même si un tel algorithme de Branch & Bound réduit considérablement le coût calculatoire de résolution du programme mixte (5.23), sa convergence vers la solution optimale en un temps raisonnable n'est pas garantie. Par conséquent, dans un premier temps, on peut commencer par utiliser l'algorithme de Branch & Bound dans le but de trouver la solution exacte du programme (5.23), et si après un certain temps, il s'avère qu'une solution optimale n'est toujours pas trouvée, il serait alors intéressant d'envisager d'utiliser une autre méthode permettant de trouver en un temps raisonnable une bonne solution de (5.23) comme une méthode de type "forward stepwise selection" [Miller (1990)].

5.4 Application de l'approche décrite ainsi que des estimateurs de densité à noyau pour estimer la dynamique globale de vésicules dans une séquence de vidéo-microscopie

Nous présentons dans cette section une méthode utilisant l'approche précédemment décrite en section 1.6 et les estimateurs de densité à noyau (Kernel Density Estimators (KDE) en anglais) afin d'estimer la dynamique globale de vésicules dans une séquence de vidéo-microscopie $3D + T$. Nous nous intéressons tout particulièrement à la vitesse "V"

(module et directionnalité) de tels vésicules impliqués dans le trafic membranaire entre le Golgi et les différents compartiments cellulaires. Pour ce faire, il est d'abord important de mettre en évidence une statistique Z qu'on peut calculer facilement sur une séquence de vidéo-microscopie et qui s'écrit sous la forme

$$Z = \Phi(X_1, X_2, U)$$

où X_1 est la variable aléatoire d'intérêt qui est cachée et pour laquelle on cherche à restaurer (estimer) la distribution de probabilité dans la séquence de vidéo-microscopie, X_2 est une variable aléatoire qui est observable (i.e. que l'on peut calculer facilement) dans la séquence, et U est supposé un bruit aléatoire statistiquement indépendant de X_1 et de X_2 . Par exemple, si l'on veut restaurer la distribution du module du vecteur de vitesse, on a $X_1 \equiv \|V\|$, et si l'on veut restaurer à la fois le module du vecteur de vitesse et sa directionnalité, alors on a $X_1 \equiv V$. Evidemment, dans les deux cas, on n'a pas la même complexité de calcul et il serait intéressant d'étudier les deux cas séparément.

5.4.1 Restauration de la distribution empirique du module de la vitesse

On suppose alors que la détection des vésicules est réalisée dans toute la séquence. On dispose alors à chaque instant $t = 1, \dots, T$ des centres de gravité des vésicules détectés qu'on note par $\zeta_t \triangleq \{x_i^t, i = 1, \dots, n_t\}$. On s'intéresse maintenant à la dynamique du processus ponctuel spatio-temporel $\zeta \triangleq \{\zeta_t, t = 1, \dots, T\}$ qui est décrit par les positions des centres de masse des vésicules à chaque instant de la séquence. Pour chaque instant t , nous écrivons l'estimateur noyau de densité de probabilité de la position du processus spatial ζ_t comme suit :

$$\hat{f}_{h,K}^t(x) = \frac{1}{n_t h^d} \sum_{i=1}^{n_t} K\left(\frac{x - x_i^t}{h}\right) \quad (5.26)$$

où $K(\cdot)$ étant un noyau, h étant la bande passante, et $d = 3$ étant la dimension d'un point x . On suppose que $K(\cdot)$ s'écrit

$$K(x) = c_{k,d} k(\|x\|^2)$$

où $c_{k,d}$ est une constante de normalisation qui fait que $K(\cdot)$ s'intègre à 1, et $k(\cdot)$ est appelé son profil.

Cette estimation de la densité s'écrit pour ζ_{t+1} comme suit :

$$\hat{f}_{h,K}^{t+1}(x) = \frac{1}{n_{t+1}h^d} \sum_{i=1}^{n_{t+1}} K\left(\frac{x - x_i^{t+1}}{h}\right) \quad (5.27)$$

On considère que le centre de masse x_i^{t+1} résulte du déplacement du centre de masse x_i^t , on a donc la relation

$$x_i^{t+1} = x_i^t + d_i^t$$

où d_i^t étant le vecteur déplacement du i -ème vésicule. Le vecteur vitesse moyen dans l'intervalle de temps $[t, t + \Delta_t]$ du i -ème vésicule v_i^t et son vecteur déplacement sont liés par l'équations $v_i^t = d_i^t/\Delta_t$, où Δ_t est le pas d'échantillonnage temporel de la séquence de vidéo-microscopie. La formule (5.27) peut alors se récrire comme suit :

$$\hat{f}_{h,K}^{t+1}(x) = \frac{1}{n_{t+1}h^d} \sum_{i=1}^{n_{t+1}} K\left(\frac{x - x_i^t - d_i^t}{h}\right)$$

Si on suppose maintenant que les centres de masses de vésicules subissent des mouvements qui ne sont pas trop rapides, on peut alors écrire l'approximation de Taylor d'ordre 1 de $\hat{f}_{h,K}^{t+1}(x)$ comme suit :

$$\hat{f}_{h,K}^{t+1}(x) \cong \frac{2c_{k,d}}{n_{t+1}h^d} \sum_{i=1}^{n_{t+1}} \left[K\left(\frac{x - x_i^t}{h}\right) + \frac{1}{h^2} (d_i^t)^T (x - x_i^t) \left[-k'\left(\left\|\frac{x - x_i^t}{h}\right\|^2\right) \right] \right]$$

où x^T désigne la transposée du vecteur x . On définit la fonction unidimensionnelle $g(\cdot)$ comme suit :

$$g(x) = -k'(x)$$

et en utilisant $g(\cdot)$ pour profil, on définit le noyau $G(\cdot)$ comme suit

$$G(x) = c_{g,d} g(\|x\|^2) \quad (5.28)$$

et avec l'hypothèse que $n_{t+1} \cong n_t$, on trouve que

$$\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) \cong \frac{2c_{k,d}}{n_t h^{d+2}} \sum_{i=1}^{n_t} (d_i^t)^T (x - x_i^t) g\left(\left\|\frac{x - x_i^t}{h}\right\|^2\right)$$

qui peut également s'écrire comme suit :

$$\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) \cong \frac{2c_{k,d}}{c_{g,d} n_t h^{d+2}} \sum_{i=1}^{n_t} d(x_i^t)^T (x - x_i^t) G\left(\frac{x - x_i^t}{h}\right)$$

ou de manière plus formelle, on peut écrire

$$\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) = \frac{2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} (d_i^t)^T (x - x_i^t) G\left(\frac{x - x_i^t}{h}\right) + U^t(x)$$

où $U^t(x)$ modélise un bruit qui résulte de l'approximation à l'ordre 1 de $\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x)$ ainsi que certains autres facteurs comme supposer que $n_t \cong n_{t+1}$, etc. . .

On remarque que l'expression de droite dans la dernière formule ne dépend statistiquement pas de la directionnalité de la vitesse mais seulement du module de la vitesse⁹. On définit alors, sur toute la séquence (i.e. sur l'ensemble de toutes les positions x de l'image et de tous les instant t de la séquence), la variable aléatoire Z comme suit :

$$Z = |\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x)|$$

on a alors la formule

$$Z = \Phi(\|V\|, U)$$

où U est supposé un bruit indépendant de $\|V\|$.

Notre but est alors de restaurer la distribution de probabilité de la variable aléatoire $\|V\|$ à partir de distribution de la probabilité empirique de Z . On peut alors utiliser la démarche décrite dans la section précédente.

5.4.2 Restauration de la distribution empirique du champ de vitesse

De la même façon qu'on l'a fait pour l'estimation de la densité de probabilité empirique du module de la vitesse des vésicules dans une séquence de vidéo-microscopie, on peut le faire également pour l'estimation de la densité de probabilité empirique du champ de vitesse (module et directionnalité à la fois). En effet, on a

$$\begin{aligned} \nabla \left[\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) \right] &= \nabla \hat{f}_{h,K}^{t+1}(x) - \nabla \hat{f}_{h,K}^t(x) \\ \nabla \hat{f}_{h,K}^t(x) &= \frac{-2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} G\left(\frac{x - x_i^t}{h}\right) (x - x_i^t) \end{aligned}$$

⁹en remarquant qu'en considérant un seul point x_i^t à l'instant "t", quelle que soit alors la direction du vecteur du mouvement de x_i^t , pourvu que son module reste le même, un tel mouvement induit la réalisation de la même variable aléatoire Z en considérant tous les voxels du volume 3D de la séquence.

et

$$\nabla \hat{f}_{h,K}^{t+1}(x) = \frac{-2c_{k,d}}{c_{g,d}n_{t+1}h^{d+2}} \sum_{i=1}^{n_{t+1}} G\left(\frac{x - x_i^{t+1}}{h}\right)(x - x_i^{t+1})$$

et en écrivant le développement de Taylor d'ordre 1 de $\nabla [\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x)]$, on trouve

$$\begin{aligned} \frac{-2c_{k,d}}{c_{g,d}n_{t+1}h^{d+2}} \sum_{i=1}^{n_{t+1}} (x - x_i^{t+1})G\left(\frac{x - x_i^{t+1}}{h}\right) &= \frac{-2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} (x - x_i^t - d_i^t)G\left(\frac{x - x_i^t - d_i^t}{h}\right) \\ &= \frac{-2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} \left[(x - x_i^t)G\left(\frac{x - x_i^t - d_i^t}{h}\right) - d_i^t G\left(\frac{x - x_i^t - d_i^t}{h}\right) \right] \end{aligned}$$

En définissant maintenant la fonction unidimensionnelle $q(\cdot)$ comme suit :

$$q(x) = -g'(x)$$

et en utilisant $q(\cdot)$ pour profil, on définit le noyau $Q(\cdot)$ comme suit :

$$Q(x) = c_{q,d}q(\|x\|^2) \quad (5.29)$$

on trouve

$$\begin{aligned} G\left(\frac{x - x_i^t - d_i^t}{h}\right) &\cong G\left(\frac{x - x_i^t}{h}\right) - \frac{2}{h^2}(d_i^t)^T(x - x_i^t)g'\left(\left\|\frac{x - x_i^t}{h}\right\|^2\right) \\ &\cong G\left(\frac{x - x_i^t}{h}\right) + \frac{2c_{g,d}}{c_{q,d}h^2}(d_i^t)^T(x - x_i^t)Q\left(\frac{x - x_i^t}{h}\right) \end{aligned}$$

donc

$$\begin{aligned} (x - x_i^{t+1})G\left(\frac{x - x_i^{t+1}}{h}\right) &\cong G\left(\frac{x - x_i^t}{h}\right)(x - x_i^t) + \frac{2}{c_{q,d}h^2}(d_i^t)^T(x - x_i^t)Q\left(\frac{x - x_i^t}{h}\right)(x - x_i^t) \\ &\quad - G\left(\frac{x - x_i^t}{h}\right)d_i^t - \frac{2c_{g,d}}{c_{q,d}h^2}(d_i^t)^T(x - x_i^t)Q\left(\frac{x - x_i^t}{h}\right)d_i^t \end{aligned}$$

d'où

$$\begin{aligned} \nabla [\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x)] &\cong \frac{2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} G\left(\frac{x - x_i^t}{h}\right)d_i^t \\ &\quad + \frac{4c_{k,d}}{c_{g,d}c_{q,d}n_t h^{d+4}} \sum_{i=1}^{n_t} (d_i^t)^T(x - x_i^t)Q\left(\frac{x - x_i^t}{h}\right)d_i^t \\ &\quad - \frac{4c_{k,d}}{c_{q,d}n_t h^{d+4}} \sum_{i=1}^{n_t} (d_i^t)^T(x - x_i^t)Q\left(\frac{x - x_i^t}{h}\right)(x - x_i^t) \end{aligned}$$

On trouve alors

$$\begin{aligned} \nabla \left[\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) \right] &\cong \frac{2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} G\left(\frac{x-x_i^t}{h}\right) d_i^t \\ &\quad + \frac{4c_{k,d}}{c_{q,d}n_t h^{d+4}} \sum_{i=1}^{n_t} (d_i^t)^T (x-x_i^t) Q\left(\frac{x-x_i^t}{h}\right) \left[d_i^t - (x-x_i^t) \right] \end{aligned}$$

et d'où

$$\begin{aligned} \nabla \left[\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) \right] &\cong \frac{2c_{k,d}}{c_{g,d}n_t h^{d+2}} \sum_{i=1}^{n_t} G\left(\frac{x-x_i^t}{h}\right) d_i^t \\ &\quad + \frac{4c_{k,d}}{c_{q,d}n_t h^{d+4}} \sum_{i=1}^{n_t} (d_i^t)^T (x-x_i^t) Q\left(\frac{x-x_i^t}{h}\right) (x_i^{t+1} - x) \end{aligned}$$

On définit alors la variable (vecteur) aléatoire Z comme suit

$$Z = \nabla \left[\hat{f}_{h,K}^{t+1}(x) - \hat{f}_{h,K}^t(x) \right]$$

Il est évident que la variable aléatoire Z ne dépend statistiquement cette fois pas seulement du module de la vitesse mais aussi de sa directionnalité. On peut alors écrire que

$$Z = \Phi(V, U)$$

où V représente le vecteur vitesse, et U représente un effet aléatoire (bruit) qui est supposé indépendant statistiquement de V . De la même façon alors qu'on a procédé tout à l'heure pour restaurer la distribution empirique du module de la vitesse dans la séquence, on peut le faire pour restaurer la distribution du champ vitesse (module et directionnalité) en se basant sur la distribution empirique de la variable aléatoire Z calculée en considérant toute la séquence.

5.5 Expériences numériques

Le but de cette section est de montrer la faisabilité de l'approche que nous avons proposée dans ce chapitre pour l'estimation de la dynamique globale de vésicules dans une séquence de vidéo-microscopie. Nous nous restreignons ici à l'estimation de la distribution de probabilité empirique du module de la vitesse dans la séquence, mais celle du champ de vitesse (i.e. directionnalité et module de la vitesse en même temps) devrait pouvoir se faire de manière similaire sans souci.

5.5.1 Choix du noyau K et apprentissage de H à partir de séquences simulées

On définit le noyau $K(\cdot)$ qu'on a utilisé dans cette étude numérique pour l'estimation de la densité à noyau du processus ponctuel induit par les positions des centres de masse (gravité) ou centroïdes des sondes fluorescentes dans une séquence de vidéo-microscopie comme suit. On définit d'abord son profil noté $k(\cdot)$ comme suit :

$$k(u) \triangleq \begin{cases} \sqrt{u}(1 - \frac{\sqrt{u}}{3}), & 0 \leq u \leq 1 \\ 0, & u > 1. \end{cases}$$

et on définit maintenant $K(\cdot)$ comme suit :

$$K(x) \triangleq c_k k(\|x\|^2) = \begin{cases} c_k \|x\| (1 - \frac{\|x\|}{3}), & \|x\| \leq 1 \\ 0, & \|x\| > 1 \end{cases}$$

où c_k est une constante de normalisation qui fait que le noyau $K(\cdot)$ s'intègre à 1. Un tel choix du noyau $K(\cdot)$ est complètement empirique puisqu'il permet d'obtenir en pratique une estimation du noyau $H(\cdot, \cdot)$ qui est assez bien conditionnée contrairement à d'autres noyaux qu'on a pu tester (tels que le noyau d'Epanechnikov ou le noyau Gaussien). Il se peut bien entendu, par le biais d'une recherche plus élaborée, que l'on puisse trouver un autre noyau $K(\cdot)$ qui permette d'obtenir une estimation de $H(\cdot, \cdot)$ qui est encore mieux conditionnée. L'estimation de $H(\cdot, \cdot)$ est réalisée en utilisant des séquences simulées comme expliqué en section 5.2.1. Dans de telles séquences simulées, on simule le mouvement de n_v vésicules sur un réseau de microtubules. Ce dernier est généré aléatoirement dans la séquence en tirant aléatoirement dans le volume $3D$ de la séquence p segments de droite qu'on connecte ensuite de de façon aléatoire les uns aux autres de telle façon à obtenir un réseau à la fin. L'estimation de $H(\cdot, \cdot)$ qu'on présente dans cette section a été réalisée avec les paramètres suivants : un nombre de vésicules $n_v = 50$, et un nombre d'arêtes du réseau $p = 60$, une discrétisation de la vitesse qui est $\Delta_v = 0.5$ voxel/sec, une vitesse minimale $v_{min} = 0.5$ voxel/sec., une vitesse maximale $v_{max} = 10$ voxels/sec, et enfin, une discrétisation de la variable Z de façon à obtenir 100 bins sur son support. La figure 5.6 montre l'estimation $\hat{H}(\cdot, \cdot)$ du noyau $H(\cdot, \cdot)$ ainsi obtenue.

5.5.2 Estimation des écarts types σ_l dans le problème inverse (5.12)

Un point qu'on a pas encore discuté dans ce chapitre est l'estimation des niveaux du bruit σ_l dans le problème inverse (5.12), on se propose alors de le discuter dans cette section. Deux cas de figure sont à considérer :

Les écart-types $\alpha_{i,l}$ sont négligeables devant les écarts types δ_l

Ceci est le cas lorsque l'on a pris le soin via un choix approprié des paramètres (la discrétisation, le noyau $H(\cdot)$, etc...) d'estimer un noyau $H(\cdot, \cdot)$ qui est lisse, ou bien que l'on a pu lisser *a posteriori*. Dans ce cas, l'estimation de σ^l se réduit à l'estimation des δ^l . On suppose l'homoscédasticité du bruit ¹⁰ qui entache l'histogramme empirique \hat{f}_z de la variable aléatoire Z et on estime alors un seul σ . L'estimateur $\hat{\sigma}$ de σ qu'on propose est le suivant. On calcule $r_l = \hat{f}_z(l+2) - 2\hat{f}_z(l+1) + \hat{f}_z(l)$, $l = 1, \dots, L-2$. On calcule ensuite $\hat{\sigma}$ en utilisant la formule

$$\hat{\sigma} := \hat{\delta} = 0.6 \operatorname{median}_{l=1, \dots, [\frac{L}{4}]} \{|r_{4l}|\} \quad (5.30)$$

Un tel estimateur de σ suppose que f_z peut être approximée par une fonction linéaire par morceaux, et tente donc de supprimer d'abord la tendance linéaire qu'il y a entre deux points successifs de \hat{f}_z pour n'en garder que la composante r qui est due au bruit. Cette dernière suit une Gaussienne centrée d'écart type $\sqrt{6}$. Ensuite, il applique l'estimateur MAD [Rousseeu] pour estimer l'écart type des observations décorréélées $\{r_{4l}, l = 1, \dots, [\frac{L}{4}]\}$, et enfin, il normalise par $\sqrt{6}$ pour estimer σ selon la formule (5.30).

Les écarts types $\alpha_{i,l}$ sont quelconques

Dans le cas général, une estimation exacte de σ^l est généralement impossible puisqu'elle dépend de l'inconnue f_v . Nos expériences numériques montrent qu'une estimation exacte des σ^l n'est pas nécessaire pour pouvoir estimer efficacement f_v et une estimation approximative généralement suffit. En effet, une estimation approximative des σ^l revient à modifier légèrement la constante de pénalisation μ_M dans l'algorithme (1.23), mais ceci est sans conséquence grave pour l'estimation finale de f_v [Foster et al. (1994)]. On suppose alors

¹⁰Une telle hypothèse est largement vérifiée en pratique (voir e.g. figures 5.3 et 5.4)

que $\alpha_{i,l} \equiv \alpha_i$ qui est une hypothèse tout à fait justifiée due à la manière dont estime le noyau $H(\cdot, \cdot)$. On suppose aussi que l'on a $\delta_l \equiv \delta$ quel que soit $l = 1, \dots, L$. L'estimateur $\hat{\delta}$ de δ est alors calculé selon la formule (5.30). Pour ce qui est de l'estimateur de $\sum_{i=1}^M \alpha_i^2 (f_i^v)^2$, celui-ci peut par exemple être remplacé par $\nu = \min_{f, \|f\|_1=1} \sum_{i=1}^M \alpha_i^2 (f_i^v)^2$. On vérifie facilement que $\nu = \frac{1}{\sum_{i=1}^M \frac{1}{\alpha_i^2}}$. Enfin, on estime σ en utilisant la formule $\hat{\sigma} = \sqrt{\nu^2 + \hat{\delta}^2}$.

5.5.3 Evaluation des performances de l'algorithme de résolution du problème inverse (5.12), Evaluation de la faisabilité de l'approche en utilisant des séquences simulées

On évaluera d'abord les performances de la méthode qu'on a proposée en section 3 qui tente de résoudre le problème inverse (5.12) par le biais de la résolution du programme mathématique mixte (5.23). Pour cela, on simule différents histogrammes f_v^l auxquels on applique pour différentes valeurs de σ la transformation :

$$f_z^l = \hat{H} f_v^l + \sigma \epsilon \quad (5.31)$$

où \hat{H} est le noyau appris comme expliqué dans la sous-section précédente qui est de taille $N \times M$, et ϵ un vecteur aléatoire de taille N suivant une Gaussienne standard (moyenne nulle et matrice de variances-covariances qui est l'identité), ensuite, on applique pour différentes valeurs de σ l'algorithme de Branch & Bound qui tente de résoudre le programme mixte (5.23) afin de restaurer les f_v^l pour $l = 1, \dots, L$ et on calcule l'erreur moyenne commise pour la restauration des f_v^l . Cette dernière est donnée par la formule

$$\overline{\mathcal{R}} = \frac{1}{L} \sum_{l=1}^L \frac{\|A \tilde{f}_v^l - A f_v^l\|_1}{\min \{\|A \tilde{f}_v^l\|_0, \|A f_v^l\|_0\}}$$

pour un certain noyau de diffusion A de taille $M \times M$, et où $\|f\|_0 = \sum_{i=1}^M |f_i|$ et $\|f\|_1 = \sum_{i=1}^M \mathbb{I}\{v_i > 0\}$. Un tel noyau A est choisi de telle manière à tolérer une certaine erreur dans l'estimation des positions des bins positifs des f_v^l et de ne pas comparer les histogrammes \tilde{f}_v^l et f_v^l bin par bin. Une discussion intéressante autour de cette idée de comparaison d'histogrammes par diffusion peut être trouvée dans [Haibin2006]. Ici, on définit A comme suit : $A_{1,1} = \frac{2}{3}$, $A_{1,2} = \frac{1}{3}$, $A_{M,M} = \frac{2}{3}$, $A_{M-1,M} = \frac{1}{3}$, et pour tous $1 \leq i, j \leq M$ tels que

$(i, j) \notin \{(1, 1); (1, 2); (M - 1, M), (M, M)\}$, on a :

$$A_{i,j} = \begin{cases} \frac{1}{4}, j \equiv i - 1 \text{ ou } j \equiv i + 1 \\ \frac{1}{2}, j \equiv i \\ 0, \text{ sinon} \end{cases}$$

Pour ce qui est de la simulation d'un histogramme f_v^l , celle-ci est effectuée comme suit :

- Le nombre de bins positifs (i.e. pour lesquels la valeur de f_v^l correspondante est strictement supérieure à zéro) de f_v^l suit une loi de Poisson $\mathcal{P}(5)$.
- La position des bins positifs suit une loi uniforme sur l'intervalle $[1, M]$.
- La valeur de f_v^l correspondant à chaque bin positif suit une loi uniforme sur l'intervalle $[0, 1]$, et à la fin, on les normalise pour que leur somme fasse exactement 1.

On a fixé $L = 50$ et on a fait alors tourner l'algorithme de Branch & Bound de restauration des histogrammes $f_v^l, l = 1, \dots, L$ pour différentes valeurs de σ , et les valeurs de $\overline{\mathcal{R}}$ correspondantes sont représentées dans le tableau suivant :

| Valeur de σ | $\overline{\mathcal{R}}$ |
|----------------------|--------------------------|
| 10^{-6} | 0.0031 |
| 2.5×10^{-6} | 0.0031 |
| 5×10^{-6} | 0.0031 |
| 7.5×10^{-6} | 0.0033 |
| 10^{-5} | 0.0038 |
| 2.5×10^{-5} | 0.0045 |
| 5×10^{-5} | 0.0083 |
| 7.5×10^{-5} | 0.0101 |
| 10^{-4} | 0.0172 |
| 2.5×10^{-4} | 0.0325 |
| 5×10^{-4} | 0.0662 |
| 7.5×10^{-4} | 0.0794 |
| 10^{-3} | 0.1038 |

On remarque que pour $\sigma \leq 10^{-4}$, la qualité de la restauration des f_v^l est excellente puisque l'erreur moyenne que commet l'algorithme de restauration des f_v^l pour telles valeurs de σ : $\overline{\mathcal{R}} \lesssim 0.01$. On remarque aussi que pour $10^{-4} \leq \sigma \leq 2.5 \times 10^{-4}$, l'algorithme réussit à faire quand même une bonne restauration des f_v^l , toutefois, pour $\sigma \geq 5 \times 10^{-4}$, l'algorithme

s'écroule purement et simplement. Afin d'avoir une idée claire de l'effet du bruit sur la distribution f_z de la variable aléatoire Z , et de comprendre pourquoi un tel algorithme s'écroule pour $\sigma \geq 5 \times 10^{-4}$, on a réalisé l'expérience représentée dans la figure 5.6 dans laquelle on s'est fixé un histogramme f_v , et on simule une telle distribution f_z en utilisant la formule (5.31) pour différentes valeurs de σ ($0, 10^{-6}, 10^{-5}, 5 \times 10^{-4}, 10^{-4}, 10^{-3}, 10^{-2}$).

Pour ce qui est des résultats qu'on a obtenus sur des séquences simulées comme on l'a expliqué plus haut, un échantillon est représenté dans les figures 5.6 et 5.6.

5.6 Conclusion

Dans ce chapitre, nous avons proposé une approche n'utilisant pas le tracking pour estimer la dynamique globale (module de la vitesse, directionnalité du mouvement, etc..) de vésicules dans une séquence de vidéo-microscopie. Nous avons alors ramené le problème à celui du calcul de l'histogramme empirique d'une statistique Z (dont on a donné une formule possible) et à la résolution d'un problème linéaire inverse avec une contrainte de non-négativité de la solution pour inférer l'histogramme de la variable aléatoire d'intérêt (module de la vitesse ou module et direction de la vitesse en même temps) à partir de celui de Z . Nous avons proposé une solution numérique au problème inverse et on a ensuite démontré la faisabilité de notre approche en utilisant des données synthétiques.

Le peu de temps qu'on a seulement pu consacrer à cette partie de la thèse traitant l'analyse et de la comparaison de dynamiques de vésicules en vidéo-microscopie ne nous a pas permis d'aller au bout de l'approche et démontrer toute sa puissance certes, mais nous sommes persuadés qu'une telle démarche constitue un effort de recherche pionnier sur les méthodes n'utilisant pas le tracking (dont on connaît les limites) en vue d'une analyse de la dynamique globale de dizaines voire de centaines de vésicules impliqués dans des processus complexes tels que le trafic membranaire.

Dans ce chapitre, nous nous sommes restreints à l'étude de la dynamique globale de vésicules dans une séquence de vidéo-microscopie, mais rien n'empêche aussi de rendre l'approche plus ou moins local par le biais de la subdivision du temps ou de l'espace (e.g. segmenter le temps en plusieurs tranches et s'intéresser à la dynamique dans chaque tranche,

et/ou segmenter le volume spatial en plusieurs sous-volumes et s'intéresser à la dynamique dans chaque volume, etc...) . En effet, dès lors qu'on a récolté spatio-temporellement suffisamment d'échantillons de la variable aléatoire Z pour pouvoir estimer efficacement sa distribution de probabilité empirique, on peut procéder à l'analyse de la dynamique telle qu'elle décrite dans le présent chapitre.

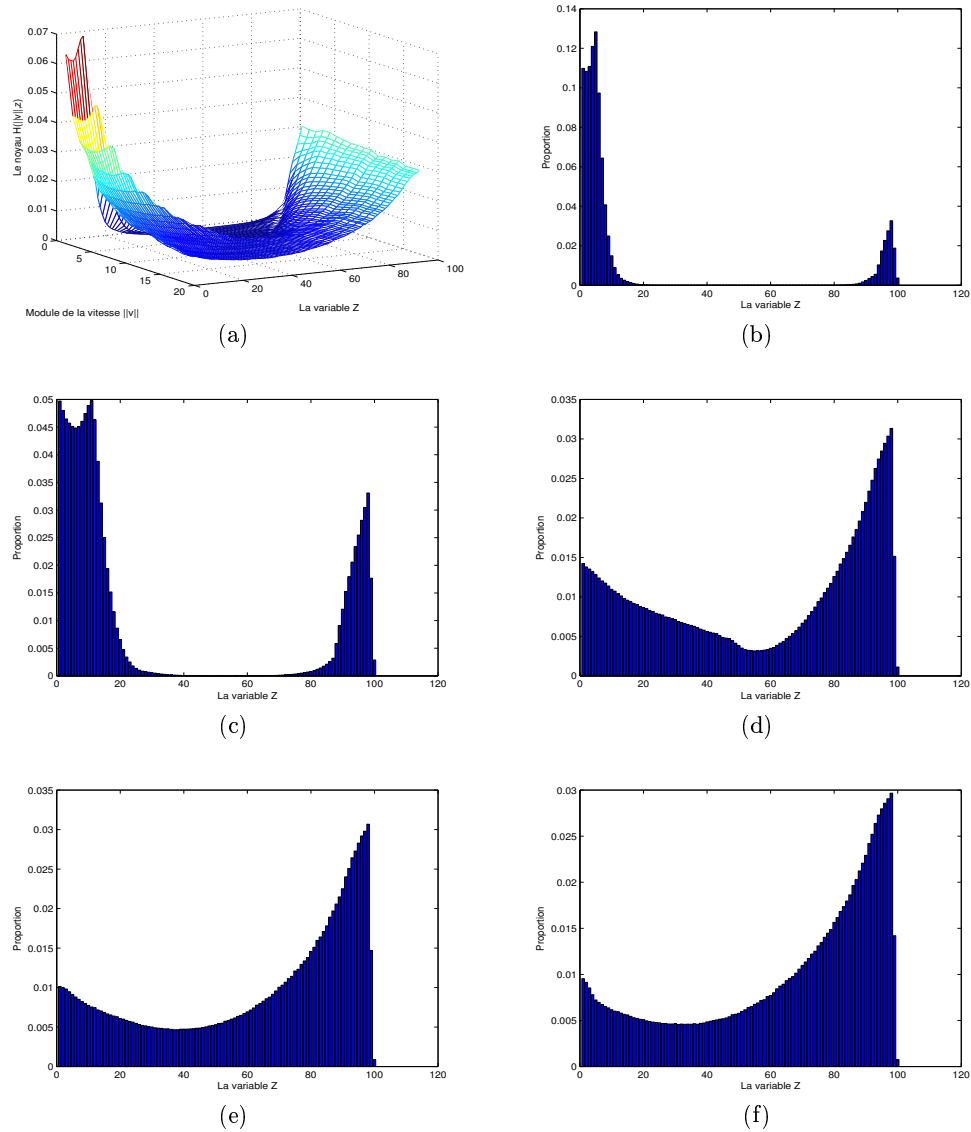


FIG. 5.2: Apprentissage du noyau $H(\cdot, \cdot)$ à partir de séquences simulées de taille $400 \times 400 \times 10 \text{ voxels} \times 120 \text{ sec.}$ avec $h = 10$, $v_{\min} = 0.5 \text{ voxels/sec.}$, $v_{\max} = 10 \text{ voxels/sec.}$, $\Delta_v = 0.5 \text{ voxel/sec.}$, et un nombre de bins $L = 100$ de la variable Z . (a) Le noyau estimé $\hat{H}(v, z)$; (b) $\hat{H}(2, \cdot)$; (c) $\hat{H}(4, \cdot)$; (d) $\hat{H}(6, \cdot)$; (e) $\hat{H}(8, \cdot)$; (f) $\hat{H}(10, \cdot)$.

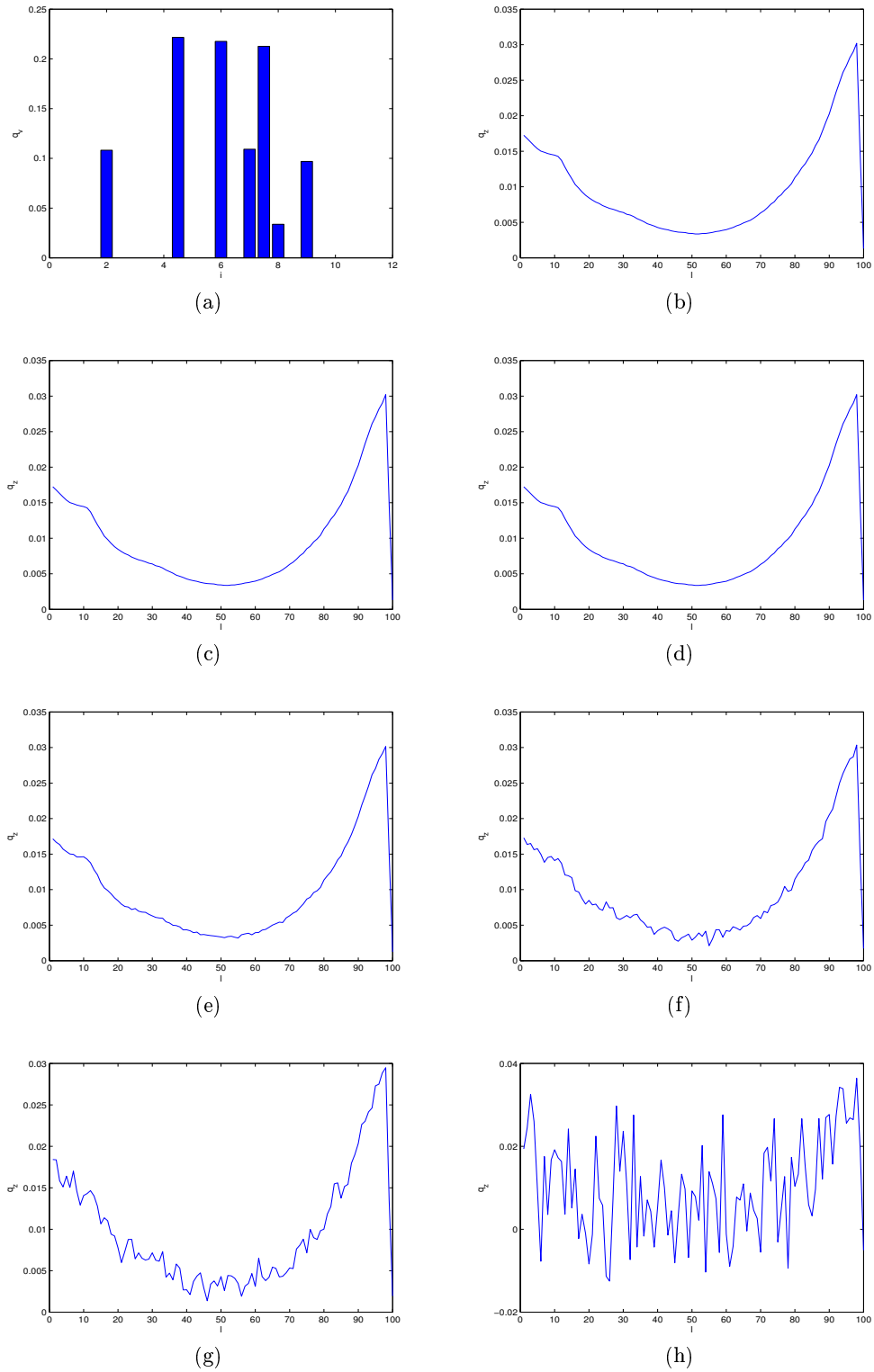


FIG. 5.3: Effet de σ sur l'histogramme empirique de la variable aléatoire Z . La figure (a) représente un histogramme de vitesse f_v pour lequel on simule l'histogramme de la variable aléatoire Z pour différentes valeurs de σ en utilisant la formule $f_z = \hat{H}f_v + \sigma\epsilon$. (b) $\sigma = 0$; (c) $\sigma = 10^{-6}$; (d) $\sigma = 10^{-5}$; (e) $\sigma = 10^{-4}$; (f) $\sigma = 5 \times 10^{-4}$; (g) $\sigma = 10^{-3}$; (h) $\sigma = 10^{-2}$

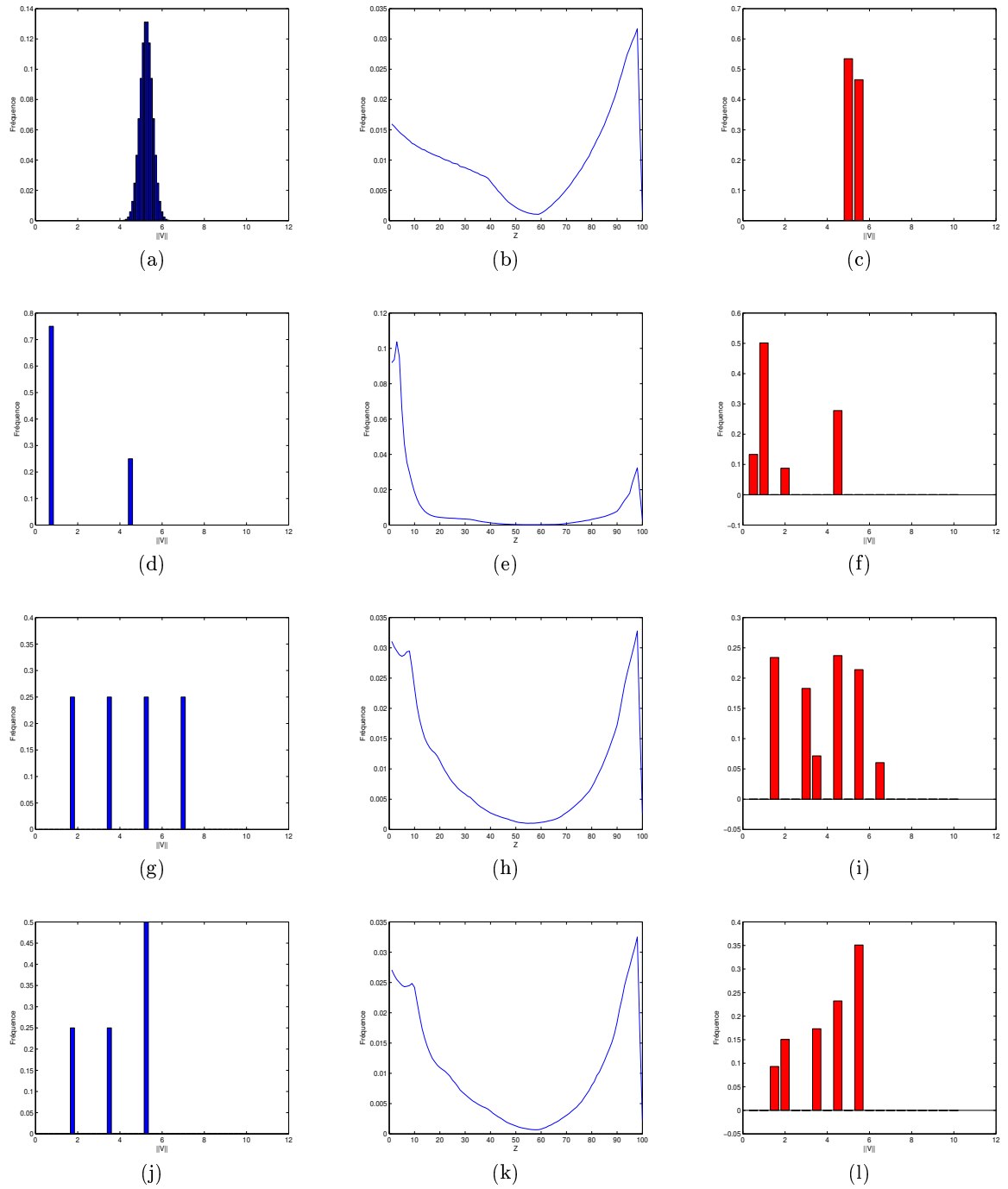


FIG. 5.4: Estimation de la dynamique globale (histogramme du module de la vitesse) de sondes fluorescentes dans une séquence simulée. (a)-(d)-(g)-(j) Les histogrammes du module de la vitesse utilisés pour générer les dynamiques de sondes fluorescentes dans quatre séquences simulées suivant des mélanges de diracs. Noter que toutes les vitesses utilisées dans les simulations n'ont pas été nécessairement apprises à l'étape d'estimation du noyau $H(\cdot, \cdot)$; (b)-(e)-(h)-(k) Les histogrammes de la variable Z correspondants; (c)-(f)-(i)-(l) Les estimations des histogrammes empiriques du module de la vitesse dans les séquences simulées en se basant sur l'histogramme empirique de Z .

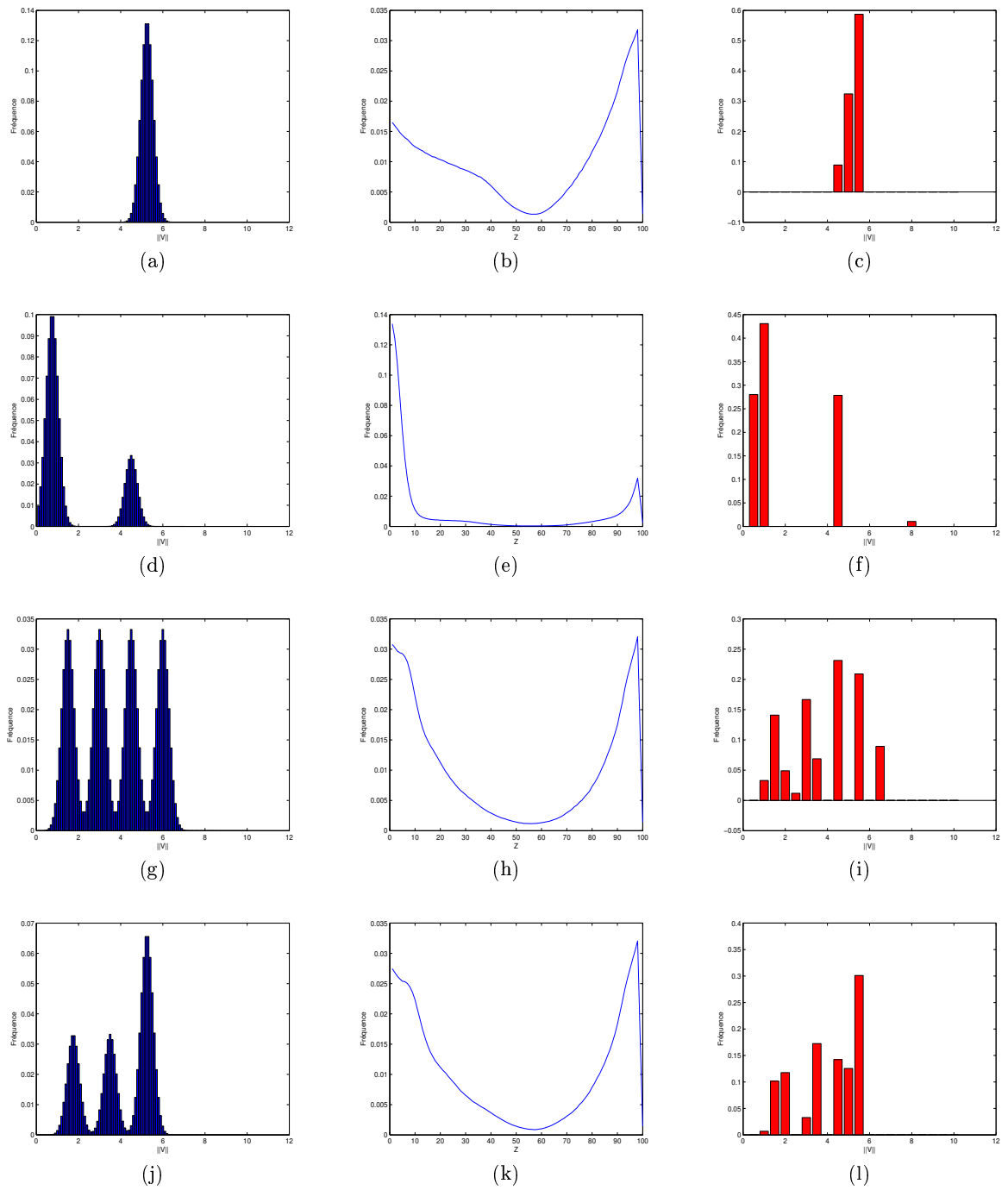


FIG. 5.5: Estimation de la dynamique globale (histogramme du module de la vitesse) de sondes fluorescentes dans des séquences simulées. (a)-(d)-(g)-(j) Les histogrammes du module de la vitesse utilisés pour générer les dynamiques de sondes fluorescentes dans quatre séquences simulées suivant des mélanges de Gaussiennes avec un écart-type de 0.3 voxel/sec. ; (b)-(e)-(h)-(k) Les histogrammes de la variable Z correspondants ; (c)-(f)-(i)-(l) Estimation des histogrammes empiriques du module de la vitesse dans les séquences simulées en se basant sur l'histogramme empirique de Z .

Chapitre 6

Conclusion générale

Cette thèse s'est inscrite dans le cadre d'un projet tri-partite (INRA, Institut Curie, INRIA) dont l'objectif était de proposer des méthodes facilitant l'analyse de dynamiques observées en microscopie de fluorescence multidimensionnelle. De fait, plusieurs méthodes d'analyse originales ont été développées par les différents partenaires de ce projet. Ces méthodes recouvrent des techniques de débruitage adaptatif spatio-temporel, des techniques de correction de photo-blanchiment, des techniques de détection d'objets mobiles, des techniques de tracking semi-automatique, des modélisations du trafic. Notre contribution concerne : a) la **correction du photo-blanchiment**, que nous avons sensiblement modifiée en autorisant aussi une correction des premiers instants observés, b) la **segmentation spatio-temporelle** des séquences d'images par des méthodes de sélection de modèle reposant sur une hypothèse locale de cohérence spatio-temporelle, c) l'analyse de dynamiques, au travers d'une **estimation de la distribution des vitesses** dans la séquence en construisant différentes statistiques basées sur le processus spatio-temporel des centres de masse des objets. Les nouveaux algorithmes développés dans cette thèse doivent encore être davantage validés et beaucoup de travail reste encore à faire et sur le plan expérimental et sur le plan théorique, notamment pour la dernière partie de cette thèse. Dans ce travail, nous avons cherché à renseigner une dynamique sans effectuer de tracking individuel des objets. Cette description peut sans doute être affinée si on échantillonne différemment l'espace et le temps d'une part et si on fusionne plusieurs statistiques d'autre part. Enfin, nous avons présenté

ici un seul usage, de la segmentation spatio-temporelle obtenue mais on peut imaginer bien d'autres utilisations, allant de la reconstruction du support du mouvement à la détection d'évènements rares au cours de la dynamique.

Pour finir, même s'il reste encore une validation de nos résultats à grande échelle sur des données images issues de la vidéo-microscopie, nous pensons qu'on peut obtenir des informations intéressantes sur les dynamiques de trafic membranaire sans avoir nécessairement recours au tracking et ceci en raison de l'intensité du trafic.

Bibliographie

- [1] H. Abdi. Bonferroni and Sidak corrections for multiple comparisons. In *N.J. Salkind (Ed.) : Encyclopedia of Measurement and Statistics. Thousand Oaks (CA) : Sage*, 2007.
- [2] F. Aguet, D. Van de Ville, and M. Unser. A maximum likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles. *Optics Express*, 13(26) :10503–10522, 2005.
- [3] H. Akaike. Information theory as an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory* : 267–281, *Akademiai, Kiado*, Budapest, 1973.
- [4] L. Alvarez, J. Weickert, and J. Sanchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1) :41–56, 2000.
- [5] N. Arhel, A. Genovesio, K.A. Kim, S. Miko, E. Perret, J.-C. Olivo-Marin, S. Shorte, and P. Charneau. Quantitative four-dimensional tracking of cytoplasmic and nuclear hiv-1 complexes. *Nature Methods*, 2006.
- [6] A.R. Barron, and T.M. Cover. Minimum complexity density estimation. *IEEE Trans. Inform. Theo.* 37 : 1034–1054, 1991.
- [7] A.R. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Prob. Theo. Relat. Fields.* 113 :301-415, 1999.
- [8] Y. Bar-Shalom and T.E. Fortmann. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11 :451–60, 1975.

- [9] I. Bechar, and A. Trubuil. A model selection approach to robust spatio-temporal analysis of dynamics in 4D fluorescence videomicroscopy. *In Proc. IEEE-EMBS'06*, NW, USA, 2006.
- [10] I. Bechar, and A. Trubuil. A bidimensional signal processing approach to vesicle trafficking analysis in $3D+T$ video-microscopy. *In Proc. MICCAI'06 Worksh. Bioph. Imag. Diag. Treat.*, Copenhagen, Denmark, 2006.
- [11] I. Bechar, and A. Trubuil. Reliable motion detection and Analysis in live-cell imaging. *In Proc. Int. Symp. on Biomedical Imaging (ISBI'07)*, Arlington, USA, 2007.
- [12] Y. Benjamini, and Y. Hochberg. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57 : 125–133, 1995.
- [13] Y. Benjamini, and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 2001.
- [14] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.
- [15] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with integrated completed likelihood, *IEEE Transactions on pattern Analysis and machine intelligence*, 719–725, 2000.
- [16] L. Birgé, and P. Massart. Gaussian Model Selection. *J. European Math. Soc.*, 3 :203–268, 2001a.
- [17] L. Birgé, and P. Massart. A generalized CP criterion for Gaussian model selection. *Techn. Rep.*, 647, Laboratoire de probabilités, Université Paris 6, 2001b.
Disponible sur le web à l'adresse :
- [http : //www.proba.jussieur.fr/mathdoc/preprints/index.html#2001](http://www.proba.jussieur.fr/mathdoc/preprints/index.html#2001)*
- [18] L. Birgé, and P.O. Massart. Minimal penalties for Gaussian model selection, *Prépublication*, 2004.

Disponible sur le web à l'adresse :

<http://www.math.u-psud.fr/massart/page6.html>

- [19] L. Birgé, and P. Massart. Minimal penalties for Gaussian model selection, *Prob. Theo. and Rel. Fields*, 138(1-2) : 33–73(41), 2007.
- [20] M. Black, and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with application in early vision. *Int. J. Computer Vision*, 19 : 57–92, 1996.
- [21] W.R. Blanding, P.K. Willett, and Y. Bar-Shalom. Offline and real-time methods for ml-pda track validation. *IEEE Trans. on Sig. Proc.*, 55(5) :1994–2006, 2007.
- [22] C. E. Bonferroni. Il calcolo delle assicurazioni su gruppi di teste. *In Studi in Onore del Professore Salvatore Ortu Carboni*, Rome, Italy, 1935.
- [23] S. Bonneau, M. Dahan, and L. Cohen. Single quantum dot tracking based on perceptual grouping using minimal paths in a spatiotemporal volume. *IEEE Trans. on Image Proc.*, 2005.
- [24] J. Boulanger, C. Kervrann, and P. Bouthemy. An adaptive statistical method for 4D fluorescence image sequences denoising with spatio-temporal discontinuities preserving. *Proc. Int. Conf. Image Processing (ICIP'05)*, Genova, Italy, 2005.
- [25] J. Boulanger, C. Kervrann, and P. Bouthemy. Estimation of dynamic background for fluorescence video-microscopy. *In Proc. Int. Conf. Image Processing (ICIP'06)*, Atlanta, USA, 2006.
- [26] J. Boulanger. *Estimation non-paramétrique et contributions à l'analyse de séquences d'images. Modélisation, simulation et estimation du trafic intra-cellulaire dans des séquences de video-microscopie*. PhD thesis, Université de Rennes I, 2006.
- [27] C.A. Bouman. Cluster : An Unsupervised Algorithm for Modeling Gaussian Mixtures, disponible sur le web :
<http://cobweb.ecn.purdue.edu/bouman/software/cluster/>.

- [28] P.-Y. Burgi and A.L. Yuille. Probabilistic motion estimation based on temporal coherence. *Neural Computation*, 2000.
- [29] D. Cai, K.J. Verhey, and E. Meyhöfer. Tracking single kinesin molecules in the cytoplasm of mammalian cell. *Biophys J.*, 92(12) :4137–44, 2007.
- [30] B.C. Carter, G.T. Shubeita, , and S.P. Gross. Tracking single particles : a user-friendly quantitative evaluation. *Physical Biology*, 2 :60–72, 2005.
- [31] M.K. Cheezum, W.F. Walker, and W.H. Guilford. Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophys. J.*, 2001.
- [32] D. Chetverikov and J. Verestoy. Comparative performance evaluation of four feature point tracking techniques. In *Proceedings of the Workshop of the Austrian Pattern Recognition*, pages 225–263, 1998.
- [33] J.-S. Cho and B.J. Yun. Selective-attention correlation measure for precision video tracking. *IEICE - Trans. Inf. Syst.*, E88-D(5) :1041–1049, 2005.
- [34] J.A. Conchello. Fluorescence Photobleaching Correction for Expectation-Maximization Algorithm. *Proced. SPIE*, 1995.
- [35] S. Courty, C. Luccardini, Y. Bellaiche, G. Cappello, and Dahan M. Tracking individual kinesin motors in living cells using single quantum-dot imaging. *Nano Letters*, 6(7) :1491–5, 2006.
- [36] I.J. Cox, S. Roy and S.L. Hingorani. Dynamic Histogram Warping of Image Pairs for Constant Image Brightness. *Proc. Intern. Conf. Imag. Proc.*, Washington, DC, USA, 1995.
- [37] I.I. Cox and S.L. Hingorani. An efficient implementation of reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. *IEEE Trans. Patt. Anal. Mach. and Int.*, 18 :138–50, 1996.
- [38] J.C. Crocker and D.G. Grier. Methods of digital video microscopy for colloidal studies. *Journal of Colloid and Interface Science.*, 179 :298–310, 1996.

- [39] E. Del Nery, S. Miserey-Lenkei, T. Falguières, C. Nizak, L. Johannes, F. Perez, and B. Goud. Rab6a and rab6a' gtpases play non-overlapping roles in membrane trafficking. *Traffic*, 2006.
- [40] A. Echard, F. Jollivet, O. Martinez, J.-J. Lacapere, A. Rousselet, I. Janoueix-Lerosey, and B. Goud. Interaction of a golgi-associated kinesin-like protein with rab6. *Science*, 1998.
- [41] J.A. Fessler and A.O. Hero. Suboptimality of the truncated svd for ill-conditioned inverse problems. Disponible sur le web à l'adresse :
- www.eecs.umich.edu/fessler/papers/files/abandon/svd.pdf*
- [42] T.E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE J. Oceanic Engineering*, 8(3) :173–84, 183.
- [43] J. Gelles, B.J. Schnapp, and M.P. Sheetz. Tracking kinesin-driven movements with nanometer scale precision. *Nature*, 331 :450–53, 1988.
- [44] A. Genovesio, T. Liedl, V. Emiliani, W.J. Parak, M. Coppey-Moisan, and J.-C. Olivo-Marin. Multiple particle tracking in 3-d+t microscopy : method and application to the tracking of endocytosed quantum dots. *IEEE Trans Image Process.*, 15(5) :1062–70, 2006.
- [45] A. Genovesio, Z. Belhassine, and J.-C. Olivo-Marin. Adaptive Gating in Gaussian bayesian Multi-Target Tracking. *Int. Conf. Image Processing (ICIP'04)*, 1 : 147–150, Singapore, 2004.
- [46] A. Genovesio, B. Zhang, and J.-C. Olivo-Marin. Interacting multiple model based method to track moving fluorescent biological spots. In *IEEE Int. Symp. on Biomed. Imaging*, volume 2, pages 1239–1242, 2004.
- [47] N. Gheissari, and A. Bab-Hadiashar. Model selection criteria in computer vision : Are they different? *Proc. 7th. Digit. Imag. Comput : Techn. and Applic.*, Sydney, 2003.

-
- [48] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, 1981.
- [49] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel. *Robust Statistics : The Approach Based on Influence Functions*, Wiley, 1986.
- [50] D.M.A. Haughton. On the choice of a mode to fit data from an exponential family. *The Annals of Statistics*, 16(1), 342–355, 1998.
- [51] E. Hill, M. Clarke, and F.A. Barr. The rab6-binding kinesin, rab6-kif1, is required for cytokinesis. *EMBO J.*, 2000.
- [52] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In J.M. Combes, A. Grossmann, and P. Tchamitchian, editors, *Wavelets*. Springer, 1989.
- [53] J. Howard. *Mechanics of motor proteins and the cytoskeleton*. 2001.
- [54] P.J. Huber. *Robust Statistics*, Wiley, New York, 1981.
- [55] S. Huet. Model selection for estimating the non-zero component of a gaussian vector. Dpt Math. 25, Université Paris Sud, 2002.
- [56] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering*, 82((D) :35–45, 1960.
- [57] K. Kanatani, and Y. Kanazawa. Sabilizing image mosaicing by the Geometric AIC. *2nd Workshop on Information-Based Induction Sciences*, Izu, Japan, 1999.
- [58] K. Kanatani. Motion segmentation by subspace separation : Model selection and reliability evaluation. *Int. Jour. Imag. Graphics*, 2 : 179–197, 2002.
- [59] C. Kervrann, D. Legland, and L. Pardini. Robust incremental compensation of the light attenuation with depth in 3D fluorescence microscopy. *J. Microscopy*, 214(3) :297–314, 2004.
- [60] G.A. Klein. *Theoretical aspects of motor protein induced filament depolymerisation*. PhD thesis, Institut fur Theoretische Physik, Technische Universitat Dresden, 2006.

- [61] C. Kural, H. Kim, S. Syed, G. Goshima, V.I. Gelfand, and P.R. Selvin. Kinesin and dynein move a peroxisome in vivo : a tug-of-war or coordinated movement. *Science*, 308 :1469–72, 2005.
- [62] B. Laurent, and P. Massart. Adaptive estimation of quadratic functional by model selection. *Ann. Statist.*, 28 :1302–1338, 2000.
- [63] E. Lebarbier, E. and M.H. Tristan. Le critère BIC : fondements théoriques et interprétation, Rapport de recherche INRIA No. RR-5315, 2004.
Disponible sur le web à l’adresse

<http://www.inria.fr/rrrt/rr-5315.html>
- [64] E. Lebarbier. Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing* 85 : 717–736, 2005.
- [65] S-C. Lee, and O. Bajcsy. Intensity correction of fluorescent confocal laser scanning microscope image by mean-weight filtering. *J. Microscopy*, 214(2) :122–136, 2005.
- [66] O.V. Lepski. On the problem of adaptive estimation in Gaussian white noise, *Theor. Probab. Applic.*, 35(3) : 459–470, 1990.
- [67] O.V. Lepski. Asymptotic minimax adaptive estimation : Upper bounds, *Theor. Probab. Applic.*, 36(4) : 645-659, 1991.
- [68] K. Li, E.D. Miller, L.E. Weiss, P.G. Campbell, and T. Kanade. Online tracking of migrating and proliferating cells imaged with phase-contrast microscopy. In *CVPRW06*, 2006.
- [69] J. Lippincott-Schwartz and G.H. Patterson. Development and use of fluorescent protein markers in living cells. *Science*, 2003.
- [70] Haiying Liu, Rama Chellappa, and Azriel Rosenfeld. Accurate dense optical flow estimation using adaptive structure tensors and a parametric model. In *ICPR '02 : Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 1*, page 10291, Washington, DC, USA, 2002. IEEE Computer Society.

- [71] S. Mallat. A theory of multiresolution signal decomposition : the wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11 :674–93, 1989.
- [72] P. Massart. Concentration inequalities and model selection, Ecole d’été de Saint Flour, 2006.
- [73] P. Massart. Concentration inequalities and model selection, *Lecture Notes in Mathematics*, Springer-Verlag, 1896, 2007.
- [74] G. McLachlan, and D. Peel. Finite mixture models. *Wiley Series in Probability and Statistics*, 2000.
- [75] E. Meijeing, I. Smal, and G. Danuser. Tracking in molecular bioimaging. *IEEE Signal Processing Magazine*, 23(3) :46–53, 2006.
- [76] H. Miki, M. Setou, K. Kaneshiro, and N. Hirokawa. All kinesin superfamily protein, kif, genes in mouse and human. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 98, page 7004, 2001.
- [77] A.J. Miller. Subset Selection in Regression. *Chapman and Hall*, 1990.
- [78] S. Miserey-Lenkei, F. Waharte, A. Boulet, M.H. Cuif, D. Tenza, A. El Marjou, G. Raposo, J. Salamero, L. Hélot, B. Goud, and S. Monier. Rab6-interacting protein 1 links rab6 and rab11 function. *Traffic*, 8(10) :1385–1403, 2007.
- [79] K. Miura and R. Pepperkok. Measurements of vesicle flow rates within cells by optical flow estimation. European Life Scientist Organization (ELSO) meeting, Sep/2004 Nice, France, 2004.
- [80] C.L. Mallows. Some comments on C_p . *Technometrics* 15 : 661–675, 1974.
- [81] S. Monier, F. Jollivet, I. Janoueix-Lerosey, L. Johannes, and B. Goud. Characterization of novel rab6-interacting proteins involved in endosome-to-tgn transport. *Traffic*, 2002.
- [82] F. Murtagh, J. Stark, and A. Bijaoui. Image restoration with noise suppression using multiresolution support. *Astronomy and Astrophysics*, 112 : 179–189, 1995.

- [83] N. Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. *J. of Cell Biol.*, 158(6) :1005–15, 2001.
- [84] O. Nestares, and D. Heeger. Robust multiresolution alignment of mri brain volumes. *Magnetic Resonance Medicine*, 43 : 705–715, 2000.
- [85] W. Ng, J.F. Li, S.J. Godsill, and J. Vermaak. A review of recent results in multiple target tracking. In *International Symposium on Image and Signal Processing and Analysis*, 2005.
- [86] J.M. Odobez, and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *J. Visual Communication Image Representation*, 6 : 348-365, 1995.
- [87] J.-C. Olivo-Marin. Extraction of spots in biological images using multi-scale products. *Pattern Recognition*, 35 :1989–1996, 2002.
- [88] F.J. Opdam, A. Echard, H.J. Croes, J.A. van den Hurk, R.A. van de Vorstenbosch, L.A. Ginsel, B. Goud, and J.A. Fransen. The small gtpase rab6b, a novel rab6 subfamily member, is cell-type specifically expressed and localised to the golgi apparatus. *J. Cell Sci.*, 113(15) :2725–35, 2000.
- [89] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2) :141–58, 2006.
- [90] V. Racine. *Quantification des dynamiques cellulaires par analyse de données de vidéo-microscopie 3D+T*. PhD thesis, Univ. P. et M. Curie, 2006.
- [91] V. Racine, A. Hertzog, J. Jouanneau, C. Kervrann, and J.-B. Sibarita. Multiple-target tracking of 3d fluorescent objects based on simulated annealing. In *IEEE Int. Symp. on Biomed. Imaging*, pages 6–9, 2006.
- [92] V. Racine, M. Sachse, J. Salamero, V. Fraissier, A. Trubuil, and J.-B. Sibarita. Visualization and quantification of vesicle trafficking on a 3d cytoskeleton network in living cells. *J. of Microscopy*, 2007.

- [93] J. Raissanen. Modelling by the shortest data description. *Automatica* 14 : 465–471, 1978.
- [94] J. Rink, E. Ghigo, Y. Kalaidzidis, and M. Zerial. Rab conversion as a mechanism of progression of early to late endosomes. *Cell*, 122 :735–749, 2005.
- [95] S. Roth and M.J. Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1) :33–50, 2007.
- [96] P.J. Rousseeuw, and A.M. Leroy. Robust Regression and Outlier Detection, Peter J. Wiley, 1987.
- [97] P.J. Rousseeuw, and C. Croux . Alternatives to the Median Absolute Deviation, *Journal of the American Statistical Association*, 88, 1993.
- [98] D. Sage, F. Hediger, S.M. Gasser, and M. Unser. Automatic tracking of particles in dynamic fluorescence microscopy. In *Proceedings of the International Symposium on Image and Signal Processing and Analysis*, pages 582–6, 2003.
- [99] I.F. Sbalzarini and P. Koumoutsakos. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struc. Biol.*, 151 :182–95, 2005.
- [100] M.C. Seabra, E.H. Mules, and A.N. Hume. Rab gtpases, intracellular traffic and disease. *Trends Mol. Med.*, 8(1) :284–288, 2002.
- [101] A. Seitz and T. Surrey. Processive movement of single kinesins on crowded microtubules visualized using quantum dots. *EMBO J.*, 25(2) :267–77, 2006.
- [102] I.K. Sethi and R. Jain. Finding trajectories of features points in a monocular image sequence. *IEEE Trans. on Pat. Anal. and Mach. Int.*, 9(1) :56–73, 1987.
- [103] C.E. Shannon. Communication Theory of Secrecy Systems. *Bell System Technical Journal*, 28 : 656–715, 1949.
- [104] P.Shi, and C.L. Tsai. A note on the unification of the Akaike information criterion. *J.R. Statist. Soc B* 60 : 551–558, 1998.

- [105] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics* 6 : 461–464, 1978.
- [106] J.-B. Sibarita. Deconvolution microscopy. *Adv. Biochem. Eng. Biotechnol.*, 95 :201–243, 2005.
- [107] J.L. Starck, F. Murtagh, and A. Bijaoui. *Image processing and data analysis*. Cambridge University Press, New York, NY, USA, 1998.
- [108] J.L. Starck, E. Pantin, and F. Murtagh. Deconvolution in astronomy : A review. *Publication of the Astronomical Society of the Pacific*, 114 :1051–69, 2002.
- [109] C. Steger. An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2) :113–125, 1998.
- [110] Stenmark H and Olkkonen VM. The rab gtpase family. *Genome Biol.*, 2(5), 2001.
- [111] R.L. Streit, M.L. Graham, and M.J. Walsh. Multitarget tracking of distributed targets using histogram-pmht. *Digital Signal Processing*, 12 :394–404, 2002.
- [112] Sugiura. Further Analysis of the data by Akaike’s information criterion and the finite corrections. *Communications in Statistics, Theory and Methods* A7 : 13–26, 1978.
- [113] M. They, Racine V., M. Piel, A. Pepin, A. Dimitrov, Y. Chen, J.-B. Sibarita, and M Bornens. Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. *Proc Natl Acad Sci U S A*, 103(52) :19771–6, 2006.
- [114] D. Thomann, D. R. Rines, P. K.O Sorger, and G. Danuser. Automatic fluorescent tag detection in 3d with super-resolution : application to the analysis of chromosome movement. *J. Microscopy*, 208(1) :49–64, 2002.
- [115] R.E. Thompson, D.R. Larson, and W.W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J.*, 2002.

- [116] S.M. Tonissen and R.J. Evans. Performance of dynamic programming technique for track-before-detect. *IEEE Trans. on Aerospace and Electronic Systems*, 32(4) :1440–51, 1996.
- [117] A. Trubuil, C. Kervrann, P. Maillet, and J. De Mey. Tracking centromeric regions in 4d. 2nd Int. Meeting & Workshop on Advanced Light Microscopy, Institut Pasteur, Paris, France, 2002.
- [118] Y. Vardi. Network tomography : estimation source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 91(433) :365–77, 1996.
- [119] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg. A survey on variational optic flow methods for small displacements. Technical Report 152, Saarland University, Saarbrücken, Germany, department of Mathematics, 2006.
- [120] J. White, L. Johannes, F. Mallard, A. Girod, S. Grill, S. Reinsch, P. Keller, B. Tzschaschel, A. Echard, B. Goud, and H.K. Stelzer. Rab6 coordinates a novel golgi retrograde transport pathway in live cells. *J. of Cell Biol.*, 147(4) :743–59, 1999.
- [121] M. Ye, and R.M. Haralick. Two-stage robust optical flow estimation Computer Vision. *Proceedings. IEEE Conference on Comp. Vis. Pat. Rec.*, 2 :623–628, 2000.
- [122] A. Yildiz, J.N. Forkey, S.A. McKinney, T. Ha, Y. Golman, and P. Selvin. Miosin v walks hand-over-hand : single fluorophore imaging with 1.5-nm localization. *Science*, 300 :2061–65, 2003.
- [123] J. Young, T. Stauber, E. del Nery, I. Vernos, R. Pepperkok, and T. Nilsson. Regulation of microtubule-dependent recycling at the trans-golgi network by rab6a and rab6a'. *Mol. Biol. Cell.*, 16(1) :162–77, 2005.
- [124] M. Zerial and H. McBride. Rab proteins as membrane organizers. *Nat. Rev. Mol. Cell Biol.*, 2(2) :107–17, 2001.
- [125] B. Zhang, M.J. Fadili, and J.-L. Starck. Multiscale variance stabilizing transform for multi-dimensional poisson count image denoising. pages 81–84, 2006.

-
- [126] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10) :1819–1829, 2007.
- [127] C. Zimmer, B. Zhang, A. Dufour, A. Thebaud, S. Berlemont, V. Meas-Yedid, and J.-C. Olivo-Marin. On the digital trail of mobile cells. *Signal Processing Magazine, IEEE*, 23(3) :54–62, 2006.

Annexe A

Estimation robuste de type M-estimation

Dans le cas général de l'estimation d'une fonction paramétrée $f_\lambda(\cdot)$ à partir des observations $y_i, i = 1, \dots, n$ aux points $x_i, i = 1, \dots, n$ respectivement, le critère robuste de type "M-estimateur" s'écrit :

$$\sum_{i=1}^n \rho\left(\frac{y_i - f_\lambda(x_i)}{\sigma_i}\right) \rightarrow \text{Min} \quad (\text{A.1})$$

On écrit alors la condition nécessaire d'optimalité de (A.1) comme suit :

$$\sum_{i=1}^n \frac{\partial f_\lambda(x_i)}{\partial \lambda} \frac{\rho'\left(\frac{y_i - f_\lambda(x_i)}{\sigma_i}\right)}{\sigma_i} = 0 \quad (\text{A.2})$$

où $\frac{\partial \rho}{\partial \lambda}$ désigne la dérivée par rapport à chaque composante scalaire λ_k du vecteur des paramètres $\lambda = (\lambda_k)_{k=1, \dots, K}$. En définissant maintenant la fonction poids $w(\cdot)$ comme suit : $w(e) = \frac{\rho'(e)}{e}$, la dernière formule devient :

$$\sum_{i=1}^n \frac{w_i}{\sigma_i^2} \frac{\partial f_\lambda(x_i)}{\partial \lambda} (y_i - f_\lambda(x_i)) = 0 \quad (\text{A.3})$$

où

$$w_i = w\left(\frac{y_i - f_\lambda(x_i)}{\sigma_i}\right)$$

L'estimation de λ s'effectue classiquement de manière itérative par le biais d'un schéma alterné de mise à jour des poids w_i et d'estimation de λ à poids fixés comme suit. L'estimation

précédente λ^{k-1} ($k = 1, 2, 3, \dots$) de λ est utilisée pour mettre à jour les poids w_i^{k-1} , $i = 1, \dots, n$ à l'itération courante à l'aide de la formule

$$w_i^{k-1} = w\left(\frac{y_i - f_{\lambda^{k-1}}(x_i)}{\sigma_i}\right)$$

et ces derniers sont ensuite injectés dans le système d'équations (A.3) comme suit :

$$\sum_{i=1}^n \frac{w_i^{k-1}}{\sigma_i^2} \frac{\partial f_{\lambda^k}(x_i)}{\partial \lambda} (y_i - f_{\lambda^k}(x_i)) = 0 \quad (\text{A.4})$$

pour mettre à jour l'estimation à l'itération k courante de λ . La condition pour qu'un tel schéma itératif converge vers la solution optimale est que l'estimation initiale λ^0 soit suffisamment proche de la solution optimale. Au fait, l'équation (A.4) peut être vue comme la condition nécessaire d'optimalité du critère quadratique pondéré

$$\sum_{i=1}^n w_i^{k-1} (y_i - f_{\lambda^k}(x_i))^2 \rightarrow \text{Min}$$

Pour ce qui est de l'estimation des écarts-types σ_i correspondant aux observations y_i , $i = 1, \dots, n$, en général, on considère que l'écart-type est homogène. Dans ce cas, il nous reste donc qu'à estimer un seul écart-type σ . Si λ^0 est l'estimation initiale de λ , on peut estimer σ à l'aide de l'estimateur MAD [Rousseeuw1987] comme suit :

$$\sigma = 1.483 \times \text{mediane}_{i=1, \dots, n} \left\{ (y_i - f_{\lambda^0}(x)) - \text{mediane}_{i=1, \dots, n} \{y_i - f_{\lambda^0}(x)\} \right\}$$

Diverses fonctions d'influence $\rho(\cdot)$ existent dans la littérature. Ces fonctions portent le nom de "M-estimateurs" et certaines parmi les plus utilisées ainsi que leurs fonctions poids sont résumées dans le tableau suivant :

| Méthode | Fonction objectif à minimiser | Fonction poids correspondante |
|--------------------------|--|--|
| <i>Moindres-carrés</i> | $\rho(e) = \frac{e^2}{2}$ | $w(e) = 1$ |
| <i>Métrique écrétée</i> | $\rho(e) = \begin{cases} \frac{e^2}{2}, & e \leq c \\ 0, & e > c \end{cases}$ | $w(e) = \begin{cases} 1, & e \leq c \\ 0, & e > c \end{cases}$ |
| <i>Huber</i> | $\rho(e) = \begin{cases} \frac{e^2}{2}, & e \leq c \\ c e - \frac{c^2}{2}, & e > c \end{cases}$ | $w(e) = \begin{cases} 1, & e \leq c \\ \frac{c}{ e }, & e > c \end{cases}$ |
| <i>Biweight de Tukey</i> | $\rho(e) = \begin{cases} \frac{e^2}{6} \left[1 - \left[1 - \left(\frac{e}{c} \right)^2 \right]^3 \right], & e \leq c \\ \frac{e^2}{6}, & e > c \end{cases}$ | $w(e) = \begin{cases} \left[1 - \left(\frac{e}{c} \right)^2 \right]^2, & e \leq c \\ 0, & e > c \end{cases}$ |
| <i>Leclerc</i> | $\rho(e) = \begin{cases} 1 - \exp\left(-\frac{e^2}{\lambda_1^2}\right), & e \leq 0 \\ 1 - \exp\left(-\frac{e^2}{\lambda_2^2}\right), & e > 0 \end{cases}$ | $w(e) = \begin{cases} \frac{2}{\lambda_1} \exp\left(-\frac{e^2}{\lambda_1^2}\right), & e \leq 0 \\ \frac{2}{\lambda_2} \exp\left(-\frac{e^2}{\lambda_2^2}\right), & e > 0 \end{cases}$ |